Fluctuations in the spectrum of random matrices

by

GIORGIO CIPOLLONI

January 2021

A thesis submitted to the
the Graduate School
of the
Institute of Science and Technology Austria
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Committee in charge:
Eva Benkova, Chair
László Erdős
Carlangelo Liverani
Robert Seiringer

IST AUSTRIA

The thesis of Giorgio Cipolloni, titled *Fluctuations in the spectrum of random matrices*, is approved by:

Supervisor: László Erdős, *IST Austria* ____________________

Committee Member: Robert Seiringer, *IST Austria* ____________________

Committee Member: Carlangelo Liverani, *University of Rome* ____________________

Defence chair: Eva Benkova, *IST Austria* ____________________

Signed page is on file

IST Austria Thesis, ISSN: 2663-337X

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Giorgio Cipolloni
January 2021

Signed page is on file

Abstract

In the first part of the thesis we consider Hermitian random matrices. Firstly, we consider sample covariance matrices XX^* with X having independent identically distributed (i.i.d.) centred entries. We prove a Central Limit Theorem for differences of linear statistics of XX^* and its minor after removing the first column of X. Secondly, we consider Wigner-type matrices and prove that the eigenvalue statistics near cusp singularities of the limiting density of states are universal and that they form a Pearcey process. Since the limiting eigenvalue distribution admits only square root (*edge*) and cubic root (*cusp*) singularities, this concludes the third and last remaining case of the Wigner-Dyson-Mehta universality conjecture. The main technical ingredients are an optimal local law at the cusp, and the proof of the fast relaxation to equilibrium of the Dyson Brownian motion in the cusp regime.

In the second part we consider non-Hermitian matrices X with centred i.i.d. entries. We normalise the entries of X to have variance N^{-1}. It is well known that the empirical eigenvalue density converges to the uniform distribution on the unit disk (*circular law*). In the first project, we prove universality of the local eigenvalue statistics close to the edge of the spectrum. This is the non-Hermitian analogue of the Tracy-Widom universality at the Hermitian edge. Technically we analyse the evolution of the spectral distribution of X along the Ornstein-Uhlenbeck flow for very long time (up to $t = +\infty$). In the second project, we consider linear statistics of eigenvalues for macroscopic test functions f in the Sobolev space $H^{2+\epsilon}$ and prove their convergence to the projection of the Gaussian Free Field on the unit disk. We prove this result for non-Hermitian matrices with real or complex entries. The main technical ingredients are: (i) local law for products of two resolvents at different spectral parameters, (ii) analysis of correlated Dyson Brownian motions.

In the third and final part we discuss the mathematically rigorous application of supersymmetric techniques (SUSY) to give a lower tail estimate of the lowest singular value of $X - z$, with $z \in \mathbf{C}$. More precisely, we use superbosonisation formula to give an integral representation of the resolvent of $(X-z)(X-z)^*$ which reduces to two and three contour integrals in the complex and real case, respectively. The rigorous analysis of these integrals is quite challenging since simple saddle point analysis cannot be applied (the main contribution comes from a non-trivial manifold). Our result improves classical smoothing inequalities in the regime $|z| \approx 1$; this result is essential to prove edge universality for i.i.d. non-Hermitian matrices.

Acknowledgment

I would like to first thank my PhD supervisor László Erdős. He has helped me to develop as a mathematician through his guidance and example. He was always available and spent much time actively working and discussing with me. I am also grateful to him for his continuous support, help and advice for my next steps in academia.

I would also like to thank my Masters supervisor Carlangelo Liverani giving me my first opportunity to enter mathematical research, and for his encouragement and support to pursue it further at PhD, as well as continuing to support me during my PhD and helping me progress to the next stage of my career.

Additionally, I am grateful to Robert Seiringer for his support and advice as part of my thesis committee for the past four years.

I would also like to thank Dominik Schröder, my past fellow PhD student and frequent collaborator, for our productive discussions and his contributions to our joint works, as well as his help and advice with other aspects of the PhD. In addition I am thankful to my other fellow colleagues Johannes Alt, Guillaume Dubach and collaborator Torben Krüger for their valuable discussions and collaborations.

I gratefully acknowledge the financial support from the *European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 665385* and my advisor's *ERC Advanced Grant No. 338804*.

Lastly, I would like thank my family and friends for their loving support over the past years, especially my parents Carla and Angelo and my girlfriend Louise and her family. My grandparents Anna, Peppe, Lello and Maria; my aunts and uncles Corrado, Fabio, Gianna, Nina and Paolo; my cousins Fulvio, Lavinia and Matilde. Also, my friends (in alphabetical order): Alberta, Alessandro C., Alessandro M., Antonella, Christian, Daniele, Dario, Davide, Diego, Federica, Francesco D., Francesco R., Giovanna, Giovanni, Giulio, Ilaria, Jacopo, Lorenzo C., Lorenzo P., Luka, Mario, Martina, Massimiliano, Matteo, Niko, Riccardo, Romolo, Roshan, Sebastian, Tommaso, Valerio R., Valerio S..

About the Author

After studying at the University of Rome Tor Vergata, Giorgio Cipolloni received his B.S. in *Mathematics* in 2015, and his M.S. degree in *Mathematics* in 2017. In September 2017, he joined IST Austria for his PhD studies in the research group of László Erdős. His research interests lie in probability theory and mathematical physics, with a focus on random matrices and dynamical systems.

Contents

List of Figures

Introduction 1

Eugene Wigner in 1955 [209] observed that energy level statistics of heavy nuclei are universal; the answer depends only on the symmetry type of the system. He also proposed real symmetric and complex Hermitian random matrices with centred independent identically distributed (i.i.d.) entries (modulo the symmetry), now known as *Wigner matrices*, as a mathematical model to describe this phenomenon. More recently, spectral properties of random matrices became important also in other areas of physics and mathematics: quantum chaos [32], disordered quantum systems [79], wireless communications [66], the error analysis of numerical algorithms [78], the zeros of the Riemann zeta function [123] and random neural networks [158].

1.1 Hermitian random matrices

In this thesis we work on *Wigner-type matrices*, which generalises *Wigner matrices*, introduced in [209], and *sample-covariance matrices* introduced by Wishart in [212] when the entries are Gaussian. To set our notation we introduce the Hermitian random matrix ensembles considered in this thesis:

GOE (Gaussian orthogonal ensemble): Symmetric matrices $G = G^t \in \mathbf{R}^{N\times N}$ such that the upper-triangular entries are centred i.i.d. real standard Gaussian random variables with $\mathbf{E}g_{ab}^2 = 1/N$, for $a < b$, and the diagonal entries are distributed as i.i.d. centred real Gaussian random variables satisfying $\mathbf{E}\, g_{aa}^2 = 2/N$.

GUE (Gaussian unitary ensemble): Hermitian matrices $G = G^* \in \mathbf{C}^{N\times N}$ such that the upper-triangular entries are centred i.i.d. standard complex Gaussian random variables with $\mathbf{E}|g_{ab}|^2 = 1/N$, and the diagonal entries are distributed as centred i.i.d. real standard Gaussian random variables satisfying $\mathbf{E}\, g_{aa}^2 = 1/N$.

Wigner matrices: Matrices $W = W^* \in \mathbf{C}^{N\times N}$ such that the upper-triangular entries $\{w_{ab}|a < b\}$ are i.i.d. real or complex random variables with $\mathbf{E}w_{ab} = 0$, $\mathbf{E}|w_{ab}|^2 = 1/N$, and the diagonal entries w_{aa} are i.i.d. with $\mathbf{E}\, w_{aa} = 0$, $c/N \leq \mathbf{E}|w_{aa}|^2 \leq$

C/N for some positive N-independent constants c, C. In addition, in the complex case $\mathbf{E}w_{ab}^2 = 0$.

Wigner-type matrices: Matrices $H = A + W \in \mathbf{C}^{N\times N}$ such that $A = A^* = \mathbf{E}\,H$ is diagonal and the upper-triangular and diagonal entries $\{w_{ab}|a \leq b\}$ of $W = W^*$ are independent and satisfy $\mathbf{E}\,w_{ab} = 0$ and $c/N \leq s_{ab} \leq C/N$, with $s_{ab} := \mathbf{E}\,|w_{ab}|^2$, for some positive N-independent constants $c, C > 0$.

Wishart matrices: Matrices $XX^* \in \mathbf{C}^{N\times N}$, where the entries of the $N \times M$ matrix X are distributed as i.i.d. Gaussian random variables with zero expectation and $\mathbf{E}|x_{ab}|^2 = (MN)^{-1/2}$. In addition, $\mathbf{E}x_{ab}^2 = 0$ in the complex case.

Sample-covariance matrices: Matrices $H = XX^* \in \mathbf{C}^{N\times N}$, where the entries of the $N \times M$ matrix X are distributed as i.i.d. real or complex random variables with zero expectation and $\mathbf{E}|x_{ab}|^2 = (MN)^{-1/2}$. In addition, $\mathbf{E}x_{ab}^2 = 0$ in the complex case.

The scaling in the random matrix ensembles presented above is such that the spectrum is contained in the interval $[-2-\epsilon, 2+\epsilon]$, for any small $\epsilon > 0$, with very high probability for large N.

The spectral properties of random matrices are analysed at three different scales: global scale, mesoscopic scale, microscopic scale. Now we explain the relevant questions on these three scales.

1.1.1 Global scale

In this section we focus on the global scale, i.e. we discuss the convergence in the large N limit of the empirical eigenvalue density (see (1.1) below). This section is divided into two subsections. In Section 1.1.1.1, we first discuss the limiting density distribution of the eigenvalues of Wigner-type matrices and then we present the classification theorem for the density of Wigner-type matrices. In Section 1.1.1.2 we focus on sample covariance matrices explaining the changes compared to Wigner matrices.

1.1.1.1 Wigner-type matrices

In order to describe the techniques used in the analysis of Hermitian matrices we first focus on the simpler Wigner matrices and then we comment on Wigner-type matrices. Let W be a Wigner matrix and denote by $\lambda_1 \leq \cdots \leq \lambda_N$ its eigenvalues. In order to study spectral properties of W, we consider the *empirical spectral density distribution* (ESD) denoted by μ_N and defined as

$$\mu_N := \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i}. \tag{1.1}$$

In [210] Wigner showed that the ESD of eigenvalues of Wigner matrices converges weakly to the celebrated Wigner semicircle law

$$\rho_{\mathrm{sc}}(x)\mathrm{d}x := \frac{\sqrt{(4-x^2)_+}}{2\pi}\mathrm{d}x.$$

Moment method

The proof of the convergence of μ_N to the semicircle law has been first achieved using the *moment method* [210]. In this method one proves that the expectation of the trace of W^k (properly rescaled) converges to the moments of the semicircle law, i.e. one proves that

$$\mathbf{E}\frac{1}{N}\mathrm{Tr}W^k = \mathbf{E}\int x^k \,\mathrm{d}\mu_N(x) = \int_{-\infty}^{+\infty} x^k \rho_{\mathrm{sc}}(x)\,\mathrm{d}x + \mathcal{O}\left(\frac{1}{N}\right).$$

The moments of the semicircle law are given by

$$\int_{-\infty}^{+\infty} x^k \rho_{\mathrm{sc}}(x)\,\mathrm{d}x = \begin{cases} C_{k/2} & \text{if } k \text{ is even,} \\ 0 & \text{if } k \text{ is odd,} \end{cases}$$

where

$$C_k := \frac{1}{k+1}\binom{2k}{k}$$

is the k-th *Catalan number*. The moment method identifies the leading term of $\mathbf{E}\frac{1}{N}\mathrm{Tr}W^k$ with C_k via a graphical expansion.

Resolvent method

Recently, in order to analyse the ESD μ_N on scales much smaller than order one, the *resolvent method* has been developed. In this method one identifies the limiting distribution μ of the empirical spectral distribution μ_N through its *Stieltjes transform*, which uniquely determines the measure. Given a measure μ, its *Stieltjes transform* is given by

$$m_\mu(z) = \int_{\mathbf{R}} \frac{1}{x-z}\mathrm{d}\mu(x), \qquad z \in \mathbf{C}\setminus\mathbf{R},$$

where z is the *spectral parameter*. For μ_N, by spectral decomposition, we have that

$$m_{\mu_N}(z) = \frac{1}{N}\sum_{i=1}^N \frac{1}{\lambda_i - z} = \frac{1}{N}\mathrm{Tr}G(z),$$

with $G(z) := (W-z)^{-1}$ being the resolvent of W. Then the global law is equivalent to

$$\lim_{N\to+\infty} m_{\mu_N}(z) = m_{\mathrm{sc}}(z)$$

with $m_{\mathrm{sc}}(z)$ being the Stieltjes transform of the semicircle law. The limiting $m_{\mathrm{sc}}(z)$ can be characterised as the unique solution of the equation

$$-\frac{1}{m_{\mathrm{sc}}(z)} = z + m_{\mathrm{sc}}(z), \qquad \Im[m_{\mathrm{sc}}]\Im z > 0.$$

This is a special case of the *matrix Dyson equation* (MDE) presented in (1.2) below (see [5, 113]).

For more general random matrix ensembles $H = W + A$, with H being a Wigner type matrix and $A = \mathbf{E}H$, we compute the deterministic approximation of the resolvent $G(z) = (H-z)^{-1}$ by the solution of the *matrix Dyson equation* (e.g. see [5, 113]):

$$-M^{-1} = z - A + \mathcal{S}[M], \qquad \Im[M]\Im z > 0, \tag{1.2}$$

where $\mathcal{S}[\cdot]$ is the *covariance operator* defined by

$$\mathcal{S}[R] := \mathbf{E} WRW, \qquad R \in \mathbf{C}^{N\times N}.$$

Then one can prove that $G \approx M$ in isotropic and average sense (see (1.6) for a precise statement). In [113] (see also [5]) it has been proven that the equation (1.2) admits a unique solution. The limiting *self-consistent density of states (scDos)* $\mu(\mathrm{d}x) = \rho(x)\mathrm{d}x$ of H is obtained from (1.2) by

$$\rho(x) := \frac{1}{\pi} \lim_{\eta \searrow 0} \langle \Im M(x + \mathrm{i}\eta) \rangle, \tag{1.3}$$

where $\langle \cdot \rangle := N^{-1}\mathrm{Tr}[\cdot]$.

Classification of the self-consistent density of states

In the classification theorem [14] it is shown that M is $1/3$-Hölder continuous in z, and that ρ has the following properties:

(i) $\operatorname{supp} \rho$ consists of finitely many compact intervals.

(ii) ρ is real analytic whenever $\rho > 0$.

(iii) If $\mathfrak{e} \in \partial \operatorname{supp} \rho$ is an *edge point*, then $\rho(\mathfrak{e} \pm x) = c\sqrt{x} + \mathcal{O}(\sqrt{x})$ and $\rho(\mathfrak{e} \mp x) = 0$ for $0 < x \ll 1$ and some constant $c = c(\mathfrak{e}) > 0$.

(iv) If $\mathfrak{c} \in \operatorname{supp} \rho$ with $\rho(\mathfrak{c}) = 0$ is a *cusp point*, then $\rho(\mathfrak{c} + x) = c|x|^{1/3} + \mathcal{O}(|x|^{1/3})$ for some constant $c > 0$.

(v) ρ cannot have other singularities than edges and cusps.

The spectral regimes corresponding to (ii), (iii) and (iv) are called *bulk*, *edge* and *cusp regime* of the scDos.

1.1.1.2 Sample covariance matrices

The analysis of sample covariance matrices follows analogous steps to Wigner matrices, hence in this section we will only explain the differences. Consider a *sample covariance matrix* XX^*, and denote by $0 \le \mu_1 \le \dots \le \mu_N$ its eigenvalues. Then the *empirical spectral density*, defined as

$$\widehat{\mu}_N := \frac{1}{N} \sum_{i=1}^N \delta_{\mu_i},$$

converges to the well known Marchenko-Pastur law [144]:

$$\rho_\phi(\mathrm{d}x) = \rho_\phi(x)\,\mathrm{d}x + (1-\phi)_+ \delta(\mathrm{d}x), \qquad \rho_\phi(x) = \frac{\sqrt{\phi}}{2\pi} \frac{\sqrt{[(x-\gamma_-)(\gamma_+ - x)]_+ x^2}}{,} \tag{1.4}$$

where

$$\gamma_\pm := \sqrt{\phi} + \frac{1}{\sqrt{\phi}} \pm 2,$$

and ϕ is so that

$$\frac{M}{N} \to \phi \in (0, +\infty), \quad \text{as} \quad N, M \to +\infty.$$

Note that $\gamma_\pm$ are the edges of the limiting spectrum and that the eigenvalues of XX^* are always non-negative. In addition, for $\phi = 1$ there is an accumulation of eigenvalues close to zero. Similarly to Wigner matrices, the Stieltjes transform of $\rho_\phi(\mathrm{d}x)$ is given by the unique solution of

$$-\frac{1}{m_\phi} = z + z\phi^{-1/2} m_\phi - (\phi^{1/2} - \phi^{-1/2}), \qquad \Im m_\phi(z)\Im z > 0.$$

Note that m_ϕ is related to the Stieltjes transform of the Wigner semicircle law by

$$w_\phi(z) = \sqrt{\phi}(1 + z m_{\phi^{-1}}(z)),$$

where w_ϕ is the Stieltjes transform of the semicircle law of radius 2 but shifted to be centred at $\phi^{1/2} + \phi^{-1/2}$.

1.1.2 Mesoscopic scale

In this section we focus on the mesoscopic scale. For concreteness we consider only Wigner-type matrices, the analysis for sample covariance matrices is analogous and thus omitted. According to (1.3), proving a global law consists in proving a bound on $\Im\langle G - M\rangle$ for $z = E + \mathrm{i}\eta$ for all N-independent $\eta < 1$, i.e. one has to prove that for all $\eta < 1$ (with η independent of N) it holds

$$\Im\langle G - M\rangle = \frac{1}{N}\sum_{i=1}^{N} \frac{\eta}{(\lambda_i - E)^2 + \eta^2} - \Im\langle M\rangle \to 0 \tag{1.5}$$

as $N \to +\infty$. Note that the main contribution to the summation in (1.5) comes from $\sim \eta N$ eigenvalues around the energy E as a consequence of the approximate delta function $\eta/[(\lambda_i - E)^2 + \eta^2]$ on a scale η. It is then natural to ask if the convergence in (1.5) still holds choosing η depending on N. In particular, since in the bulk of the spectrum of the limiting density the level spacing (distance between two neighbouring eigenvalues) is proportional to N^{-1}, we expect that the convergence in (1.5) holds for any $\eta \gg N^{-1}$. This is optimal, indeed we do not expect that the concentration result (1.5) holds for $\eta \sim N^{-1}$ since on this scale the fluctuation of single eigenvalues matter and so one cannot expect the convergence of $\Im\langle G\rangle$ to a deterministic quantity.

In the last decade optimal local laws have been proven uniformly in the spectrum of Wigner-type matrices, or even of matrices with some correlation structure (e.g. see[15, 83, 84]):

$$|\langle \boldsymbol{x}, (G - M)\boldsymbol{y}\rangle| \prec \|\boldsymbol{x}\|\|\boldsymbol{y}\|\sqrt{\frac{\rho}{N\eta}}, \qquad |\langle B(G - M)\rangle| \prec \frac{\|B\|}{N\eta}, \tag{1.6}$$

where $\prec$ is a suitable notion of high probability bound up to N^ϵ-factors (e.g. see Definition 3.4.2 for the precise definition), $\rho(z) := \pi^{-1}\Im\langle M(z)\rangle$, and $\boldsymbol{x}, \boldsymbol{y}, B$ are arbitrary deterministic vectors and matrices. Note that the bound in average sense in (1.6) is of order $(\sqrt{N\eta})^{-1}$ better (in the bulk, i.e. when $\rho \sim 1$) than the one on individual entries $(G - M)_{aa}$ (*fluctuation averaging* feature).

We now explain the three main implications of the local laws in (1.6): eigenvalue rigidity, eigenvector delocalization, absence of eigenvalues outside the limiting spectrum.

1.1.2.1 Eigenvalue rigidity

Let H be a Wigner-type matrix, denote by $\lambda_1 \le \cdots \le \lambda_N$ its eigenvalues and by ρ the limiting distribution of the eigenvalues. We define the classical eigenvalue locations (*quantiles*) by

$$\int_{-\infty}^{\gamma_i} \rho(x)\,\mathrm{d}x = \frac{i}{N}, \qquad i \in \{1, \dots, N\}.$$

Using Cauchy-integral formula, by a standard argument (e.g. see [81, Lemma 7.1, Theorem 7.6] or [93, Section 5]), one can prove that the eigenvalues λ_i are *rigid* in the following sense

$$|\lambda_i - \gamma_i| \prec \eta_{\mathrm{f}}(\gamma_i),$$

with $\eta_{\mathrm{f}} = \eta_{\mathrm{f}}(\gamma_i)$ being the fluctuation scale around γ_i, which is implicitly defined by

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

The fluctuation scale η_{f} is of order N^{-1} in the bulk, $N^{-2/3}$ at the edge, and $N^{-3/4}$ at the cusp. The fact that the eigenvalues fluctuate on these scales is a consequence of their strong correlations.

1.1.2.2 Eigenvector delocalization

As a consequence of the local law (1.6) for the entries of the resolvent, it is possible to conclude that the ℓ^2-normalized eigenvectors $\mathbf{u}_i$ of a Wigner-type matrix are fully delocalized, in the sense that $|\mathbf{u}_i(a)| \prec N^{-1/2}$ for any $1 \le a \le N$. This phenomenon is called *eigenvector delocalization* because the mass of the eigenvector $\mathbf{u}_i$ is (almost) equally distributed to all its entries $\mathbf{u}_i(a)$. This is an easy consequence of the local law (1.6):

$$|\mathbf{u}_i(a)|^2 \le C \sum_{a=1}^{N} \frac{\eta^2 |\mathbf{u}_i(a)|^2}{(E-\lambda_a)^2 + \eta^2} = C\eta (\Im G)_{ii} \prec \eta,$$

for some constant $C > 0$, choosing $\eta = N^{-1+\epsilon}$ and $E = \gamma_a$, for some arbitrary small $\epsilon > 0$.

1.1.2.3 Absence of eigenvalues outside of the limiting spectrum

By the local law (1.6) and a stronger version of (1.6) outside the spectrum in the *edge* (see [15, Eq (2.6c)]) and the *cusp* (see [83, Eq. (2.8b)]) regime one can exclude the existence of eigenvalues well outside the support of the limiting eigenvalue density, i.e. one can prove that with very high probability there are no eigenvalues at a distance much bigger than $N^{-2/3}$ from the spectral edges. Additionally, in case of the support of the limiting density consists of several components, one can also prove that the number of eigenvalues in each component is deterministic with very high probability.

1.1.3 Linear statistics

The global law and the local law on mesoscopic scales prove that each matrix element of the resolvent converges to a deterministic quantity. This phenomenon can be rephrased in terms of linear statistics. Let W be a Wigner matrix and denote by $\lambda_1, \dots, \lambda_N$ its eigenvalues, then we define the *centred linear statistics* by

$$L_N(f) := \sum_{i=1}^N f(\lambda_i) - \mathbf{E}\sum_{i=1}^N f(\lambda_i), \tag{1.7}$$

where $f(x) = f_{a,E}(x) = g(N^a(x-E))$, with g a smooth compactly supported test function, $a \in [0, 1-\delta]$ and $|E| \le 2-\epsilon$, for some small fixed $\epsilon, \delta > 0$. To make the presentation clearer we focus on linear statistics of eigenvalues in the bulk of the limiting spectrum; analogous results hold at the edges of the spectrum too.

To analyse the linear statistics in (1.7) we will often use the following convenient integral representation of any smooth function (Helffer-Sjöstrand formula):

$$f(\lambda) = \frac{1}{\pi}\int_{\mathbf{C}} \frac{\partial_{\overline{z}} f_{\mathbf{C}}(z)}{\lambda - z}\,\mathrm{d}^2 z, \qquad \lambda \in \mathbf{R}, \tag{1.8}$$

with $\mathrm{d}^2 z := \mathrm{d}\Re z \mathrm{d}\Im z$. Here $f_{\mathbf{C}}$ is an almost analytic extension of f, defined by

$$f_{\mathbf{C}}(z) = f_{\mathbf{C}}(x + \mathrm{i}\eta) := [f(x) + \mathrm{i}\eta\partial_x f(x)]\chi(N^a\eta),$$

with χ being a smooth cut-off function equal to one on $[-5, 5]$ and identically zero on $[-10, 10]^c$. Note that if the test function f is in C^k then we can define an almost analytic extension of f such that

$$\partial_{\overline{z}} f_{\mathbf{C}}(z) = \mathcal{O}\left(|\Im z|^k\right). \tag{1.9}$$

By the averaged local law in (1.6), choosing $B = I$, and Helffer-Sjöstrand formula (1.8) it is easy to see that

$$\left| \frac{1}{N}\sum_{i=1}^N N^a f(\lambda_i) - \mathbf{E}\frac{1}{N}\sum_{i=1}^N N^a f(\lambda_i)\right| \prec \frac{N^a}{N}. \tag{1.10}$$

Note the unusually small error term N^{-1+a} in (1.10). By standard CLT scaling one would expect the fluctuations around the expectation to be of order $N^{-(1-a)/2} = N^{-1}N^a(N^{1-a})^{1/2}$ (since the sums in (1.10) effectively involve N^{1-a} terms). This is a consequence of the strong correlation of the eigenvalues $\{\lambda_i\}_{i=1}^N$.

Using Helffer-Sjöstrand formula we find that

$$L_N(f) = \frac{N}{\pi}\int_{\mathbf{R}}\int_{\mathbf{R}} \partial_{\overline{z}} f_{\mathbf{C}}(z)\langle G(z) - \mathbf{E}G(z)\rangle\,\mathrm{d}^2 z. \tag{1.11}$$

The key feature of this integral representation is that, due to the bound (1.9), we can trade in a higher smoothness of the test function f for a poorer control of $\langle G(z) - \mathbf{E}G(z)\rangle$ for small $|\Im z|$. In particular, if we consider a test function f supported on a scale N^{-a}, only the regime $\eta \sim N^{-a}$ gives an order one contribution in Helffer-Sjöstrand.

The analysis of linear statistics goes back to the 90's. The first results proved the Gaussianity of the linear statistics for complex analytic [124, 180] or real analytic [19, 22] test functions. The analiticity of the test functions enables us to have an integral representation of the linear statistics on a contour that is (almost) order one away from the spectrum, hence the resolvent is completely stable, hence a simple application of the moment method is enough for the analysis. For macroscopic test functions (i.e. $a = 0$), after several preliminary results [117, 143, 169], the best result up to date is the proof of the asymptotic Gaussianity of $L_N(f)$ for test functions in the Sobolev space $H^{1+\epsilon}$ [187]. The Gaussianity of the linear statistics of Wigner matrices has been also proven for test function supported down to the optimal mesoscopic scale, that is $N^{-1+\epsilon}$ in the bulk and $N^{-2/3+\epsilon}$ at the edge [110–112, 149].

1.1.4 Microscopic scale

In this section we focus only on Wigner-type matrices, since sample covariance matrices can be analysed exactly in the same way. Indeed, the eigenvalues of XX^*, with X a matrix with i.i.d. entries, are the squares of the eigenvalues of its linearization L, which is an Hermitian matrix defined by

$$L := \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}.$$

On the microscopic scale, which is N^{-1} in the *bulk*, $N^{-2/3}$ at regular *edges*, and $N^{-3/4}$ at *cusps*, the fluctuation of individual eigenvalues becomes relevant. The universality of the local statistics of the eigenvalues has been first conjectured by Wigner in 1955 in the bulk of the spectrum, and then it has been formalized as the Wigner-Dyson-Mehta (WDM) conjecture [146]. More precisely, the WDM universality conjecture states that the local eigenvalue statistics are independent of the details of the model. They depend only on the symmetry class of the matrix (complex Hermitian or real symmetric) and on the local singularity type of the limiting density of states (bulk, edge, cusp).

In contrast with the local law in the mesoscopic regime in Section 1.1.2, which can be interpreted as a law of large numbers (LLN) for the resolvent, the emergence of universal spectral statistics in random matrix theory can be interpreted as the analogue of the universality of Gaussian fluctuations, i.e. CLT, in weakly correlated systems. However, the new universal statistics is not Gaussian.

In order to formulate the universality of local eigenvalue statistics we define the k-point correlation functions $p_k^{(N)}$ implicitly via

$$\int_{\mathbf{R}^k} f(\boldsymbol{x}) p_k^{(N)}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \binom{N}{k}^{-1} \sum_{i_1,\dots i_k=1}^{N} f(\lambda_{i_1}, \dots, \lambda_{i_k}),$$

where the summation is over distinct indices, and f is any smooth compactly supported test function.

We now formulate the WDM conjecture in the complex Hermitian case. In the following we say that $\mathfrak{b}$ is a bulk point if $\rho(\mathfrak{b}) \geq \delta$, for some N-independent $\delta > 0$.

Conjecture (WDM conjecture for the Hermitian symmetry class). *Assume that* $\mathfrak{b}$, $\mathfrak{e}$ *and* $\mathfrak{c}$ *are bulk, edge and cusp points, respectively, of some density* ρ *with parameters* $\gamma_{\mathfrak{e}}, \gamma_{\mathfrak{c}}$ *defined in such a way that*

$$\rho(\mathfrak{e} \pm x) = \gamma_{\mathfrak{e}}^{3/2} x^{1/2}/\pi + \mathcal{O}(x^{1/2}), \quad \rho(\mathfrak{c} + x) = \sqrt{3}\gamma_{\mathfrak{c}}^{4/3} |x|^{1/3}/2\pi + \mathcal{O}(|x|^{1/3}).$$

Then, for any fixed $k \in \mathbf{N}$, the universal correlation functions are given by

$$\frac{1}{\rho(\mathfrak{b})^k} p_k^{(N)}\Big(\mathfrak{b} + \frac{\boldsymbol{x}}{\rho(\mathfrak{b})N}\Big) \approx \det\Big(\frac{\sin \pi(x_i - x_j)}{\pi(x_i - x_j)}\Big)_{i,j\in[k]}, \tag{Bulk}$$

$$\frac{N^{k/3}}{\gamma_{\mathfrak{e}}^k} p_k^{(N)}\Big(\mathfrak{e} + \frac{\boldsymbol{x}}{\gamma_{\mathfrak{e}} N^{2/3}}\Big) \approx \det\Big(K_{\mathrm{Airy}}(x_i, x_j)\Big)_{i,j\in[k]}, \tag{Edge}$$

$$\frac{N^{k/4}}{\gamma_{\mathfrak{c}}^k} p_k^{(N)}\Big(\mathfrak{c} + \frac{\boldsymbol{x}}{\gamma_{\mathfrak{c}} N^{3/4}}\Big) \approx \det\Big(K_{\mathrm{Pearcey}}(x_i, x_j)\Big)_{i,j\in[k]}, \tag{Cusp}$$

where the approximation is meant up to an error of $N^{-c(k)}$ when integrated against smooth compactly supported test functions in $\boldsymbol{x} = (x_1, \dots, x_k)$.

The limiting kernels in the WDM conjecture (edge and bulk) were explicitly computed for the Gaussian Unitary Ensemble (GUE). The kernel in the bulk case is known as the *sine kernel* [147]. The kernel at the edge is given by the *Airy kernel* and it has been first computed in [96]. Finally, in the cusp case the limiting kernel is given by the *Pearcey kernel*, which was computed in [50] for a GUE matrix with diagonal expectation $\mathrm{diag}(1, \dots, 1, -1, \dots, -1)$ using saddle point analysis of an explicit contour integral formula obtained via the *Harish-Chandra-Itzykson-Zuber* integral over the unitary group. We stated the conjecture for complex Hermitian matrices, but the same conjecture holds for real symmetric matrices as well. The limiting kernels in the real case are also known in the bulk and at the edge, but the explicit formula of the kernel at the cusps of real symmetric matrices is not known (due to the lack of *Harish-Chandra-Itzykson-Zuber* integral representation).

The WDM universality conjecture has been an open problem for about fifty years. Universality at the edge of the spectrum of a special class of Wigner matrices has been firstly proven in 1999 using moment method [186]. Only about ten years ago the WDM conjecture was solved in the bulk of the spectrum of Wigner matrices in a series of papers [85, 86, 92, 193]. More recently WDM universality conjecture has been proven also for more general random matrix ensembles both in the bulk and at the edge of the spectrum of the limiting density of states [15, 84]. Close to the cusps the universality of the local statistics has been proven only very recently [57, 83], concluding the third and last remaining case of WDM universality conjecture.

The most powerful technique to prove universality is the so called *three-step-strategy* (see [90] for a pedagogical introduction):

1. Eigenvalue rigidity.
2. Addition of a small Gaussian (GOE/GUE) component via Green function comparison theorem (GFT).
3. Proof of universality for matrices with a small Gaussian component.

The local law in (1.6) is model dependent and it is often quite challenging to prove such a result for very general random matrix ensembles (e.g. Wigner-type matrices, or even more generally, matrices with correlated entries). These local laws have been proven in [15, 83, 84]. In the three-step-strategy the local law is used in (1) to prove the rigidity of the eigenvalues, and in (2) to add a small Gaussian component using a perturbative argument. In the remainder of this chapter we give a few more details about (2), in Section 1.1.4.1, and (3), in Section 1.1.4.2.

1.1.4.1 Green function comparison theorem (GFT)

Given a Wigner-type matrix $H = W + A$, the goal of the second step is to add a small Gaussian component to H without changing much the k-point correlation functions $p_k^{(N)}$. We consider the Ornstein-Uhlenbeck (OU) flow

$$\mathrm{d}H_t = -\frac{1}{2}(H_t - A)\mathrm{d}t + \Sigma^{1/2}[\mathrm{d}B_t], \qquad H_0 = H, \tag{1.12}$$

where B_t is the standard complex Hermitian/real symmetric matrix valued Brownian motion independent of H, and we defined the non-negative operator

$$\Sigma[\cdot] := \mathbf{E}W\mathrm{Tr}[\cdot W].$$

The solution of (1.12) is given by

$$H_t = A + e^{-t/2}W + \int_0^t e^{-(t-s)/2}\Sigma^{1/2}[\mathrm{d}B_s], \tag{1.13}$$

hence one can readily see that the key feature of the flow (1.12) is that the expectation of $\mathbf{E}H_t = A$ and its covariance operator

$$\mathcal{S}_t[\cdot] = \mathbf{E}W_t \cdot W_t$$

are independent of t, where $W_t := H_t - A$. In particular, as a consequence of (1.2)-(1.3), this implies that also the density of states of the eigenvalues of H_t does not change, i.e. $\rho_t \equiv \rho$ for any $t \geq 0$.

Analysing the joint distribution of the resolvents $G_t := (H_t - z)^{-1}$, for different nearby z's with $\Im z \ll \eta_{\mathrm{f}}$, one can see that if the time is not too big then the k-point correlation functions are unchanged at leading order. Using a simple continuity argument (Green function comparison theorem) for G_t one can prove that an upper bound for the time we are allowed to run the OU-flow, without changing the local statistics, is given by

$$t \ll \begin{cases} N^{-1/2} & \text{bulk}, \\ N^{-1/6} & \text{edge}, \\ N^{-1/4} & \text{cusp}. \end{cases} \tag{1.14}$$

We remark that by (1.13) it follows that along the OU-flow we add a Gaussian component proportional to $\sqrt{t}$. Indeed, using fairly easy computations one can construct a Wigner-type matrix $\widetilde{H}_t$ such that

$$H_t \stackrel{\mathrm{d}}{=} \widetilde{H}_t + \sqrt{ct}U, \tag{1.15}$$

with c a constant very close to one, and U being a GUE/GOE matrix independent of $\widetilde{H}_t$. We conclude this section noticing that thanks to the Green function comparison theorem and (1.15) it is enough to prove universality for the matrix H_t, which has a small Gaussian component of size proportional to $\sqrt{t}$.

1.1.4.2 Dyson Brownian motion (DBM)

The final step to prove universality in the complex case can be achieved using explicit computations via the *Harish-Chandra-Itzykson-Zuber* integral representation. This integral representation is available only for complex Hermitian matrices, hence to prove universality for real symmetric matrices one has to rely on the Dyson Brownian motion (DBM) introduced in random matrix theory in [86]. The proof of universality via Dyson Brownian motion works both in the real and in the complex case. The DBM is a system of coupled stochastic differential equations (SDE) introduced by Dyson in [74]. In the remainder of this section we give a sketch of the analysis of the DBM.

Given $T = N^\epsilon \eta_{\mathrm{f}}$, for some small fixed $\epsilon > 0$, consider the matrix flow

$$\mathrm{d}\widehat{H}_t = \frac{\mathrm{d}\widehat{B}_t}{\sqrt{N}}, \qquad \widehat{H}_0 = \widetilde{H}_T, \tag{1.16}$$

for any $t \geq 0$, where $\widehat{B}_t$ is a standard real symmetric or complex Hermitian matrix valued Brownian motion independent of $\widetilde{H}_T$, with $\widetilde{H}_T$ defined in (1.15). The solution of (1.16) is such that

$$\widehat{H}_t \stackrel{\mathrm{d}}{=} \widehat{H}_0 + \sqrt{t}\widehat{U}, \tag{1.17}$$

with $\widehat{U}$ a GUE/GOE matrix independent of $\widehat{H}_0$. In particular, combining (1.15) and (1.17) it follows that

$$\widehat{H}_{cT} \stackrel{\mathrm{d}}{=} H_T, \tag{1.18}$$

with c defined in (1.15).

In order to conclude the proof of universality for matrices with a small Gaussian component we are left with the spectral analysis of the flow (1.16). Using fairly simple computations one can see that the flow (1.16) induces the following DBM-flow on the eigenvalues $\lambda_i(t)$ of H_t:

$$\mathrm{d}\lambda_i(t) = \sqrt{\frac{2}{\beta N}}\mathrm{d}b_i(t) + \frac{1}{N}\sum_{j\neq i}\frac{1}{\lambda_i(t)-\lambda_j(t)}\,\mathrm{d}t, \tag{1.19}$$

with $\{b_i(t)\}_{i=1}^N$ being a family of standard real i.i.d. Brownian motions. Here β is a parameter such that $\beta = 1$ in the real symmetric case, and $\beta = 2$ in the complex Hermitian case. The key idea in the analysis of (1.19) is to use the fact that the GOE/GUE ensembles are a strong attractive equilibrium for the eigenvalue dynamics (1.19), and that the local statistics of GOE/GUE are explicitly computable. In order to exploit this fact we introduce a comparison process

$$\mathrm{d}\mu_i(t) = \sqrt{\frac{2}{\beta N}}\mathrm{d}b_i(t) + \frac{1}{N}\sum_{j\neq i}\frac{1}{\mu_i(t)-\mu_j(t)}\,\mathrm{d}t, \tag{1.20}$$

with $\mu_i(t)$ being the eigenvalues of the evolution of a GOE/GUE matrix along the flow (1.16). Note that the driving Brownian motions in (1.20) are exactly the same as in (1.19). In order to compare the processes $\{\lambda_i(t)\}_{i=1}^N$ and $\{\mu_i(t)\}_{i=1}^N$ directly we take their difference and see that $w_i(t) := \lambda_i(t) - \mu_i(t)$ is a solution of the following parabolic equation

$$\mathrm{d}\mathbf{w} = \mathcal{B}\mathbf{w}\mathrm{d}t, \qquad \mathcal{B}_{ij} = \mathcal{B}_{ij}(t) := \frac{\mathbf{1}(j\neq i)}{(\lambda_i(t)-\lambda_j(t))(\mu_i(t)-\mu_j(t))}. \tag{1.21}$$

Then using heat kernel decay estimates for (1.21) one can show that that $|w_i(t)| \ll \eta_{\mathrm{f}}(\gamma_i)$, with very high probability, after sufficiently long time. This phenomenon is referred to as the *fast relaxation to equilibrium of the DBM*. The time scale for the relaxation to the equilibrium is $\gg N^{-1}$ in the bulk, $\gg N^{-1/3}$ at the edge, and $\gg N^{-1/2}$ at the cusp. These time scales leave quite a big room for the choice of the time T such that verifies (1.14), concluding the third and last step of the *three-step-strategy*.

1.2 Non-Hermitian matrices

Despite several applications [52, 145, 184], non-Hermitian random matrices are much less studied than Hermitian ones. Similarly to Hermitian matrices, it is conjectured that local eigenvalue statistics exhibit a universal behaviour. The analysis of non-Hermitian matrices is much harder for two fundamental reasons: (i) the resolvent is very unstable, (ii) lack of a good analogue of the Dyson Brownian motion. We will comment about these two main difficulties later in this section.

We introduce the non-Hermitian random matrix ensembles considered in this thesis:

Ginibre matrices: Non-Hermitian matrices $X \in \mathbf{C}^{N\times N}$ such that the entries are i.i.d. standard real or complex Gaussian random variables.

i.i.d. matrices Non-Hermitian matrices $X \in \mathbf{C}^{N\times N}$ such that the entries are i.i.d. real or complex centred random variables with variance $\mathbf{E}\,|x_{ab}|^2 = N^{-1}$. In addition, in the complex case the entries x_{ab} are such that $\mathbf{E}\,x_{ab}^2 = 0$.

Analogously to Hermitian matrices, we define the empirical spectral distribution as

$$\mu_N := \frac{1}{N}\sum_{i=1}^N \delta_{\sigma_i}, \tag{1.22}$$

with $\{\sigma_i\}_{i=1}^N$ being the eigenvalues of X; note that σ's are typically complex. The measure μ_N can again be analysed at macroscopic, mesoscopic, and microscopic scales. In the remainder of this section we explain which are the relevant questions in these three regimes.

1.2.1 Global and Mesoscopic scales

In 1984 Girko proved that the empirical eigenvalue density (1.22) of *i.i.d. matrices* converges weakly to the uniform distribution on the unit disk [103] (see also [18, 34, 191, 198]), i.e.

$$\lim_{N\to+\infty}\frac{1}{N}\sum_i f(\sigma_i) = \frac{1}{\pi}\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z, \qquad \mathbf{D} := \{z\in\mathbf{C} : \mathbf{1}(|z|\le 1)\}, \tag{1.23}$$

with f a smooth test function and $\mathrm{d}^2 z := \mathrm{d}\Re z\mathrm{d}\Im z$. Later, the convergence in (1.23) has been generalised to test function supported on a mesoscopic set both in the bulk and in the edge regime, i.e. it has been proven that (1.23) holds replacing f with

$$f_{z_0,a}(z) := N^{2a} g(N^a(z-z_0)) \tag{1.24}$$

with g a smooth and compactly supported test function, for any $a \in [0, 1/2)$ and $|z_0| \leq 1$ (e.g. see [44, 46, 213]). Recently, these results have been generalised to more general random matrix ensembles allowing inhomogeneous variance profile or even some correlation structure [11, 13, 16].

While it is not possible to analyse the non-Hermitian resolvent directly, as a consequence of its instability, one can analyse the linear statistics in (1.23) relying on Girko formula (see [103, 195]):

$$\frac{1}{N}\sum_{i=1}^{N} f_{z_0,a}(\sigma_i) - \frac{1}{\pi}\int_{\mathbf{D}} f_{z_0,a}(z)\,\mathrm{d}^2 z \approx -\frac{1}{2\pi}\int_{\mathbf{C}} \Delta f_{z_0,a}(z)\int_0^T \Im\langle G^z(\mathrm{i}\eta) - m^z(\mathrm{i}\eta)\rangle \mathrm{d}\eta\,\mathrm{d}^2 z, \tag{1.25}$$

with $T = N^{100}$ being a regularisation parameter. The approximation in (1.25) means that the l.h.s. and the r.h.s. are equal modulo a negligible error smaller than T^{-1}. Here G^z denotes the resolvent $G^z = G^z(\mathrm{i}\eta) := (H^z - \mathrm{i}\eta)^{-1}$, with H^z the so called Hermitisation of X defined by

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

The 2×2 block structure of $H^z = (H^z)^*$ induces a spectrum that is symmetric with respect to zero. In addition, note that

$$z \in \mathrm{spec}(X) \iff 0 \in \mathrm{spec}(H^z).$$

The deterministic approximation $m^z(w)$ of $G^z(w)$ in (1.25) can be found as the unique solution of the following scalar cubic equation:

$$-\frac{1}{m^z(w)} = w + m^z(w) - \frac{|z|^2}{w + m^z(w)}, \qquad \Im[m^z(w)]\Im w > 0. \tag{1.26}$$

The limiting eigenvalue distribution ρ^z of H^z is given by

$$\rho^z(x) := \frac{1}{\pi}\lim_{\eta \searrow 0} \Im m^z(x + \mathrm{i}\eta).$$

By a detailed analysis of (1.26) it follows that ρ^z develops a *cusp singularity* as z approaches the unit circle (i.e. for $|z| \approx 1$). This key fact is used in [59] to connect the non-Hermitian edge analysis with the cusp analysis in the Hermitian case.

The good news about Girko's formula (1.25) is that H^z is a Hermitian matrix, so one can expect that the Hermitian theory can be adapted to analyse G^z as well. However, the z-dependence of the resolvent is a difficulty not present in the usual Hermitian case.

Similarly to the Hermitian case, one can conclude further information from mesoscopic spectral analysis of H^z: (i) left and right eigenvectors of X are delocalized [166, Theorem 1.1, Corollary 1.5], (ii) the spectral radius $\rho(X)$ converges to 1 with very high probability with a speed at least $N^{-1/2+\epsilon}$, for some small fixed $\epsilon > 0$ (see [13, Theorem 2.1]).

1.2.2 Linear statistics

Similarly to Hermitian matrices, one can ask what can be said about the fluctuation around the circular law:

$$\left|\frac{1}{N}\sum_{i=1}^{N} f(\sigma_i) - \frac{1}{\pi}\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z\right| \prec \frac{1}{N}, \tag{1.27}$$

with f a sufficiently smooth test function, and $\{\sigma_i\}_{i=1}^N$ being the eigenvalues of the non-Hermitian matrix X. Note that also in the non-Hermitian case, as a consequence of the strong correlation of the eigenvalues, the fluctuation around the expectation are of order N^{-1}, i.e. much smaller than the usual $N^{-1/2}$ for the standard CLT. Define the centred linear statistics

$$L_N(f) := \sum_{i=1}^N f(\sigma_i) - \mathbf{E}\sum_{i=1}^N f(\sigma_i).$$

Note that in the definition of $L_N(f)$ we subtracted the expectation of $\sum_i f(\sigma_i)$ and not the deterministic answer in (1.27) given by the circular law. This is because

$$\mathbf{E}\sum_{i=1}^N f(\sigma_i) = \frac{1}{\pi}\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z + \mathcal{O}\left(\frac{1}{N}\right).$$

For the explicit computation of the sub-leading order correction to the circular law see (2.3) later. In order to analyse $L_N(f)$, one can once again rely on Girko's formula (1.25):

$$L_N(f) \approx -\frac{1}{2\pi}\int_{\mathbf{C}} \Delta f(z) \int_0^T \Im \mathrm{Tr}[G^z(\mathrm{i}\eta) - \mathbf{E}G^z(\mathrm{i}\eta)]\,\mathrm{d}\eta\,\mathrm{d}^2 z, \tag{1.28}$$

with $T = N^{100}$ being a regularisation parameter.

There are two main unrelated difficulties in (1.28): (i) we have to study the resolvent $G^z(\mathrm{i}\eta)$ also for very tiny η's close to zero, (ii) to study the distribution of $L_N(f)$ we have to know the joint distribution of $\langle G^z(\mathrm{i}\eta)\rangle$ for different z's simultaneously. Note that (i) is a fundamental difference compared to Helffer-Sjöstrand formula (1.11) where a smoother f compensates for less information on the resolvent for spectral parameters with small imaginary part. In Girko's formula there is no (known) way to compensate a poorer control for small η's with a smoother test function. In particular, even if one wants to prove a CLT only for macroscopic test functions f it is needed to study $\langle G^z(\mathrm{i}\eta)\rangle$ at microscopic scales.

The analysis of centred linear statistics $L_N(f)$ goes back to 1999 when Forrester [95] proved the Gaussianity of $L_N(f)$ for radial test functions and for complex Ginibre matrices (i.e. the entries of X are standard Gaussian random variables). In [95] Forrester also predicted the exact formula of the variance for complex Ginibre matrices (see (2.2) for $\kappa_4 = 0$ later) and generic test functions, which has been confirmed in [164] by Rider and Virag. In [164] they also interpreted the fluctuation around the circular law as the projection of the Gaussian Free Field (GFF) on the unit disk. This result has been extended to matrices matching the first four moments with the Gaussian ones [126], using the *four moment matching method developed in [195] for non-Hermitian matrices*. For i.i.d. matrices X with generic entry distribution the Gaussianity of $L_N(f)$ has been proven for analytic test functions in the disk of radius 4 [152, 162], and only very recently we proved it for $f \in H^{2+\epsilon}$ in [58, 60]. In these papers we proved that the variance of the limiting Gaussian process depends on the fourth cumulant of the entries of the matrix X (see (2.2) later); the dependence of the variance on the fourth cumulant was previously unknown. We remark that the analysis in [152, 162] for analytic test functions is much easier than the general case $f \in H^{2+\epsilon}$, since for analytic functions in an order one neighborhood of the unit disk one can use an integral representation of f over a contour that is order one away from the spectrum and so it is possible to analyse the non-Hermitian resolvent $(X-z)^{-1}$ directly (outside the spectrum it is completely stable).

1.2.3 Microscopic scale

On the microscopic scale, which is $N^{-1/2}$, the fluctuation of single eigenvalues becomes relevant, hence, similarly to the Hermitian case, one cannot expect that the linear statistics (with $a = 1/2$) converges to a deterministic quantity as in (1.23). Instead, it is expected that the local eigenvalue statistics converge to a universal distribution, which depends only on the matrix being real or complex. We point out that on this scale real matrices exhibit an interesting behaviour close to the real axis: $\sim \sqrt{N}$ eigenvalues accumulate on the real axis (e.g. see [77, 97, 195]).

We now formulate the universality conjecture for non-Hermitian complex matrices. For this purpose we define the k-point correlation functions $p_k^{(N)}$ implicitly as

$$\int_{\mathbf{C}^k} F(\mathbf{w}) p_k^{(N)}(\mathbf{w})\, \mathrm{d}^2 w = \binom{N}{k}^{-1} \mathbf{E} \sum_{i_1,\dots,i_k} F(\sigma_{i_1},\dots,\sigma_{i_k}), \tag{1.29}$$

where $\mathbf{w} := (w_1, \dots, w_k) \in \mathbf{C}^k$, F is a smooth test function, and the summation is over distinct indices.

Conjecture. *Let $p_k^{(N)}$ be the k-point functions defined in* (1.29)*, then, for any fixed base points $\mathbf{z} = (z_1, \dots, z_k) \in \mathbf{C}$, there exists a universal function $p_{\mathbf{z}}^{\mathrm{Gin}}(\mathbf{w})$ such that*

$$p_k^{(N)}\left(\mathbf{z} + \frac{\mathbf{w}}{\sqrt{N}}\right) \approx p_{\mathbf{z}}^{\mathrm{Gin}(\mathbf{R}/\mathbf{C})}(\mathbf{w}), \tag{1.30}$$

where the approximation is meant up to an error $N^{-c(k)}$ when integrated against smooth compactly supported test functions in $\mathbf{w} = (w_1, \dots, w_k) \in \mathbf{C}^k$.

For complex Ginibre matrices the universal function $p_{\mathbf{z}}^{\mathrm{Gin}(\mathbf{C})}$ is determinantal and it has been explicitly computed in [102]. More precisely, $p_{\mathbf{z}}^{\mathrm{Gin}(\mathbf{C})}(\mathbf{w})$ is given by

$$p_{\mathbf{z}}^{\mathrm{Gin}(\mathbf{C})}(w_1, \dots, w_k) = \det\left(K_{z_i,z_j}(w_i, w_j)\right)_{1 \le i,j \le k}, \tag{1.31}$$

where $\mathbf{z} = (z_1, \dots, z_k)$ and the kernel $K_{z_i,z_j}(w_i, w_j)$ is defined by

(i) For $z_1 \neq z_2$, $K_{z_1,z_2}(w_1, w_2) = 0$.

(ii) For $z_1 = z_2$ and $|z_1| > 1$, $K_{z_1,z_2}(w_1, w_2) = 0$.

(iii) For $z_1 = z_2$ and $|z_1| < 1$,

$$K_{z_1,z_2}(w_1, w_2) = \frac{1}{\pi} e^{-\frac{|w_1|^2}{2} - \frac{|w_2|^2}{2} + w_1 \overline{w_2}}.$$

(iv) For $z_1 = z_2$ and $|z_1| = 1$,

$$K_{z_1,z_2}(w_1, w_2) = \frac{1}{2\pi}\left[1 + \mathrm{erf}\left(-\sqrt{2}(z_1 \overline{w_2} + w_1 \overline{z_2})\right)\right] e^{-\frac{|w_1|^2}{2} - \frac{|w_2|^2}{2} + w_1 \overline{w_2}},$$

where

$$\mathrm{erf}(z) := \frac{2}{\sqrt{\pi}} \int_{\gamma_z} e^{-t^2}\, \mathrm{d}t,$$

for any $z \in \mathbf{C}$, with γ_z any contour from 0 to z.

The universal function $p_{\mathbf{z}}^{\mathrm{Gin}(\mathbf{R})}(\mathbf{w})$ is explicitly known also for real matrices [35], but it is much more involved as a consequence of the special role of the real axis.

In [59] we proved the universality conjecture (1.30) at the edge of the spectrum, i.e. for $|z_i| = 1$. This is the non-Hermitian analogue of the Tracy-Widom universality at the edge of the limiting spectrum of Hermitian matrices. This conjecture in the bulk of the spectrum is still an outstanding open problem.

1.2.3.1 Dyson Brownian motion

In order to prove the universality conjecture of local eigenvalue statistics for non-Hermitian random matrices, one would naturally try to use a non-Hermitian analogue of the three-step-strategy developed for Hermitian matrices. This approach so far failed because there is no good analogue of the Dyson Brownian motion (DBM) flow for the non-Hermitian eigenvalues.

Consider the matrix flow

$$\mathrm{d}X_t = \frac{\mathrm{d}B_t}{\sqrt{N}}, \qquad X_0 = X, \tag{1.32}$$

with B_t a matrix with entries being real or complex standard Brownian motions independent of X. From now on we only consider the complex case, the real case is similar but more involved. One can see that if X_t is a solution of (1.32) then its eigenvalues $\{\sigma_i(t)\}_{i=1}^N$ are a solution of the following system of SDE (see [38, Appendix A]):

$$\mathrm{d}\sigma_i(t) = \mathrm{d}M_i(t), \tag{1.33}$$

with $M_i(t)$ being a collection of martingales such that their quadratic covariations are given by

$$\mathrm{d}\langle M_i, \overline{M}_j\rangle = \mathcal{O}_{ij}(t)\frac{\mathrm{d}t}{N}, \qquad \mathcal{O}_{ij}(t) := \langle R_i(t), R_j(t)\rangle\langle L_i(t), L_j(t)\rangle. \tag{1.34}$$

Here $L_i(t)$, $R_i(t)$ are the left and right biorthogonal eigenvectors of X_t.

The analysis of the flow (1.33) is much harder than the standard Hermitian DBM (1.19), because of the correlations of the driving martingales which strongly depend on the eigenvector overlaps $\mathcal{O}_{ij}$. Nothing is known about the distribution of $\mathcal{O}_{ij}$ for general *i.i.d. matrices*; their distribution is known only for complex Ginibre matrices [38] and for overlaps corresponding to real eigenvalues for real Ginibre matrices [99].

To circumvent this problem one can go back to the Hermitisation idea and use the DBM methods for H^z. Then, relying on Girko's formula (1.25) for microscopic test function (i.e. $a = 1/2$):

$$\sum_{i=1}^N f(\sqrt{N}(\sigma_i - z_0)) \approx -\frac{1}{2\pi}\int_{\mathbf{C}} \Delta f(\sqrt{N}(z - z_0)) \int_0^T \sum_{|i|\le N} \frac{\eta}{(\lambda_i^z)^2 + \eta^2}\mathrm{d}\eta\, \mathrm{d}^2 z, \tag{1.35}$$

where $\{\lambda_i^z\}_{-N}^N$, with $\lambda_{-i}^z = -\lambda_i^z$, denote the eigenvalues of H^z, and analyse the eigenvalues λ_i^z close to zero. The flow (1.32) induces the following flow on $\lambda_i^z(t)$ (the eigenvalues of the Hermitisation of $X_t - z$):

$$\mathrm{d}\lambda_i^z(t) = \frac{\mathrm{d}b_i^z(t)}{\sqrt{2N}} + \frac{1}{2N}\sum_{|j|\le N} \frac{1}{\lambda_i^z(t) - \lambda_j^z(t)}\mathrm{d}t, \tag{1.36}$$

where $\{\mathrm{d}b_i^z(t)\}_{i=1}^N$ is an N-dimensional Brownian motion and $b_{-i}^z(t) = -b_i^z(t)$ for $i \in [N]$. The good news of the flow (1.36) is that for a fixed z it behaves as the Hermitian DBM (1.19). However, in order to study the distribution of (1.35) we need to understand the joint distribution of $(\lambda^{z_1}(t), \dots, \lambda^{z_k}(t))$, for a finite collection of parameters $z_1, \dots, z_k \in \mathbf{C}$. For this purpose we compute the quadratic variation of the driving Brownian motions in (1.36):

$$\mathrm{d}\langle b_i^{z_l}, b_j^{z_m}\rangle = 4\Re[\langle \mathbf{u}_i^{z_l}, \mathbf{u}_j^{z_m}\rangle\langle \mathbf{v}_j^{z_m}, \mathbf{v}_i^{z_l}\rangle]\,\mathrm{d}t, \tag{1.37}$$

with $\mathbf{u}_i^{z_l}$, $\mathbf{v}_i^{z_l}$ being left and right singular vectors of $X - z_l$. The families $\{\sqrt{2}\mathbf{u}_i^{z_l}\}_{i=1}^N$, $\{\sqrt{2}\mathbf{v}_i^{z_l}\}_{i=1}^N$ are orthonormal for each fixed $l \in \{1, \dots, k\}$. Hence, in order to compute the joint distribution of $(\lambda^{z_1}(t), \dots, \lambda^{z_k}(t))$ one would need to know the distribution of

$$\langle \mathbf{u}_i^{z_l}, \mathbf{u}_j^{z_m}\rangle\langle \mathbf{v}_j^{z_m}, \mathbf{v}_i^{z_l}\rangle.$$

In [58, 60] we proved that

$$|\langle \mathbf{u}_i^{z_l}, \mathbf{u}_j^{z_m}\rangle| + |\langle \mathbf{v}_j^{z_m}, \mathbf{v}_i^{z_l}\rangle| \le N^{-\epsilon/10},$$

with very high probability, if $|z_l - z_m| \ge N^{-\epsilon}$, for some small $\epsilon > 0$. This shows that the driving Brownian motions $\{b_i^{z_l}\}_{i=1}^N$ are almost independent for different z_l's. This enabled us to prove the asymptotic independence of $(\lambda^{z_1}(t), \dots, \lambda^{z_k}(t))$ for $|z_l - z_m| \ge N^{-\epsilon}$. The case when the z_l's are close is not known, in particular it is expected that

$$|\langle \mathbf{u}_i^{z_l}, \mathbf{u}_j^{z_m}\rangle|^2 \approx \frac{1}{N|z_l - z_m|^2},$$

and a similar result for $|\langle \mathbf{v}_j^{z_m}, \mathbf{v}_i^{z_l}\rangle|^2$. Hence $\{b_i^{z_l}\}_{i=1}^N$ are expected to have a non-trivial correlation when the z_l's are at the distance of the level spacing $N^{-1/2}$ away from each other; this regime would be necessary for bulk universality.

Overview of results 2

This chapter contains a concise summary of the main results of the PhD Thesis. For this summary we selected only the most representative statement from each chapter, several other results and refinements will be presented later.

2.1 CLT for Linear Statistics of Minors of Sample Covariance Matrices (Paper [56])

Let $\widetilde{X}$ be an $M \times N$ matrix with independent identically distributed (i.i.d.) entries, and denote by X the matrix obtained by $\widetilde{X}$ after removing its first column. Fix $\phi = M/N$ to be such that

$$c_1 \leq \phi \leq c_2,$$

for some N-independent constants $c_1, c_2 > 0$. We consider

$$f_N := \mathrm{Tr} f[\widetilde{X}^* \widetilde{X}] - \mathrm{Tr} f[X^* X],$$

with $f \in H^2$, and prove that its fluctuation is much smaller than the one of the centred linear statistics

$$L_N := \mathrm{Tr} f[\widetilde{X}^* \widetilde{X}] - \mathbf{E}\, \mathrm{Tr} f[\widetilde{X}^* \widetilde{X}].$$

Indeed, in [56] (see Chapter 3) we prove that f_N fluctuates on a scale $N^{-1/2}$, whilst L_N fluctuates on a much bigger scale of order one. This is a consequence of the strong correlation of the eigenvalues of $\widetilde{X}^* \widetilde{X}$ and the ones of its minor $\widetilde{X}^* \widetilde{X}$.

Theorem 2.1.1. *For any $f \in H^2$, f_N converges in probability to the constant*

$$\Omega_f := \int_{\gamma_-}^{\gamma_+} f(x) \frac{\sqrt{\phi}}{4\pi^2 x \rho_\phi(x)} \left(1 + \frac{\phi^{1/2} - \phi^{-1/2}}{x} \right) \mathrm{d}x + \frac{f(0)}{2} \mathbf{1}(\phi = 1),$$

where $\phi := M/N \geq 1$, $\gamma_\pm = \phi^{1/2} + \phi^{-1/2} \pm 2$, and ρ_ϕ is the Marchenko-Pastur law in (1.4). *Additionally, we have the following Central Limit Theorem (CLT):*

$$\sqrt{N}(f_N - \Omega_f) \Longrightarrow \Delta_f,$$

where Δ_f is a centred Gaussian random variable of variance

$$V_f := V_{f,1} + (\sigma_4 - 1)V_{f,2} + |\sigma_2|^2,$$

where $\sigma_2 := \sqrt{MN}\mathbf{E}\widetilde{X}_{i\mu}^2$, $\sigma_4 := MN\mathbf{E}|\widetilde{X}_{i\mu}|^4$,

$$V_{f,1} = \int_{\gamma_-}^{\gamma_+} f'(x)^2 x\rho_\phi(x)\phi^{-1/2}\,\mathrm{d}x - \left(\int_{\gamma_-}^{\gamma_+} f'(x)x\rho_\phi(x)\phi^{-1/2}\,\mathrm{d}x\right)^2,$$

$$V_{f,2} = \left(\int_{\gamma_-}^{\gamma_+} f'(x)^2 x\rho_\phi(x)\phi^{-1/2}\,\mathrm{d}x\right)^2,$$

and V_{σ_2} is explicit (see (3.119) *later) and such that $V_{\sigma_2} = V_{f,1}$ for $|\sigma_2| = 1$.*

2.2 Cusp Universality for Wigner-type Matrices (Paper [57])

We consider *Wigner-type* matrices H. Denote by ρ the density profile of the eigenvalues of H. The classification theorem (see below (1.3)) shows that the eigenvalue density ρ may vanish only as a square root at *regular edges* or as a cubic root at *cusps*, and no other singularities may occur. The main result of [57] (see Chapter 4) is the proof of cusp universality for Wigner-type matrices. This proves the third and last remaining case of *Wigner-Dyson-Mehta* universality conjecture.

Theorem 2.2.1. *Assume that ρ has a cusp point at $\mathfrak{c}$, then the local k-point correlation function $p_k^{(N)}$ at $\mathfrak{c}$ is universal, i.e. there exists a universal determinantal k-point correlation function of the form*

$$p_k^{\text{Pearcey}}(x_1, \ldots, x_k) = \det\left(K_{\text{Pearcey}}(x_i, x_j)\right)_{1\le i,j,\le k}$$

with $x_1, \ldots, x_k \in \mathbf{R}$, such that

$$N^{k/4} p_k^{(N)}\left(\mathfrak{c} + \frac{\mathbf{x}}{N^{3/4}}\right) \rightharpoonup p_k^{\text{Pearcey}}(\mathbf{x}), \qquad x \in \mathbb{R}^k, \tag{2.1}$$

irrespective of any details of the distribution of H, except symmetry.

The proof of cusp universality in [57, 83] follows the *three-step-strategy*. The local law part which proves the rigidity of the eigenvalues (Step 1) and the Green function comparison theorem (Step 2) have been proven in [83]. The main novelty in [57] is the analysis of the Dyson Brownian motion (DBM) to prove universality of the local statistics of matrices with a small Gaussian component (Step 3). We remark that Step 3 for the complex case in [83] has been proven relying on a saddle point analysis in the *Harish-Chandra-Itzykson-Zuber* integral representation of the k-point function. Such an explicit formula is not known for real symmetric matrices, hence in [57] we rely on the analysis of the DBM that works for both complex Hermitian and real symmetric Wigner-type matrices.

In [57] we extended the DBM analysis to the cusp regime. The main difficulties lied in the rigidity analysis of the DBM, and in the careful analysis of the shape of the highly unstable eigenvalue density along the DBM. The main novelty is a dynamical proof of eigenvalue rigidity near the cusp along the DBM using a novel PDE-based method which relies on the maximum principle (previous results applied only to the bulk [114] and edge [1] regimes). The cusp is much harder, in fact it represents an entire one parameter family of universality classes (see Chapter 4 for more details).

2.3 Edge Universality for non-Hermitian Matrices (Paper [59])

We consider $N \times N$ matrices X with i.i.d. entries. Similarly to the Hermitian case, it is conjectured that local spectral statistics of non-Hermitian matrices X are universal. In [59] (see Chapter 5) we proved this conjecture at the *edge* of the spectrum; this is the non-Hermitian analogue of Tracy-Widom universality.

Theorem 2.3.1. *Fix spectral parameters* $\mathbf{z} := (z_1, \dots, z_k) \in \mathbf{C}^k$ *such that* $|z_i| = 1$. *Then the* k*-point correlation function* $p_k^{(N)}$ *is universal, i.e. it holds*

$$p_k^{(N)}\left(\mathbf{z} + \frac{\mathbf{w}}{N^{1/2}}\right) \rightharpoonup p_{\mathbf{z}}^{\mathrm{Gin}(\mathbf{R}/\mathbf{C})}(\mathbf{w}), \qquad \mathbf{w} \in \mathbf{C}^k,$$

where $p_{\mathbf{z}}^{\mathrm{Gin}}$ *is explicitly computed for Ginibre matrices (e.g. see* (1.31) *for the complex case).*

The only previous result is by Tao and Vu [195]. They prove *bulk* and *edge* universality for X with entries matching the first four moments with the corresponding Gaussian ones. Matching the first four moment with Gaussian random variables allows for a purely perturbative argument; this is not possible when the entries of X match only the first two moments with the ones of standard Gaussians.

As explained in Chapter 1, non-Hermitian matrices are much harder to analyse than Hermitian matrices since successful techniques in Hermitian theory, e.g. resolvents and DBM, do not have useful non-Hermitian counterparts. We circumvent this problem by using Girko's formula (1.25). The main problem in (1.25) is that we need to control $G^z(\mathrm{i}\eta)$ even for very tiny η's. We follow the Ornstein-Uhlenbeck (OU) flow for a very long time (up to infinity) to interpolate between the distribution of the matrix X (at time $t = 0$) and Ginibre (at time $t = +\infty$).

The main novelties are: (i) lower tail estimate of the smallest singular value of $X - z$ (See Section 2.5), (ii) precise analysis of the resolvent G^z along the OU flow for long time exploiting the extra smallness close to the edge of the spectrum of X. The extra smallness for $|z| \approx 1$ is a consequence of the fact that the density of states ρ^z of H^z develops a cusp singularity in zero and so that we have a stronger local law for $G^z(\mathrm{i}\eta)$ (see [13]).

2.4 CLT for Linear Eigenvalue Statistics of non-Hermitian Matrices (Papers [58, 60])

We consider real or complex $N \times N$ *i.i.d. matrices* X. For sufficiently regular test functions f on $\mathbf{C}$, it is expected that the centred linear statistics of the non-Hermitian eigenvalues

$$L_N(f) = \sum_i f(\sigma_i) - \mathbf{E} \sum_i f(\sigma_i),$$

have Gaussian fluctuations of order one. This may be viewed as an anomalous version of the CLT, since the usual $N^{1/2}$ scaling factor is missing as a consequence of the strong correlation of the eigenvalues of X. In [58, 60] (see Chapter 6 and Chapter 7) we proved a CLT for linear statistics $L_N(f)$ for test functions in the Sobolev space $f \in H^{2+\epsilon}$. In the remainder of this section we state the result in the complex case, the real case is similar but more complicated (see Theorem 7.2.1 later).

Theorem 2.4.1. *Denote by* $\kappa_4 := \mathbf{E}|\sqrt{N}X_{ij}|^4 - 2$ *the fourth cumulant of the entries of* X. *Then* $L_N(f)$ *converges*

$$L_N(f) \Longrightarrow L(f)$$

to a Gaussian random variable $L(f)$ *with expectation* $\mathbf{E}L(f) = 0$ *and variance*

$$\mathbf{E}|L(f)|^2 := \frac{1}{4\pi}\|\nabla f\|_{L^2(\mathbf{D})} + \frac{1}{2}\sum_{k\in\mathbf{Z}} |\widehat{f}(k)|^2 + \kappa_4 \left| \frac{1}{\pi}\int_{\mathbf{D}} f(z)\mathrm{d}^2 z - \frac{1}{2\pi}\int_0^{2\pi} f(e^{\mathrm{i}\theta})\mathrm{d}\theta \right|^2, \tag{2.2}$$

where $\mathbf{D}$ *denotes the unit disk, and* $\widehat{f}$ *denotes the Fourier transform of the restriction of* f *to* $\partial\mathbf{D}$. *We also find the subleading order corrections to the circular law:*

$$\mathbf{E}\sum_{i=1}^N f(\sigma_i) = \frac{N}{\pi}\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z - \frac{\kappa_4}{\pi}\int_{\mathbf{D}} f(z)(2|z|^2 - 1)\,\mathrm{d}^2 z \tag{2.3}$$

We remark that the dependence on the fourth cumulant κ_4 in (2.2)-(2.3) was previously unknown.

Previous results considered either only Ginibre matrices (with radial test functions [95] or generic test functions [164]); or only i.i.d. matrices with analytic test functions f [152, 162]; or only X with entries matching the first four moments with the corresponding Gaussian ones [126, 195]. Our result needs none of these restrictions.

Our proof relies on two main novel ingredients: (i) local law for products of resolvents at different spectral parameters $G^{z_1}(w_1)AG^{z_2}(w_2)B$, with A, B deterministic matrices (previous local laws involved only single G); (ii) coupling of several dependent DBMs (previously only indepedent DBMs have been analysed).

2.5 Optimal Lower Bound on the Least Singular Value of the Shifted Ginibre Ensemble (Paper [61])

Classical smoothing inequalities [168] prove a lower tail bound for the lowest singular value $\lambda_1(A)$ of $A = A_0 + X$, with A_0 deterministic, and X being a Ginibre matrix:

$$\mathbf{P}\left(\lambda_1(A) \le xN^{-1}\right) \lesssim \begin{cases} x & \text{if } X \sim \mathrm{Gin}(\mathbf{C}), \\ \sqrt{x} & \text{if } X \sim \mathrm{Gin}(\mathbf{R}). \end{cases}$$

We proved [61] (see Chapter 8) the optimal bound for the lowest singular value of $X - z$ improving the bound [168] for the particular shift $A_0 = zI$ in the regime $|z| \approx 1$. This improvement is essential for our proof of non-Hermitian edge universality (Section 2.3).

Theorem 2.5.1. *Let* $|z| \le 1 + CN^{-1/2}$*, for some constant* $C > 0$. *Then, for any* $x > 0$, *it holds*

$$\mathbf{P}(\lambda_1(X - z) \le xc(N, z)) \lesssim \begin{cases} x & \text{if } X \sim \mathrm{Gin}(\mathbf{C}), \\ x + e^{-\frac{1}{2}N(\Im z)^2}\sqrt{x} & \text{if } X \sim \mathrm{Gin}(\mathbf{R}), \end{cases} \tag{2.4}$$

where

$$c(N, z) := \min\left\{ \frac{1}{N^{3/4}}, \frac{1}{N\sqrt{|1 - |z|^2|}} \right\}.$$

To prove (2.4) we relied on supersymmetric (SUSY) techniques. Supersymmetric formalism has been a very useful computational tool in physics, even beyond Random Matrix Theory, although most of these physics arguments lack mathematical rigour. However, we managed to use rigorous SUSY analysis for a key problem that was partly motivated by non-Hermitian edge universality (Section 2.3).

Using SUSY, we gave an integral representation of $\mathrm{Tr}[(X-z)(X-z)^*-\mathrm{i}\eta]^{-1}$, which for small η's yields (2.4). Then, by the superbosonization formula [142], we reduced the representation to two contour integrals in the complex case, and to three contour integrals in the real case (see (2.5) for the representation in the complex case).

$$\begin{aligned} \mathbf{E}\mathrm{Tr}[(X-z)(X-z)^*-\mathrm{i}\eta]^{-1} &= \frac{N^2}{2\pi \mathrm{i}}\int_0^{\mathrm{i}\infty}\mathrm{d}x\oint \mathrm{d}y\, e^{-Nf(x)+Nf(y)} y\cdot G(x,y), \\ G(x,y) &:= \frac{1}{xy}-\frac{1}{(1+x)(1+y)}\left[1+\frac{|z|^2}{1+x}+\frac{|z|^2}{1+y}\right], \\ f(x) &:= \log(1+x)-\log x-\frac{|z|^2}{1+x}-\mathrm{i}\eta x, \end{aligned} \tag{2.5}$$

where the x-integration is over $(0,\mathrm{i}\infty)$, and the y-integration is over a circle of radius N^{-1} around the origin.

The main difficulty is the rigorous analysis of the contour integrals, which are not accessible via saddle point analysis. The answer comes from very careful estimates along the whole integration regime, in contrast with standard SUSY analysis where the answer comes only from the saddle point.

Fluctuations for linear eigenvalue statistics of sample covariance matrices 3

We prove a central limit theorem for the difference of linear eigenvalue statistics of a sample covariance matrix $\widetilde{W}$ and its minor W. We find that the fluctuation of this difference is much smaller than those of the individual linear statistics, as a consequence of the strong correlation between the eigenvalues of $\widetilde{W}$ and W. Our result identifies the fluctuation of the spatial derivative of the approximate Gaussian field in the recent paper by Dumitru and Paquette [73]. Unlike in a similar result for Wigner matrices, for sample covariance matrices the fluctuation may entirely vanish.

Published as G. Cipolloni and L. Erdős, *Fluctuations for differences of linear eigenvalue statistics for sample covariance matrices*, Random Matrices: Theory and Applications **9**, 2050006 (2020).

3.1 Introduction

We consider sample covariance matrices of the form $\widetilde{W} = \widetilde{X}^*\widetilde{X}$, where the entries of the $M \times N$ matrix $\widetilde{X}$ are i.i.d. random variables with mean zero and variance $\frac{1}{\sqrt{MN}}$. In the Gaussian case this ensemble was introduced by Wishart [212]. Besides Wigner matrices, this is the oldest and the most studied family of random matrices.

Let $\lambda_1, \dots, \lambda_N$ be the eigenvalues of $\widetilde{W} = \widetilde{X}^*\widetilde{X}$, then the empirical distribution $\frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$ converges in probability to the Marchenko-Pastur distribution [144]. This asymptotics can be refined by examining the centered linear statistics

$$\mathrm{Tr} f(\widetilde{W}) - \mathbf{E}\mathrm{Tr}\, f(\widetilde{W}) = \sum_{i=1}^N \left[f(\lambda_i) - \mathbf{E} f(\lambda_i) \right], \tag{3.1}$$

with a sufficiently smooth function f, which has been shown to have Gaussian fluctuation (see e.g. [22], [119], [169]). Notice that (3.1) does not carry the usual $\frac{1}{\sqrt{N}}$ normalization of the conventional central limit theorem. In particular this result indicates a very strong correlation between eigenvalues. Apart from understanding an interesting mathematical phenomenon, the asymptotic properties of centered linear statistics for sample covariance matrices also have potential applications [159].

All the previously cited works on the centered linear statistics of a sample covariance matrix $\widetilde{W}$ concern the study of a single random matrix. The recent paper of Dumitru and Paquette [73] considers the joint eigenvalue fluctuations of a sample covariance matrix and its minors, by picking submatrices whose dimensions differ macroscopically. They show that their centered linear eigenvalue statistics converge to spatial averages of a two dimensional Gaussian free field. Similar results for Wigner matrices have been achieved earlier in [36].

In the current work we study this phenomenon for submatrices whose dimensions differ only by one. This requires a detailed analysis on the local spectral scale while [73] concerns only the global scale. In particular, we prove a central limit theorem (CLT) for the difference of linear eigenvalue statistics of a sample covariance matrix $\widetilde{W} = \widetilde{X}^*\widetilde{X}$ and its minor $W = X^*X$, obtained by deleting the first row and column. This difference fluctuates on a scale $N^{-\frac{1}{2}}$, which is much smaller than the order one fluctuations scale of the individual linear statistics, demonstrating a strong correlation between the eigenvalues of $\widetilde{W}$ and its minor W. The statistical interpretation of our result is that changing the sample size by one in a statistical data has very little influence on the fluctuations of the linear eigenvalue statistics. Motivated by Gorin and Zhang [104], another interpretation is that we prove a CLT for the *spatial derivative* of the approximate Gaussian field in [73].

This result extends a CLT, proved in [89] for Wigner matrices, to sample covariance random matrices, with the difference that in this latter case it is also possible not to have random fluctuations at all, see Remark 3.2.4 in Section 2.

In the proof of the CLT for sample covariance matrices there are two main differences compared to the proof given in [89] for the Wigner case. Firstly, we have to handle the singularity of the Marchenko-Pastur law at zero, which also gives an additional contribution to the leading order term of (3.8). Secondly, the entries of the matrix $\widetilde{W} = \widetilde{X}^*\widetilde{X}$ are not independent and the analogy occurs on the level of $\widetilde{X}$. Besides linearizing the problem and using recent local laws for Gram matrices [12, 30], we need to approximate sums of the form $\sum_{ij} G_{ij}G'_{ji}$ and $\sum_{ij} G_{ij}G'_{ij}$ where G and G' are the resolvents of XX^* at two different spectral parameter. While the first sum is tracial, the second one is not and thus cannot be directly analyzed by existing local laws: we need to derive a novel self-consistent equation for it.

Notation

We introduce some notation we use throughout the paper. For positive quantities f, g, we write $f \lesssim g$ if $f \le Cg$, for some $C > 0$ which depends only on the parameter ϕ defined in (3.4). Similarly, we define $f \gtrsim g$. For any $\alpha, \beta > 0$, with $\alpha \asymp \beta$ we denote that there exists two ϕ independent constants $r_*, r^* > 0$ such that $r_*\beta \le \alpha \le r^*\beta$.

Acknowledgments

The authors are grateful to Dominik Schröder for valuable discussions. We also thank the referees for many useful comments and for pointing out a computation mistake in the first version of the paper.

3.2 Main Results

All along the paper we will refer to the $N \times N$ matrix with $\widetilde{W} = \widetilde{X}^* \widetilde{X}$ and to the $(N-1) \times (N-1)$ matrix obtained after removing its first row and column with $W = X^* X$, where X is the matrix obtained by $\widetilde{X}$ after removing its first column. It may look unconventional, but we chose to put the tilde on the original matrix $\widetilde{W}$ and no tilde on the minor W in order to simplify formulas.

Remark 3.2.1. *We follow the convention that Latin letters $i \in \{1, \ldots, M\}$ denote the rows of the matrix $\widetilde{X}$ and Greek letters $\mu \in \{1, \ldots, N\}$ its columns.*

Let $\widetilde{X}$ be an $M \times N$ matrix whose entries $\widetilde{X}_{i\mu}$ are i.i.d. complex valued random variables satisfying:

$$\mathbf{E}\widetilde{X}_{i\mu} = 0, \ \mathbf{E}|\widetilde{X}_{i\mu}|^2 = \frac{1}{\sqrt{MN}}, \quad 1 \le i \le M, \, 1 \le \mu \le N. \tag{3.2}$$

Furthermore, for any $p \in \mathbf{N}$ there exists a constant $C_p > 0$ such that

$$\mathbf{E}\left|(NM)^{\frac{1}{4}} \widetilde{X}_{i\mu}\right|^p \le C_p, \quad 1 \le i \le M, \, 1 \le \mu \le N. \tag{3.3}$$

We assume that M and N are comparable, i.e. there exist N-independent constants $c_1, c_2 > 0$ such that

$$c_1 \le \phi := \frac{M}{N} \le c_2. \tag{3.4}$$

For fixed ϕ and large N the empirical distribution of the eigenvalues of the $N \times N$ matrix $\widetilde{W} = \widetilde{X}^* \widetilde{X}$ is given by the Marchenko-Pastur law [144]:

$$\rho_\phi(dx) = \rho_\phi(x)dx + (1-\phi)_+ \delta(dx), \text{ with } \rho_\phi(x) := \frac{\sqrt{\phi}}{2\pi} \sqrt{\frac{[(x-\gamma_-)(\gamma_+ - x)]_+}{x^2}}, \tag{3.5}$$

where we defined

$$\gamma_\pm := \sqrt{\phi} + \frac{1}{\sqrt{\phi}} \pm 2$$

to be the edges of the limiting spectrum. The Stieltjes transform of $\rho_\phi(dx)$ is

$$m_\phi(z) := \int_{\mathbf{R}} \frac{\rho_\phi(dx)}{x - z} = \frac{\phi^{1/2} - \phi^{-1/2} - z + i\sqrt{(z-\gamma_-)(\gamma_+ - z)}}{2\phi^{-1/2} z}, \tag{3.6}$$

where the square root is chosen so that m_ϕ is holomorphic in the complex upper half plane $\mathbf{H}$ and satisfies $m_\phi(z) \to 0$ as $z \to \infty$. The function $m_\phi = m_\phi(z)$ may also be characterized as the unique solution of the equation

$$m_\phi + \frac{1}{z + z\phi^{-1/2} m_\phi - (\phi^{1/2} - \phi^{-1/2})} = 0 \tag{3.7}$$

satisfying $\Im m_\phi(z) > 0$ for $\Im z > 0$. Our main result is the following:

Theorem 3.2.2. *Let $d_* > 0$ and $\widetilde{W} = \widetilde{X}^* \widetilde{X}$, with $\widetilde{X}$ an $M \times N$ matrix whose i.i.d. entries satisfy* (3.2) *and* (3.3). *Furthermore we assume* (3.4) *and that either $\phi = 1$ or $|\phi - 1| \ge d_*$. Let $\sigma_2 := \sqrt{MN}\mathbf{E}\widetilde{X}_{i\mu}^2$ and $\sigma_4 := MN\mathbf{E}|\widetilde{X}_{i\mu}|^4$ and assume that σ_2 and σ_4 are N-independent .*

Moreover, let $f \in H^2([\gamma_- - 3, \gamma_+ + 3])$ *be some real valued function in the* H^2*-Sobolev space. Then the random variable*

$$f_N := Trf(\widetilde{W}) - Trf(W) \tag{3.8}$$

converges in probability to the constant

$$\Omega_f := \int_{\gamma_-}^{\gamma_+} f(x) \frac{\sqrt{\phi}}{4\pi^2 x \rho_\phi(x)} \left(1 + \frac{\sqrt{\phi} - \frac{1}{\sqrt{\phi}}}{x} \right) dx \tag{3.9}$$

for $|\phi - 1| \geq d_*$*, and to*

$$\Omega_f := \int_0^4 \frac{f(x)}{4\pi^2 x \rho_1(x)} \, dx + \frac{f(0)}{2}.$$

for $\phi = 1$*. More precisely, for any fixed* $\epsilon > 0$,

$$\mathbf{E} f_N = \Omega_f + \mathcal{O}\left(N^{-\frac{2}{3}+\epsilon}\right)$$

and f_N *fluctuates on a scale* $N^{-\frac{1}{2}}$*, i.e.*

$$\mathbf{E}\left(\sqrt{N}(f_N - \Omega_f)\right)^2 = V_f + \mathcal{O}\left(N^{-\frac{1}{6}+\epsilon}\right).$$

The limit variance V_f *can be computed explicitly:*

$$V_f := V_{f,1} + (\sigma_4 - 1)V_{f,2} + |\sigma_2|^2 V_{\sigma_2}, \tag{3.10}$$

with

$$V_{f,1} = \int_{\gamma_-}^{\gamma_+} f'(x)^2 x \rho_\phi(x) \phi^{-\frac{1}{2}} \, dx - \left(\int_{\gamma_-}^{\gamma_+} f'(x) x \rho_\phi(x) \phi^{-\frac{1}{2}} \, dx \right)^2,$$

$$V_{f,2} = \left(\int_{\gamma_-}^{\gamma_+} f'(x) x \rho_\phi(x) \phi^{-\frac{1}{2}} \, dx \right)^2,$$

where $\rho_\phi(x)$ *is the density of the Marchenko-Pastur law* (3.5)*, and* V_{σ_2} *defined as in* (3.119) *if* $|\sigma_2| < 1$ *and* $V_{\sigma_2} := V_{f,1}$ *if* $|\sigma_2| = 1$.

Furthermore,

$$\sqrt{N}(f_N - \Omega_f) \Rightarrow \Delta_f,$$

where Δ_f *is a centered Gaussian random variable of variance* V_f *and " $\Rightarrow$ " denotes the convergence in distribution. Finally, any fixed moment converges at least at a rate* $\mathcal{O}\left(N^{-\frac{1}{6}+\epsilon}\right)$ *to the corresponding Gaussian moments.*

Remark 3.2.3. *The non-negativity of* $V_{f,1}$ *follows by applying Schwarz inequality using that* $\int_{\gamma_-}^{\gamma_+} x \rho_\phi(x) \phi^{-\frac{1}{2}} \, dx = 1$.

Remark 3.2.4. *One can easily check that the variance* V_f *is zero if and only if* $\sigma_2 = 0$, $\sigma_4 = 1$ *and* $f'(x) \equiv 1$*. This is the case, for example, when the entries of* $\widetilde{X}$ *are i.i.d complex Bernoulli random variables, i.e. the distribution of each* $\widetilde{X}_{i\mu}$ *is* $(MN)^{-\frac{1}{4}} e^{iU}$*, with* U *a uniform random* *variable in* $[0, 2\pi]$. *In particular, since the entries of* $\widetilde{X}$ *have modulus* $(MN)^{-\frac{1}{4}}$, *the difference of the traces of* $\widetilde{W}$ *and* W *is deterministic:*

$$Trf(\widetilde{W}) - Trf(W) = Tr\widetilde{W} - TrW = \boldsymbol{x}^*\boldsymbol{x} = \sqrt{\phi},$$

where $\boldsymbol{x}$ *is the first column of* $\widetilde{X}$. *The possibility of* $V_f = 0$ *is a fundamental difference compared to the Wigner case in [89] where the analogous quantity always had a non trivial fluctuation.*

Remark 3.2.5. *We stated our result in Theorem 3.2.2 for the matrix* $\widetilde{X}^*\widetilde{X}$, *but it obviously holds for* $\widetilde{X}\widetilde{X}^*$ *as well. Indeed all computations and results remain valid after the swapping:* $\widetilde{X} \leftrightarrow \widetilde{X}^*$, $M \leftrightarrow N$ *and* $\phi \leftrightarrow \phi^{-1}$. *The empirical distribution of the eigenvalues of* $\widetilde{X}\widetilde{X}^*$ *is asymptotically* $\rho_{\phi^{-1}}(dx)$, *whose Stieltjes transform is*

$$m_{\phi^{-1}}(z) = \frac{1}{\phi}\left(m_\phi(z) + \frac{1-\phi}{z}\right). \tag{3.11}$$

Remark 3.2.6. *Notice that in the statement of Theorem 3.2.2 we assumed that* $\widetilde{X}$ *is either a square matrix,* $\phi = 1$, *or a proper rectangular matrix,* $|\phi - 1| > d_*$. *The reason is that to prove Theorem 3.2.2 we use optimal local laws for all* $z \in \mathbf{H}$ *which are available in these cases only (see [12]). If* ϕ *is close to one, our proof still yields Theorem 3.2.2 assuming that the function* $f \in H^2$ *is supported away from zero.*

3.3 Preliminaries

Our main result pertains to the matrix $\widetilde{X}^*\widetilde{X}$, but in the proof we will also need the matrix $\widetilde{X}\widetilde{X}^*$, so for each $z \in \mathbf{H}$ we define both resolvents

$$\widetilde{R}(z) := (\widetilde{X}^*\widetilde{X} - z)^{-1}, \quad \widetilde{G}(z) := (\widetilde{X}\widetilde{X}^* - z)^{-1}. \tag{3.12}$$

Next, we define the $M \times (N-1)$ matrix X as the matrix $\widetilde{X}$ after removing its first column, which we denote by $\boldsymbol{x}$, i.e. $\widetilde{X} = [\boldsymbol{x}|X]$. Moreover, for $\mu, \nu \notin \{1\}$, we define the resolvent entries

$$R_{\mu\nu}(z) := (X^*X - z)^{-1}_{\mu\nu}, \quad G^{[T]}_{ij}(z) := (XX^* - z)^{-1}_{ij}.$$

Remark 3.3.1. *In the following sections, without loss of generality, we will always assume that* $\phi \geq 1$, *i.e.* $M \geq N$. *Indeed, if* $\phi \leq 1$ *then the proof proceeds exactly in the same way having in mind that* $m_{\phi^{-1}}$ *and* m_ϕ *are related by* (3.11).

Since $\phi \geq 1$ and the spectrum of $\widetilde{X}\widetilde{X}^*$ is equal to the spectrum of $\widetilde{X}^*\widetilde{X}$ plus $M - N$ zero eigenvalues, we have

$$\phi\frac{1}{M}\mathrm{Tr}\widetilde{G} = \frac{1}{N}\mathrm{Tr}\widetilde{R} + \frac{1-\phi}{z} \tag{3.13}$$

and that

$$\mathrm{Tr}R - \mathrm{Tr}G = \frac{M-(N-1)}{z}. \tag{3.14}$$

Furthermore, setting $\eta = \Im z > 0$, we have the *Ward identity*

$$\sum_{j=1}^{M} |G_{ij}(z)|^2 = \frac{1}{\eta}\Im G_{ii}(z). \tag{3.15}$$

Finally, we record some properties of the Stieltjes transform defined in (3.6) in the following lemma, which will be proved in Appendix A.

Lemma 3.3.2. *There exist positive constants* $c, \tilde{c}, \hat{c}$ *such that for any* $\phi \geq 1$ *and for each* $z = x + i\eta \in \mathbf{H}$ *such that* $|z - \sqrt{\phi}| \leq 10$ *we have the following bounds*

$$c \leq \left| \frac{z}{\sqrt{\phi}} m_\phi(z)^2 \right| \leq 1 - \tilde{c}\eta, \tag{3.16}$$

$$|m_\phi(z)'| \leq \frac{\hat{c}\sqrt{\phi}}{|z|\sqrt{\kappa_x + \eta}}, \tag{3.17}$$

$$\left|1 - z\phi^{-\frac{1}{2}} m_\phi(z)^2\right| \asymp \frac{\phi^{\frac{1}{4}}}{|z|^{\frac{1}{2}}} \sqrt{\kappa_x + \eta},, \tag{3.18}$$

where $\kappa_x := \min\{|\gamma_+ - x|, |\gamma_- - x|\}$.

In Lemma 3.3.2 we explicitly wrote the ϕ-dependence in the bounds since they hold uniformly in ϕ. But all along the proof of Theorem 3.2.2 we will omit the explicit dependence on ϕ, since we work under the assumption $c_1 \leq \phi \leq c_2$ (see (3.4)).

3.4 Mean and variance computation

In this section we prove Theorem 3.2.2 in the sense of mean and variance. We recall that with $\boldsymbol{x}$ we denote the first column of $\widetilde{X}$. To study $f_N = \mathrm{Tr} f(\widetilde{W}) - \mathrm{Tr} f(W)$, with $\widetilde{W} = \widetilde{X}^* \widetilde{X}$ and $W = X^* X$, we consider the quantity

$$\Delta_N(z) := \mathrm{Tr}\widetilde{R}(z) - \mathrm{Tr}R(z), \quad z \in \mathbf{H}. \tag{3.19}$$

Clearly $\widetilde{X}\widetilde{X}^*$ is a rank-one perturbation of the matrix XX^*, hence to compute $\widetilde{G}(z)$ we use the following lemma whose proof is a direct calculation.

Lemma 3.4.1. *Let* A *be an* $M \times M$ *matrix with* $\Im A < 0$ *and* $h \in \mathbf{C}^M$ *a column vector, then*

$$\frac{1}{A + hh^*} = \frac{1}{A} - \frac{1}{1 + \left\langle h, \frac{1}{A} h \right\rangle} \cdot \frac{1}{A} hh^* \frac{1}{A}.$$

We now find an explicit formula for $\Delta_N(z)$. Using (3.13), (3.14) and (3.19) we get

$$\Delta_N(z) = \mathrm{Tr}\widetilde{G}(z) - \mathrm{Tr}R(z) - \frac{N(1-\phi)}{z} = \mathrm{Tr}\frac{1}{XX^* + \boldsymbol{x}\boldsymbol{x}^* - z} - \mathrm{Tr}G(z) - \frac{1}{z}.$$

Using Lemma 3.4.1 for the first term in the right-hand side, we conclude that

$$\Delta_N(z) = -\frac{\langle \boldsymbol{x}, G^2(z)\boldsymbol{x} \rangle}{1 + \langle \boldsymbol{x}, G(z)\boldsymbol{x} \rangle} - \frac{1}{z}. \tag{3.20}$$

We introduce a commonly used notion of high probability bound.

Definition 3.4.2. *If*

$$X = \left(X^{(N)}(u)|N \in \mathbf{N},\, u \in U^{(N)}\right) \quad \textit{and} \quad Y = \left(Y^{(N)}(u)|N \in \mathbf{N},\, u \in U^{(N)}\right)$$

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

$$\sup_{u \in U^{(N)}} \mathbf{P}\left(X^{(N)}(u) > N^{\epsilon} Y^{(N)}(u)\right) \le N^{-D}$$

for large enough $N \ge N_0(\epsilon, D)$. In this case we use the notation $X \prec Y$. Moreover, if we have $|X| \prec Y$, we also write $X = \mathcal{O}_{\prec}(Y)$.

We will say that a sequence of events $A = A^{(N)}$ holds with overwhelming probability if $\mathbf{P}\left(A^{(N)}\right) \ge 1 - N^{-D}$ for any $D > 0$ and $N \ge N_0(D)$. In particular, under the conditions (3.2) and (3.3), we have $X_{i\mu} \prec (MN)^{\frac{1}{4}}$ uniformly in i, μ and that $\max_k \lambda_k \le \gamma_+ + 1$, $\min_k \lambda_k \ge \max\{0, \gamma_- - 1\}$ with overwhelming probability (see Theorem 2.10, Lemma 4.11 in [30]).

Let $\chi : \mathbf{R} \to \mathbf{R}$ be a smooth cut-off function which is constant 1 in $[\gamma_- - 1, \gamma_+ + 1]$ and constant 0 outside $[a, b] := [\gamma_- - 3, \gamma_+ + 3]$. We define $f_\chi(x) := f(x)\chi(x)$ and its almost analytic extension

$$f_{\mathbf{C}}(x + i\eta) := \left(f_\chi(x) + i\eta f'_\chi(x)\right)\tilde{\chi}(\eta), \tag{3.21}$$

where $\tilde{\chi} : \mathbf{R} \to \mathbf{R}$ is a smooth cut-off function which is constant 1 in $[-5, 5]$ and constant 0 outside $[-10, 10]$. By this definition it follows that $f_{\mathbf{C}}$ is bounded and compactly supported. Furthermore for small η we have that

$$\partial_{\bar{z}} f_{\mathbf{C}}(x + i\eta) = \mathcal{O}(\eta) \text{ and } \partial_\eta \partial_{\bar{z}} f_{\mathbf{C}}(x + i\eta) = \mathcal{O}(1). \tag{3.22}$$

We use the following representation of f_N from [89]:

$$f_N = \frac{2}{\pi} \Re \int_{\mathbf{R}} \int_{\mathbf{R}_+} \partial_{\bar{z}} f_{\mathbf{C}}(x + i\eta) \Delta_N(x + i\eta)\, dx d\eta. \tag{3.23}$$

We first exclude a critical area very close to the real line in the integral in (3.23). From the resolvent identities $|\eta \langle \boldsymbol{x}, G^2 \boldsymbol{x} \rangle| \le \Im \langle \boldsymbol{x}, G\boldsymbol{x} \rangle$. Then, we have that

$$\left|\eta z \left\langle \boldsymbol{x}, G^2 \boldsymbol{x} \right\rangle + \eta \langle \boldsymbol{x}, G\boldsymbol{x} \rangle + \eta\right| \le 2 \left|z + z \langle \boldsymbol{x}, G\boldsymbol{x} \rangle\right|.$$

Hence, we conclude that

$$|\eta \Delta_N(x + i\eta)| \le 2. \tag{3.24}$$

To study f_N we restrict our integration to the domain $\Im z \in [\eta_0, 10]$, with $\eta_0 := N^{-\frac{2}{3}}$. Thanks to (3.22) and (3.24), we find that

$$f_N = \frac{2}{\pi} \Re \int_{\mathbf{R}} \int_{\eta_0}^{10} \partial_{\bar{z}} f_{\mathbf{C}}(x + i\eta) \Delta_N(x + i\eta)\, dx d\eta + \mathcal{O}_{\prec}(\eta_0).$$

Then, for $\Im z = \eta \geq \eta_0$ we claim that the leading order term of $\Delta_N(z)$ is given by

$$\widehat{\Delta}_N(z) := \frac{1 + \frac{1}{N\sqrt{\phi}}\mathrm{Tr}G(z) + \tilde{z}\frac{1}{N}\mathrm{Tr}G^2(z)}{-z - \tilde{z}\frac{1}{N}\mathrm{Tr}G(z)}, \tag{3.25}$$

with the notation $\tilde{z} := \phi^{-\frac{1}{2}} z$ for brevity. Note that (3.25) is related to (3.20) by taking expectation with respect to $\boldsymbol{x}$ in the numerator and denominator separately.

We split the analysis of f_N into two parts: the leading order term

$$\widehat{\Omega}_f := \frac{2}{\pi}\Re \int_{\mathbf{R}} \int_{\eta_0}^{10} \partial_{\bar{z}} f_{\mathbf{C}}(x+i\eta)\widehat{\Delta}_N(x+i\eta)\, d\eta dx \tag{3.26}$$

and the fluctuation term

$$F_N := \frac{2}{\pi}\Re \int_{\mathbf{R}} \int_{\eta_0}^{10} \partial_{\bar{z}} f_{\mathbf{C}}(x+i\eta) \left(\Delta_N(x+i\eta) - \widehat{\Delta}_N(x+i\eta)\right) d\eta dx. \tag{3.27}$$

In this way we have that

$$f_N = \widehat{\Omega}_f + F_N + \mathcal{O}_\prec\left(N^{-\frac{2}{3}}\right).$$

In the following two sections, we will show that $\widehat{\Omega}_f = \Omega_f + \mathcal{O}_\prec(N^{-\frac{2}{3}})$ and $\mathbf{E}(F_N^2) = \frac{1}{N}V_f + \mathcal{O}_\prec(N^{-\frac{7}{6}})$, with some N-independent constant V_f, which will prove Theorem 3.2.2 in the sense of mean and variance.

3.4.1 Leading term: calculation of the mean.

The main tool we will use is the local law for the Marchenko-Pastur distribution in its averaged and entry-wise from. These results have first been proven in [30] (see Theorem 2.4 and Theorem 2.5) uniformly for each $z \in \mathbf{S}$, where

$$\mathbf{S} \equiv \mathbf{S}(\omega, \eta_0) := \left\{ z = x + i\eta \in \mathbf{C} : \kappa_x \leq \omega^{-1},\, \eta_0 \leq \eta \leq \omega^{-1},\, |z| \geq \omega \right\},$$

with some $\omega \in (0,1)$ fixed and $\kappa_x := \min\{|\gamma_+ - x|,\, |\gamma_- - x|\}$. In our proof, instead, we rely on local laws which hold true for each $z \in \mathbf{H}$, hence, combining the results in [30] with Theorem 2.7 and Theorem 2.9 respectively for $\phi = 1$ and $d_* \leq |\phi - 1| \leq \hat{d}$ in [12], we get the Marchenko-Pastur local law in the averaged form

$$\begin{aligned} m_R(z) &:= \frac{1}{N}\mathrm{Tr}R(z) = m_\phi(z) + \mathcal{O}_\prec\left(\frac{1}{N\eta}\right), \\ m_G(z) &:= \frac{1}{M}\mathrm{Tr}G(z) = m_{\phi^{-1}}(z) + \mathcal{O}_\prec\left(\frac{1}{N\eta}\right), \end{aligned} \tag{3.28}$$

and its entry-wise form

$$|R_{\mu\nu}(z) - \delta_{\mu\nu} m_\phi(z)| \prec \frac{1}{\sqrt{N\eta|z|}}, \quad \left|G_{ij}(z) - \delta_{ij} m_{\phi^{-1}}(z)\right| \prec \frac{1}{\sqrt{N\eta|z|}} \tag{3.29}$$

uniformly for each $z \in \mathbf{H}$.

Remark 3.4.3. *Notice that in* (3.28) *and* (3.29) *the error term from [12] is smaller in some particular cases, but we will not need these optimal bounds and we write local laws in a unified form which hold true for both the cases* $\phi = 1$ *and* $d_* \leq |\phi - 1| \leq \hat{d}$.

By (3.14), we have that

$$\tilde{z} m_G(z) = \tilde{z} m_R(z) - \phi^{\frac{1}{2}} + \phi^{-\frac{1}{2}} + \frac{\phi^{\frac{1}{2}}}{N}. \tag{3.30}$$

Hence, using the equality above, (3.7) and (3.28), we write (3.25) as follows

$$\begin{aligned} \widehat{\Delta}_N(z) &= \frac{1 + \frac{1}{N\sqrt{\phi}} \mathrm{Tr} G(z) + \tilde{z} \frac{1}{N} \mathrm{Tr} G^2(z)}{-z + \phi^{\frac{1}{2}} - \phi^{-\frac{1}{2}} - \tilde{z} m_R(z)} \\ &= m_\phi(z) \left(1 + \frac{1}{N\sqrt{\phi}} \mathrm{Tr} G(z) + \tilde{z} \frac{1}{N} \mathrm{Tr} G^2(z) \right) + \mathcal{O}_\prec \left(\frac{1}{N\eta} \right). \end{aligned} \tag{3.31}$$

Hence, thanks to (3.31) and (3.22), we obtain

$$\widehat{\Omega}_f = \frac{2}{\pi} \Re \int_{\mathbf{R}} \int_{\eta_0}^{10} \partial_{\bar{z}} f_{\mathbf{C}}(z) m_\phi(z) \left(1 + \frac{1}{N\sqrt{\phi}} \mathrm{Tr} G(z) + \frac{\tilde{z}}{N} \mathrm{Tr} G^2(z) \right) d\eta dx + \mathcal{O}_\prec \left(\frac{1}{N} \right),$$

where from now on we will use the notation $z = x + i\eta$ and $z_0 = x + i\eta_0$. Furthermore, we notice that, using (3.30) and the identity $\partial_z \mathrm{Tr} G(z) = \mathrm{Tr} G^2(z)$, we get

$$\begin{aligned} 1 + \frac{1}{N\sqrt{\phi}} \mathrm{Tr} G(z) + \tilde{z} \frac{1}{N} \mathrm{Tr} G^2(z) &= \partial_\eta \left(\eta - i\phi^{-\frac{1}{2}} (x + i\eta) \frac{1}{N} \mathrm{Tr} G(x + i\eta) \right) \\ &= \partial_\eta \left(\eta - i\phi^{-\frac{1}{2}} (x + i\eta) m_R(x + i\eta) \right). \end{aligned}$$

Hence, integrating by parts twice in η, using that the upper limit of the $\eta-$integration is zero since $\partial_{\bar{z}} f_{\mathbf{C}}(x + 10i) = 0$ by the definition of $\tilde{\chi}$, we have

$$\begin{aligned} \widehat{\Omega}_f = &-\frac{2}{\pi} \Re \int_{\mathbf{R}} \partial_{\bar{z}} f_{\mathbf{C}}(z_0) m_\phi(z_0) (\eta_0 - i\tilde{z}_0 m_R(z_0))\, dx &(3.32)\\ &- \frac{2}{\pi} \Re \int_{\mathbf{R}} \int_{\eta_0}^{10} \partial_\eta (\partial_{\bar{z}} f_{\mathbf{C}}(z) m_\phi(z)) (\eta - i\tilde{z} m_R(z))\, d\eta dx + \mathcal{O}_\prec \left(N^{-1} \right) &(3.33)\\ = &-\frac{2}{\pi} \Re \int_{\mathbf{R}} \int_{\eta_0}^{10} \partial_\eta (\partial_{\bar{z}} f_{\mathbf{C}}(z) m_\phi(z)) (\eta - i\tilde{z} m_\phi(z))\, d\eta dx &(3.34)\\ &- \frac{2}{\pi} \Re \int_{\mathbf{R}} \int_{\eta_0}^{10} \partial_\eta (\partial_{\bar{z}} f_{\mathbf{C}}(z) m_\phi(z)) (-i\tilde{z} m_R(z) + i\tilde{z} m_\phi(z))\, d\eta dx + \mathcal{O}_\prec (\eta_0) &(3.35)\\ = &\frac{2}{\pi} \Re \int_{\mathbf{R}} \int_{\eta_0}^{10} \partial_{\bar{z}} f_{\mathbf{C}}(z) m_\phi(z) (1 + (\tilde{z} m_\phi(z))') d\eta dx + \mathcal{O}_\prec (\eta_0) + \mathcal{O}_\prec \left(\frac{|\log \eta_0|}{N} \right), &(3.36) \end{aligned}$$

where we used that $\partial_{\bar{z}} f_{\mathbf{C}}(x + i\eta)$ scales like η near the real axis by (3.22), the local law from (3.28) and that $|z\phi^{-\frac{1}{2}} \partial_\eta (\partial_{\bar{z}} f_{\mathbf{C}} m_\phi(z))| \leq C$ from the bounds (3.16) and (3.17). In the last step we also used that $-i\partial_\eta h(z) = \partial_z h(z)$ for any analytic function h.

In summary, by (3.34), we conclude that

$$\widehat{\Omega}_f = \frac{2}{\pi}\Re \int_{\mathbf{R}} \int_{\eta_0}^{10} \partial_{\bar{z}} f_{\mathbf{C}}(z) p_\phi(z)\, d\eta dx + \mathcal{O}_\prec(\eta_0), \tag{3.37}$$

where for brevity we introduced

$$p_\phi(z) := m_\phi(z)[1 + (\tilde{z} m_\phi(z))'], \qquad z \in \mathbf{H}. \tag{3.38}$$

For the main term we need the following lemma (see Lemma 3.4 in [89]).

Lemma 3.4.4. *Let $\varphi, \psi : [a,b] \times [0, 10i] \to \mathbf{C}$ be functions such that $\partial_{\bar{z}}\psi(z) = 0$, $\varphi, \psi \in H^1$ and φ vanishes at the left, right and top of the boundary of the integration region. Then for any $\tilde{\eta} \in [0, 10]$, we have*

$$\int_a^b \int_{\tilde{\eta}}^{10} (\partial_{\bar{z}}\varphi(z))\psi(z)\, d\eta dx = \frac{1}{2i}\int_a^b \varphi(x+i\tilde{\eta})\psi(x+i\tilde{\eta})\, dx.$$

In order to compute the leading term defined in (3.26) we extend the integral in (3.37) to the real axis. For this purpose we introduce a tiny auxiliary scale η_1, say $\eta_1 := N^{-10}$. We recall that $f_{\mathbf{C}}$ is supported in $[a,b] \times [-10, 10]$, with $a = \gamma_- - 3$ and $b = \gamma_+ + 3$, where γ_-, γ_+ are the spectral edges, and $\kappa_x = \min\{|x-\gamma_-|, |x-\gamma_+|\}$.

Since by (3.16), (3.17) and (3.22), we have that

$$\left| \frac{2}{\pi}\Re \int_{\mathbf{R}} \int_{\eta_1}^{\eta_0} \partial_{\bar{z}} f_{\mathbf{C}}(z) p_\phi(z)\, d\eta dx \right| \lesssim \int_a^b \int_{\eta_1}^{\eta_0} \left(\frac{\eta}{|z|} + \frac{\eta}{|z|^{\frac{1}{2}}\sqrt{\kappa_x+\eta}} \right) d\eta dx \lesssim \eta_0^{\frac{3}{2}},$$

we conclude that

$$\widehat{\Omega}_f = \frac{2}{\pi}\Re \int_{\mathbf{R}} \int_{\eta_1}^{10} \partial_{\bar{z}} f_{\mathbf{C}}(z) p_\phi(z)\, d\eta dx + \mathcal{O}_\prec(\eta_0). \tag{3.39}$$

Next, applying Lemma 3.4.4 to the integral in the r.h.s. of (3.39), we conclude

$$\widehat{\Omega}_f = \frac{1}{\pi}\Im \int_{\mathbf{R}} f_{\mathbf{C}}(x+i\eta_1) p_\phi(x+i\eta_1)\, dx + \mathcal{O}_\prec(\eta_0). \tag{3.40}$$

By (3.21) and (3.22), using the bounds (3.16)–(3.17), it easily follows that

$$\widehat{\Omega}_f = \frac{1}{\pi}\int_{\mathbf{R}} f(x) \Im p_\phi(x+i\eta_1)\, dx + \mathcal{O}_\prec(\eta_0). \tag{3.41}$$

We notice that

$$w_\phi(z) := \sqrt{\phi}(1 + z m_{\phi^{-1}}(z)) = \frac{\phi^{\frac{1}{2}} + \phi^{-\frac{1}{2}} - z + i\sqrt{(z-\gamma_-)(\gamma_+ - z)}}{2} \tag{3.42}$$

is the Stieltjes transform of the Wigner semicircle law centered at $\phi^{\frac{1}{2}} + \phi^{-\frac{1}{2}}$. Hence, w_ϕ is also characterized as the unique solution of

$$w_\phi(z) + \frac{1}{z - \phi^{\frac{1}{2}} - \phi^{-\frac{1}{2}} + w_\phi(z)} = 0, \quad \Im w_\phi > 0. \tag{3.43}$$

Notice that $w_\phi(z) = w_{\phi^{-1}}(z)$ and that, using the self consistent equation (3.7) and the relation between m_ϕ and $m_{\phi^{-1}}$ in (3.11), we have

$$w_\phi(z) = -zm_\phi(z)m_{\phi^{-1}}(z). \tag{3.44}$$

We now distinguish the cases $\phi = 1$ and $|\phi - 1| \geq d_*$, since for $\phi = 1$ the integral in (3.41) has an additional singularity in zero which we have to take into account.

We start with the case $|\phi - 1| \geq d_*$. In this case $\gamma_- \geq \tau(d_*)$, for some $\tau(d_*) > 0$. By equations (3.7) and (3.11), expressing $(\tilde{z}m_\phi)' = w_\phi'$ from differentiating the self consistent equation for w_ϕ in (3.43), it follows that

$$w_\phi(z)' = \frac{w_\phi^2(z)}{1 - w_\phi^2(z)}, \tag{3.45}$$

and so we may write p_ϕ from (3.38) as

$$p_\phi(z) = \frac{m_\phi(z)}{1 - w_\phi^2(z)}. \tag{3.46}$$

Furthermore, by Lemma 3.6 of [31] we have that

$$|1 - w_\phi^2(z)| \asymp \sqrt{\kappa_x + \eta}, \qquad c \leq |w_\phi(z)| \leq 1, \tag{3.47}$$

with some ϕ-independent constant $c > 0$, for any $z = x + i\eta$ such that $|z - \sqrt{\phi}| \leq 10$.

To evaluate $\widehat{\Omega}_f$ in (3.41), we first remove the η_1 in the argument of p_ϕ. We proceed writing $p_\phi(x + i\eta_1) - p_\phi(x)$ as follows

$$p_\phi(x + i\eta_1) - p_\phi(x) = \frac{1}{1 - w_\phi(x + i\eta_1)^2} \int_0^{\eta_1} m_\phi(x + i\eta)'\, d\eta \tag{3.48}$$

$$+ \frac{m_\phi(x)(w_\phi(x + i\eta_1) + w_\phi(x))}{(1 - w_\phi(x + i\eta_1)^2)(1 - w_\phi(x)^2)} \int_0^{\eta_1} w_\phi(x + i\eta)'\, d\eta. \tag{3.49}$$

Then, by (3.16)–(3.17) and (3.45)–(3.48), simple estimates give that

$$\Big| p_\phi(x + i\eta_1) - p_\phi(x) \Big| \lesssim \frac{\eta_1^{1/4}}{|x|^{1/2}\kappa_x^{1/4}\sqrt{\kappa_x + \eta_1}} + \frac{\sqrt{\eta_1}}{|x|^{1/2}\sqrt{\kappa_x(\kappa_x + \eta_1)}} \lesssim \frac{\eta_1^{1/4}}{|x|^{1/2}\kappa_x^{3/4}},$$

for any $x \in \mathbf{R}$. Hence, if $|\phi - 1| \geq d_*$, integrating over x, we conclude that

$$\left| \frac{1}{\pi} \int_a^b f(x) \Im[p_\phi(x + i\eta_1) - p_\phi(x)]\, dx \right| \lesssim \eta_1^{1/4}. \tag{3.50}$$

In particular, this implies that $\Im p_\phi(z)$ is of order $\eta_1^{1/4}$ outside the interval $[\gamma_-, \gamma_+]$, since $\Im p_\phi(x) = 0$ for $x \notin [\gamma_-, \gamma_+]$. Moreover, (3.38), (3.41) and (3.50) imply that

$$\begin{aligned}
\widehat{\Omega}_f &= \frac{1}{\pi} \int_{\gamma_-}^{\gamma_+} f(x) \Im \left[m_\phi(x)(1 + (x\phi^{-\frac{1}{2}} m_\phi(x))') \right] dx + \mathcal{O}_\prec \left(N^{-\frac{2}{3}} \right) \\
&= \int_{\gamma_-}^{\gamma_+} f(x) \frac{\sqrt{\phi}}{4\pi^2 x \rho_\phi(x)} \left(1 + \frac{\sqrt{\phi} - \frac{1}{\sqrt{\phi}}}{x} \right) dx + \mathcal{O}_\prec \left(N^{-\frac{2}{3}} \right),
\end{aligned}$$

concluding the estimate for the leading term of $\mathbf{E} f_N$ when $|\phi - 1| \geq d_*$.

Now we consider the case $\phi = 1$, when $\gamma_- = 0$ and $\gamma_+ = 4$. In this case, the computation of the integral (3.41) is a bit more delicate since the singularities around $x \approx 0$ and $\kappa_x \approx 0$ overlap. For brevity, in the rest of this section we use the notation $m = m(z) := m_{\phi=1}(z)$ and $w = w(z) := w_{\phi=1}(z)$ for any $z \in \mathbf{H}$. Expressing m' from differentiating the self consistent equation (3.7), using (3.7) repeatedly and the relation (3.44), a simple calculation gives that

$$p = m(1 + (zm)') = -\frac{1}{z} \cdot \frac{1}{1 - zm^2} = -\frac{1}{z} \cdot \frac{1}{1 + w}, \tag{3.51}$$

with $p = p(z) := p_{\phi=1}(z)$. We also define

$$q(z) := \frac{1}{1 + w(z)}, \qquad z \in \mathbf{H}. \tag{3.52}$$

As a consequence of (3.51)-(3.52), it follows that

$$\left| \int_a^b f(x) \Im [p(x + i\eta_1) - p(x)] \, dx - \frac{\pi f(0)}{2} \right| \tag{3.53}$$

$$\leq \left| \int_a^b f(x) \Big[\frac{x}{x^2 + \eta_1^2} \Im q(x + i\eta_1) - \frac{1}{x} \Im q(x) \Big] \, dx \right| \tag{3.54}$$

$$+ \left| \int_a^b f(x) \frac{\eta_1}{x^2 + \eta_1^2} \Re q(x + i\eta_1) \, dx - \frac{\pi f(0)}{2} \right|. \tag{3.55}$$

We start estimating (3.55). Using explicit computations, by the expression in (3.42) for $\phi = 1$, we conclude that

$$(3.55) \leq \left| \frac{1}{2} \int_a^b f(x) \frac{\eta_1}{x^2 + \eta_1^2} \, dx - \frac{\pi f(0)}{2} \right| + \mathcal{O}(\sqrt{\eta_1}) \lesssim \sqrt{\eta_1}. \tag{3.56}$$

Furthermore, since

$$|1 + w(z)| = |1 - zm(z)^2| \asymp \frac{\sqrt{\kappa_x + \eta}}{|z|^{\frac{1}{2}}},$$

by (3.18), using (3.42) and the definition of q in (3.52), it also follows that the integrand in (3.54) is bounded by

$$\frac{f(x)|x|^{3/2}\sqrt{\eta_1}}{(x^2 + \eta_1^2)^{\frac{3}{4}} \sqrt{\kappa_x(\kappa_x + \eta_1)}} + \frac{f(x)\eta_1^2}{|x|^{1/2}|4 - x|^{1/2}(x^2 + \eta_1^2)}, \tag{3.57}$$

for any $x \in \mathbf{R}$. Then, combining (3.56) with the integral of (3.57), we conclude

$$\left| \frac{1}{\pi} \int_a^b f(x) \Im [p(x + i\eta_1) - p(x)] \, dx - \frac{f(0)}{2} \right| \lesssim \eta_1^{1/4}. \tag{3.58}$$

Similarly to the case $|\phi - 1| \geq d_*$, this bound implies that $\Im p(x + i\eta_1)$ is of order $\eta_1^{1/4}$ outside $[0, 4]$. Hence, the above inequality implies that

$$\widehat{\Omega}_f = \frac{1}{\pi} \int_0^4 \frac{f(x)}{4\pi^2 x \rho_1(x)} \, dx + \frac{f(0)}{2} + \mathcal{O}_{\prec}\left(N^{-\frac{2}{3}}\right),$$

concluding the computation of Ω_f, the leading term of $\mathbf{E} f_N$ in Theorem 3.2.2.

3.4.2 Fluctuation term

We write the difference $\Delta_N(z) - \widehat{\Delta}_N(z)$ in a more convenient form to study the integral in (3.27). The key point is to express it as a derivative (up to an error) to prepare it for an integration by parts. Let $\hat{z}$ be defined as $\hat{z} := z\phi^{\frac{1}{2}}$.

Lemma 3.4.5. *For any $\eta > \eta_0$ we have that*

$$\Delta_N(z) - \widehat{\Delta}_N(z) = \partial_z \frac{z\langle \boldsymbol{x}, G\boldsymbol{x}\rangle - \hat{z}m_G(z)}{-z - \hat{z}m_G(z)} + \mathcal{O}_\prec\left(\frac{1}{N\eta^2}\right). \tag{3.59}$$

Proof. This lemma, using (3.3), relies on the following large deviation bound (see, e.g. Lemma 3.1 in [30])

$$\langle \boldsymbol{x}, G\boldsymbol{x}\rangle = \frac{1}{\sqrt{MN}}\mathrm{Tr}G + \mathcal{O}_\prec\left(\sqrt{(MN)^{-1}\mathrm{Tr}|G|^2}\right), \tag{3.60}$$

and a similar formula for $\langle \boldsymbol{x}, G^2\boldsymbol{x}\rangle$.

In the following part of the proof, in order to abbreviate our notation, we use $G := G(z)$, $m_G := m_G(z)$. Using (3.20) and (3.25), we have

$$\begin{aligned}\Delta_N(z) - \widehat{\Delta}_N(z) = &\frac{(z\langle \boldsymbol{x}, G^2\boldsymbol{x}\rangle + \langle \boldsymbol{x}, G\boldsymbol{x}\rangle + 1)(-z - \hat{z}m_G)}{(-z - z\langle \boldsymbol{x}, G\boldsymbol{x}\rangle)(-z - \hat{z}m_G)} \\ &+ \frac{(-1 - \phi^{\frac{1}{2}}m_G - \hat{z}m_G')(-z - z\langle \boldsymbol{x}, G\boldsymbol{x}\rangle)}{(-z - z\langle \boldsymbol{x}, G\boldsymbol{x}\rangle)(-z - \hat{z}m_G)}.\end{aligned} \tag{3.61}$$

Now we claim that

$$\Delta_N(z) - \widehat{\Delta}_N(z) = \partial_z \frac{z\langle \boldsymbol{x}, G\boldsymbol{x}\rangle - \hat{z}m_G}{-z - \hat{z}m_G} + \mathcal{E},$$

with an error term $\mathcal{E}$ we will determine along the proof. We start with

$$\partial_z \frac{z\langle \boldsymbol{x}, G\boldsymbol{x}\rangle - \hat{z}m_G}{-z - \hat{z}m_G} = \frac{(-z - \hat{z}m_G)\left(\langle \boldsymbol{x}, G\boldsymbol{x}\rangle + z\langle \boldsymbol{x}, G^2\boldsymbol{x}\rangle - \phi^{\frac{1}{2}}m_G - \hat{z}m_G'\right)}{(-z - \hat{z}m_G)^2} \tag{3.62}$$

$$- \frac{(-1 - \phi^{\frac{1}{2}}m_G - \hat{z}m_G')(z\langle \boldsymbol{x}, G\boldsymbol{x}\rangle - \hat{z}m_G)}{(-z - \hat{z}m_G)^2}. \tag{3.63}$$

Using $m_G(z) = \frac{1}{M}\mathrm{Tr}G(z)$ and $m_G'(z) = \frac{1}{M}\mathrm{Tr}G^2(z)$ we write the r.h.s. of (3.61) as

$$\begin{aligned}\Delta_N(z) - \widehat{\Delta}_N(z) = &\frac{\langle \boldsymbol{x}, G\boldsymbol{x}\rangle + z\langle \boldsymbol{x}, G^2\boldsymbol{x}\rangle - \phi^{\frac{1}{2}}m_G - \hat{z}m_G'}{(-z - \hat{z}m_G) - (z\langle \boldsymbol{x}, G\boldsymbol{x}\rangle - \hat{z}m_G)} \\ &- \frac{(-1 - \phi^{\frac{1}{2}}m_G - \hat{z}m_G')(z\langle \boldsymbol{x}, G\boldsymbol{x}\rangle - \hat{z}m_G)}{(-z - \hat{z}m_G)^2 - (-z - \hat{z}m_G)(z\langle \boldsymbol{x}, G\boldsymbol{x}\rangle - \hat{z}m_G)}.\end{aligned} \tag{3.64}$$

By (3.28), (3.60) and the bound in (3.16) it follows that

$$z\langle \boldsymbol{x}, G\boldsymbol{x}\rangle - \hat{z}m_G(z) \prec \frac{|z|}{\sqrt{MN}}\sqrt{\mathrm{Tr}|G(z)|^2} \le \frac{|z|}{\sqrt{MN}}\sqrt{\frac{1}{\eta}\Im\mathrm{Tr}G(z)} \prec \frac{|z|^{\frac{3}{4}}}{\sqrt{N\eta}} \tag{3.65}$$

and also

$$z\left\langle \boldsymbol{x}, G^2\boldsymbol{x}\right\rangle - \hat{z}m_G'(z) \prec \frac{|z|}{\sqrt{MN}}\sqrt{\mathrm{Tr}|G|^4} \le \frac{|z|}{\sqrt{MN\eta}}\sqrt{\mathrm{Tr}|G(z)|^2} \prec \frac{|z|^{\frac{1}{2}}}{\sqrt{N\eta^3}}. \quad (3.66)$$

Note that the leading term in the denominators in (3.64) is separated away from zero since $-z-\hat{z}m_{\phi^{-1}}(z) = [m_{\phi^{-1}}(z)]^{-1}$, by (3.7) and (3.11). Thus these denominators are stable under small perturbations. Hence, replacing $z\langle \boldsymbol{x}, G\boldsymbol{x}\rangle$ in the denominator with $\hat{z}m_G(z) + \mathcal{O}_\prec\left(\frac{1}{\sqrt{N\eta}}\right)$ and comparing (3.63) and (3.64), we conclude that

$$\Delta_N(z) - \widehat{\Delta}_N(z) = \partial_z \frac{z\langle \boldsymbol{x}, G\boldsymbol{x}\rangle - \hat{z}m_G}{-z-\hat{z}m_G} + \mathcal{O}_\prec\left(\frac{1}{N\eta^2}\right).$$

In estimating various error terms along the proof we used that $zm_G(z) = \mathcal{O}_\prec(1)$ (by (3.28) and (3.16)) and that $zm_G'(z) = \mathcal{O}_\prec\left(\eta^{-1}\right)$ by (3.15) and (3.16). □

Next, we use (3.59) to estimate the fluctuation term F_N as defined in (3.27) via an integration by parts

$$\begin{aligned} F_N = &-\frac{2}{\pi}\Re\int_{\mathbf{R}} \partial_{\bar{z}} f_{\mathbf{C}}(z_0) i \frac{z_0\langle \boldsymbol{x}, G(z_0)\boldsymbol{x}\rangle - \hat{z}_0 m_G(z_0)}{-z_0-\hat{z}_0 m_G(z_0)}\, dx \\ &+ \frac{2}{\pi}\Re\int_{\mathbf{R}}\int_{\eta_0}^{10} \partial_\eta\partial_{\bar{z}} f_{\mathbf{C}}(z) i \frac{z\langle \boldsymbol{x}, G(z)\boldsymbol{x}\rangle - \hat{z}m_G(z)}{-z-\hat{z}m_G(z)}\, d\eta dx + \mathcal{O}_\prec\left(\frac{|\log\eta_0|}{N}\right), \end{aligned}$$

with $\hat{z}_0 := \phi^{\frac{1}{2}} z_0$. Then, we continue with the estimate

$$\frac{z\langle \boldsymbol{x}, G(z)\boldsymbol{x}\rangle - \hat{z}m_G(z)}{-z-\hat{z}m_G(z)} = m_\phi(z)\left(z\langle \boldsymbol{x}, G(z)\boldsymbol{x}\rangle - \hat{z}m_G(z)\right) + \mathcal{O}_\prec\left(\frac{1}{(N\eta)^{\frac{3}{2}}}\right)$$

from (3.28), (3.13), (3.7) and (3.65) to find that

$$F_N = -\frac{2}{\pi}\Re\int_{\mathbf{R}} m_\phi(z_0)\partial_{\bar{z}} f_{\mathbf{C}}(z_0) i\left(z_0\langle \boldsymbol{x}, G(z_0)\boldsymbol{x}\rangle - \hat{z}_0 m_G(z_0)\right) dx \quad (3.67)$$

$$+\frac{2}{\pi}\Re\int_{\mathbf{R}}\int_{\eta_0}^{10} m_\phi(z)\partial_\eta\partial_{\bar{z}} f_{\mathbf{C}}(z) i\left(z\langle \boldsymbol{x}, G(z)\boldsymbol{x}\rangle - \hat{z}m_G(z)\right) d\eta dx + \mathcal{O}_\prec\left(N^{-\frac{2}{3}}\right) \quad (3.68)$$

$$= -\frac{2}{\pi}\Im\int_{\mathbf{R}}\int_{\eta_0}^{10} m_\phi(z)\partial_\eta\partial_{\bar{z}} f_{\mathbf{C}}(z)\left(z\langle \boldsymbol{x}, G(z)\boldsymbol{x}\rangle - \hat{z}m_G(z)\right) d\eta dx + \mathcal{O}_\prec\left(N^{-\frac{2}{3}}\right), \quad (3.69)$$

where in the last step we used that by (3.22) and (3.65) it follows

$$\left|\partial_{\bar{z}} f_{\mathbf{C}}(z_0) i\left(z_0\langle \boldsymbol{x}, G(z_0)\boldsymbol{x}\rangle - \hat{z}_0 m_G(z_0)\right)\right| \prec \sqrt{\frac{\eta_0}{N}} \le N^{-\frac{2}{3}}.$$

The leading order expression for F_N has zero mean, hence we can start computing the variance $\mathrm{Var}(F_N) = \mathbf{E}F_N^2 + \mathcal{O}_\prec\left(N^{-\frac{4}{3}}\right)$ as

$$\mathbf{E}F_N^2 = \mathbf{E}\left(\frac{2}{\pi}\Im\int_{\mathbf{R}}\int_{\eta_0}^{10} m_\phi(z)\partial_\eta\partial_{\bar{z}} f_{\mathbf{C}}(z)\left(z\langle \boldsymbol{x}, G(z)\boldsymbol{x}\rangle - \hat{z}m_G(z)\right) d\eta dx\right)^2 + \mathcal{O}_\prec\left(N^{-\frac{7}{6}}\right).$$

When we use the expectation $\mathbf{E}$ we frequently use the property that if X and Y are random variables with $X = \mathcal{O}_{\prec}(Y)$, $Y \geq 0$ and $|X| \leq N^C$ for some constant C, then $\mathbf{E}|X| \prec \mathbf{E}Y$, or, equivalently, $\mathbf{E}|X| \leq N^{\epsilon}\mathbf{E}Y$ for any $\epsilon > 0$ and $N \geq N_0(\epsilon)$. To compute the leading term F_N' in $\mathbf{E}F_N^2$ we introduce the short-hand notations

$$g(z) := \frac{2}{\pi} z m_\phi(z) \partial_\eta \partial_{\overline{z}} f_{\mathbf{C}}(z), \quad A(z) := \sqrt{N}\left(\langle \boldsymbol{x}, G(z)\boldsymbol{x}\rangle - \phi^{\frac{1}{2}} m_G(z)\right) \tag{3.70}$$

to write

$$F_N' := \frac{1}{N}\mathbf{E}\left(\Im \int_{\mathbf{R}} \int_{\eta_0}^{10} g(z)A(z)\, d\eta dx\right)^2.$$

We will often use the following identity for any $z, w \in \mathbf{C}$:

$$(\Im z)(\Im w) = \frac{1}{2}\Re(\overline{z}w - zw). \tag{3.71}$$

Thanks to (3.71) we write

$$F_N' = \frac{1}{2N}\Re \iint_{\mathbf{R}} \iint_{\eta_0}^{10} \left[g(z)g(\overline{z}')\mathbf{E}\left(A(z)A(\overline{z}')\right) - g(z)g(z')\mathbf{E}\left(A(z)A(z')\right)\right] d\eta d\eta' dx dx', \tag{3.72}$$

where we used that $\overline{X(z)} = X(\overline{z})$ and $\overline{g(z)} = g(\overline{z})$. In the following we use the short notation $G = G(z),\ G' = G(z')$.

To study the expectation of $A(z)A(z')$, we consider

$$\begin{aligned} A(z)A(z') =& N\left(\sum_{i,j=1,\, i\neq j}^{M} \overline{\boldsymbol{x}_i} G_{ij} \boldsymbol{x}_j + \sum_{i=1}^{M}\left(|\boldsymbol{x}_i|^2 - \frac{1}{\sqrt{MN}}\right) G_{ii}\right) \\ &\times \left(\sum_{l,k=1,\, l\neq k}^{M} \overline{\boldsymbol{x}_l} G'_{lk} \boldsymbol{x}_k + \sum_{l=1}^{M}\left(|\boldsymbol{x}_l|^2 - \frac{1}{\sqrt{MN}}\right) G'_{ll}\right). \end{aligned}$$

The conditional expectation $\mathbf{E}_1 = \mathbf{E}(\cdot|X)$ conditioned on the matrix X gives

$$\mathbf{E}_1(A(z)A(z')) = \frac{1}{\phi N}\sum_{i,j=1,\, i\neq j}^{M}\left(G_{ij}G'_{ji} + |\sigma_2|^2 G_{ij}G'_{ij}\right) + \frac{\sigma_4 - 1}{M}\sum_{i=1}^{M} G_{ii}G'_{ii} \tag{3.73}$$

$$= \frac{1}{\phi N}\sum_{i,j=1,\, i\neq j}^{M}\left(G_{ij}G'_{ji} + |\sigma_2|^2 G_{ij}G'_{ij}\right) + (\sigma_4 - 1) m_{\phi^{-1}}(z) m_{\phi^{-1}}(z') \tag{3.74}$$

$$+ \mathcal{O}_{\prec}\left(\frac{1}{|zz'|^{\frac{1}{2}}}\left(\frac{1}{\sqrt{N\eta}} + \frac{1}{\sqrt{N\eta'}} + \frac{1}{N\sqrt{\eta\eta'}}\right)\right), \tag{3.75}$$

where we used that $\mathbf{E}\boldsymbol{x}_i^2 = \mathbf{E}\widetilde{X}_{i1}^2 = \frac{\sigma_2}{\sqrt{MN}}$ and $\mathbf{E}|\boldsymbol{x}_i|^4 = \mathbf{E}|\widetilde{X}_{i1}|^4 = \frac{\sigma_4}{MN}$ for each $i = 1, \dots, M$. In the last step we also used (3.29).

To continue with the study of the fluctuation term we need to find an expression for $\frac{1}{\phi N}\sum_{i,j=1,\, i\neq j}^{M} G_{ij}G'_{ji}$ and $\frac{1}{\phi N}\sum_{i,j=1,\, i\neq j}^{M} G_{ij}G'_{ij}$ in terms of m_ϕ and $m_{\phi^{-1}}$.

Lemma 3.4.6. *For $z = x + i\eta$, $z' = x' + i\eta'$, $\eta, \eta' > \eta_0$, with $|z - \sqrt{\phi}| \leq 10$ and $|z' - \sqrt{\phi}| \leq 10$, it holds*

$$\frac{1}{\phi N} \sum_{\substack{i,j=1 \\ i \neq j}}^{M} G_{ij} G'_{ji} = \frac{zz' m_\phi(z) m_\phi(z') m_{\phi^{-1}}(z)^2 m_{\phi^{-1}}(z')^2}{1 - zz' m_\phi(z) m_\phi(z') m_{\phi^{-1}}(z) m_{\phi^{-1}}(z')} + \mathcal{O}_\prec\left(\frac{\Psi}{|zz'|^{\frac{1}{2}}}\right), \tag{3.76}$$

$$\frac{1}{\phi N} \sum_{\substack{i,j=1 \\ i \neq j}}^{M} G_{ij} G'_{ij} = \frac{|\sigma_2|^2 zz' m_\phi(z) m_\phi(z') m_{\phi^{-1}}(z)^2 m_{\phi^{-1}}(z')^2}{1 - |\sigma_2|^2 zz' m_\phi(z) m_\phi(z') m_{\phi^{-1}}(z) m_{\phi^{-1}}(z')} + \mathcal{O}_\prec\left(\frac{\Psi}{|zz'|^{\frac{1}{2}}}\right), \tag{3.77}$$

where

$$\Psi := \frac{1}{\eta + \eta'}\left(\frac{1}{\sqrt{N\eta\eta'^2}} + \frac{1}{\sqrt{N\eta^2\eta'}} + \frac{1}{N\eta\eta'}\right).$$

Proof. To prove this lemma we change our point of view and we study the linearized problem. We remark that (3.76), being a tracial quantity, could still be analyzed without linearization, but (3.77) cannot. For brevity we use the proof with linearization for both cases.

Let the $[(N-1)+M] \times [(N-1)+M]$ matrix $\mathcal{H}$ be defined as

$$\mathcal{H} := \begin{pmatrix} 0 & X^* \\ X & 0 \end{pmatrix}. \tag{3.78}$$

We introduced this bigger matrix $\mathcal{H}$ to study W, since $\mathcal{H}$ has the advantage that all nonzero elements are i.i.d. random variables (modulo symmetry) and it carries all information on the matrices $W = X^*X$ and XX^* we are studying. Indeed, $\mathcal{H}^2$ with diagonal blocks X^*X and XX^* has the same non zero spectrum as W (with double multiplicity).

To prove (3.76) we define the resolvents

$$\mathcal{G}(z) := (\mathcal{H}^2 - z)^{-1} \text{ and } \mathfrak{G}(\zeta) := (\mathcal{H} - \zeta)^{-1}. \tag{3.79}$$

Note that

$$\mathcal{G}(z) = \frac{1}{2\sqrt{z}} \cdot \left(\frac{1}{\mathcal{H} - \sqrt{z}} - \frac{1}{\mathcal{H} + \sqrt{z}}\right) = \frac{1}{2\sqrt{z}} \cdot \left(\mathfrak{G}(\sqrt{z}) - \mathfrak{G}(-\sqrt{z})\right), \tag{3.80}$$

where we chose the branch of $\sqrt{z}$ which lies in $\mathbf{H}$.

In the following we state some fundamental properties of the Gram matrix $\mathcal{H}$ and of its resolvent $\mathfrak{G}$ (for a detailed description see [10] and [12]). Let $m_1, m_2 : \mathbf{H} \to \mathbf{H}$ be the unique solutions of the system

$$\begin{cases} -\frac{1}{m_1} = \zeta + \phi^{\frac{1}{2}} m_2, \\ -\frac{1}{m_2} = \zeta + \phi^{-\frac{1}{2}} m_1. \end{cases} \tag{3.81}$$

Then, for each $\zeta \in \mathbf{H}$ (see [12]) we have

$$\begin{aligned} |\mathfrak{G}_{ij}(\zeta) - \delta_{ij} m_1(\zeta)| &\prec \frac{1}{\sqrt{N\Im\zeta}}, \qquad i,j = 2, \ldots N, \\ |\mathfrak{G}_{ij}(\zeta) - \delta_{ij} m_2(\zeta)| &\prec \frac{1}{\sqrt{N\Im\zeta}}, \qquad i,j = N+1, \ldots, N+M. \end{aligned} \tag{3.82}$$

Notice that if $z = x + i\eta$ is such that $\zeta^2 = z$ then $\frac{1}{\sqrt{N\Im\zeta}} \lesssim \frac{1}{\sqrt{N\eta}}$. Indeed, $\Im\zeta = \frac{\eta}{\Re\zeta} \gtrsim \eta$, since $|\zeta| \lesssim 1$ under the hypothesis $|z - \sqrt{\phi}| \le 10$ and (3.4). Hence all along the proof we will estimate the error terms only in terms of η. We will use ζ as the argument of the resolvent $\mathfrak{G}$, with $\zeta = \sqrt{z}$.

In particular m_1 and m_2 are Stieltjes transforms of symmetric probability measures on $\mathbf{R}$, whose support is contained in $[-2\phi^{\frac{1}{4}}, 2\phi^{-\frac{1}{4}}]$ (see Theorem 2.1 in [7]). Furthermore, we have that

$$m_\phi(z) = \frac{m_1(\zeta)}{\zeta}, \quad m_{\phi^{-1}}(z) = \frac{m_2(\zeta)}{\zeta} \tag{3.83}$$

and they are related in the following way:

$$\begin{cases} -\frac{1}{m_\phi(z)} = z + z\phi^{\frac{1}{2}} m_{\phi^{-1}}(z) \\ -\frac{1}{m_{\phi^{-1}}(z)} = z + z\phi^{-\frac{1}{2}} m_\phi(z). \end{cases} \tag{3.84}$$

By (3.83), using that an analogue of (3.16) holds substituting ϕ with ϕ^{-1} (see proof of Lemma 3.3.2 in Appendix A), we have that

$$|\phi^{-\frac{1}{4}} m_1(z)| \le 1 - c\eta, \quad |\phi^{\frac{1}{4}} m_2(z)| \le 1 - c\eta. \tag{3.85}$$

Next, we use a resolvent expansion to express the resolvents of $\mathcal{H}$ and $\mathcal{H}^2$ in terms of resolvents of their minors. For each $T \subset \{2, \dots, N+M\}$ we define

$$\mathcal{G}^{[T]}(z) := \left(\left(\mathcal{H}^{[T]} \right)^2 - z \right)^{-1} \quad \text{and} \quad \mathfrak{G}^{[T]}(\zeta) := (\mathcal{H}^{[T]} - \zeta)^{-1}, \tag{3.86}$$

where $\mathcal{H}^{[T]}$ is the matrix $\mathcal{H}$ with the rows and columns labeled with T set to zero:

$$\left(\mathcal{H}^{[T]} \right)_{ij} := \mathbf{1}(i \notin T)\mathbf{1}(j \notin T)\mathcal{H}_{ij}. \tag{3.87}$$

Let γ_{ij} denote the entries of the matrix $\mathcal{H}$, i.e. $\gamma_{ij} = X_{ij}$ for $i = N+1, \dots N+M$, $j = 2, \dots, N$, $\gamma_{ij} = \overline{\gamma_{ji}}$ for $i = 2, \dots, N$, $j = N+1, \dots N+M$ and $\gamma_{ij} = 0$ otherwise. From now on we abandon the convention in Remark 3.2.1 about Greek letters for columns indices and we use only $i, j, k, \dots$ We use the one sided expansion for the resolvent of $\mathcal{H}$, i.e. for each $i \neq j$ we have

$$\mathfrak{G}_{ij} = -\mathfrak{G}_{ii} \sum_{\substack{k=2 \\ k \neq j}}^{N+M} \mathfrak{G}_{ik}^{[j]} \gamma_{kj}. \tag{3.88}$$

Notice that here $\mathfrak{G}_{ik}^{[j]}$ is independent of γ_{kj} since $\mathcal{H}$ has independent elements.

By the definition of $\mathcal{H}^2$ and (3.80), using the identification $\zeta = \sqrt{z}$ choosing the branch of $\sqrt{z}$ which lies in $\mathbf{H}$, it follows that

$$\begin{aligned} \frac{1}{N} \sum_{\substack{i,j=1 \\ i \neq j}}^{M} G_{ij} G'_{ji} &= \frac{1}{N} \sum_{\substack{i,j=N+1 \\ i \neq j}}^{N+M} \mathcal{G}_{ij}(z) \mathcal{G}_{ji}(z') \\ &= \frac{1}{N} \sum_{\substack{i,j=N+1 \\ i \neq j}}^{N+M} \frac{1}{4\zeta\zeta'} \left(\mathfrak{G}(\zeta)_{ij} \mathfrak{G}(\zeta')_{ji} - \mathfrak{G}(\zeta)_{ij} \mathfrak{G}(-\zeta')_{ji} \right) \\ &\quad + \frac{1}{N} \sum_{\substack{i,j=N+1 \\ i \neq j}}^{N+M} \frac{1}{4\zeta\zeta'} \left(\mathfrak{G}(-\zeta)_{ij} \mathfrak{G}(-\zeta')_{ji} - \mathfrak{G}(-\zeta)_{ij} \mathfrak{G}(\zeta')_{ji} \right). \end{aligned} \tag{3.89}$$

We introduce the shorthand notation $\mathfrak{G}_{ij} := \mathfrak{G}_{ij}(\zeta)$, $\mathfrak{G}'_{ij} := \mathfrak{G}_{ij}(\zeta')$. By (3.82), for any i, j, k all distinct, it holds

$$\mathfrak{G}_{ik} = \mathfrak{G}_{ik}^{[j]} + \frac{\mathfrak{G}_{ij}\mathfrak{G}_{jk}}{\mathfrak{G}_{jj}} = \mathfrak{G}_{ik}^{[j]} + \mathcal{O}_{\prec}\left(\frac{1}{N\eta}\right). \tag{3.90}$$

We now derive a self consistent equation for $\sum_{i\neq j} \mathfrak{G}_{ij}\mathfrak{G}'_{ji}$, that is the first term in the second equality of (3.89).

For this purpose, we start proving that $\sum_{i\neq j} \mathfrak{G}_{ij}\mathfrak{G}'_{ji}$ is close to $\sum_{i\neq j} \mathbf{E}_j\mathfrak{G}_{ij}\mathfrak{G}'_{ji}$ where $\mathbf{E}_j(\cdot) := \mathbf{E}(\cdot|\mathcal{H}^{[j]})$ denotes the conditional expectation with respect to the matrix $\mathcal{H}^{[j]}$. This result is a special case of the fluctuation averaging analysis presented in [80], in fact its very elementary version given in Proposition 6.1 of [80] suffices. No other input from the technically involved paper [80] is used for the proof of (3.91). More precisely, for any fixed i, we have the bound

$$\frac{1}{N}\sum_{\substack{j=N+1 \\ j\neq i}}^{N+M} (1-\mathbf{E}_j)\mathfrak{G}_{ij}\mathfrak{G}'_{ji} = \mathcal{O}_{\prec}\left(\frac{1}{\sqrt{N\eta}}\frac{1}{\sqrt{N\eta'}}\left(\frac{1}{\sqrt{N\eta}} + \frac{1}{\sqrt{N\eta'}}\right)\right). \tag{3.91}$$

In particular, (3.91) shows that the operator $(1-\mathbf{E}_j)$ reduces the naive size of $\frac{1}{N}\sum_{i\neq j} \mathfrak{G}_{ij}\mathfrak{G}'_{ji}$ coming from (3.82) by an additional factor $1/\sqrt{N\eta} + 1/\sqrt{N\eta'}$. Indeed, by [80, Eq. (4.5)], the left hand side of (3.91) is exactly the left hand side of [80, Eq. 6.1] after the associations $\boldsymbol{a} = (i)$, $\boldsymbol{\mu} = (j)$, $w(\boldsymbol{a}) = w(i) = N^{-1}$, $F = \{j\}$ and Δ being the graph of degree $\deg(\Delta) = 2$ corresponding to $\mathfrak{G}_{ij}\mathfrak{G}'_{ji}$. Now we explain the single modification in the proof of Proposition 6.1 in [80] that leads to (3.91).

We recall that the main strategy in the proof of Proposition 6.1 in [80] is to compute the p-th moment of the sum $\sum_j (1-\mathbf{E}_j)\mathfrak{G}_{ij}\mathfrak{G}'_{ji}$. Expanding the p-th power yields a p-fold summation $\sum_{j_1,j_2,\dots,j_p}$. For any fixed choice of these indices, we successively expand the resolvent entries as much as possible, in order to create factors partially independent of each other using the resolvent expansion (3.90) for terms of the form $\mathfrak{G}_{ik}^{[T]}$, with $i, k \notin T$, and its analogues for $1/\mathfrak{G}_{ii}^{[T]}$ from [80, Eq. (3.13)]. Here the set T is a subset of the actual summation indices $j_1, j_2, \dots, j_p$. After taking the expectation and using that $\mathbf{E}(1-\mathbf{E}_j) = 0$, a simple power counting shows that only those terms remain nonzero that have many resolvent factors. Then, after that each factor is expanded as described above, we use the bound $|\mathfrak{G}_{ij}(z)| \leq 1/\sqrt{N\Im z}$, given by the local law in (3.82) for $i \neq j$. In particular, in the proof of Proposition 6.1 in [80] the resolvent expansions and the bounds given by the local law are used only for single resolvent entries. Hence, the proof of Proposition 6.1 [80] works verbatim for our case when different spectral parameter are considered, just in the estimates the different η's have to be carried. As a consequence, the error term in the r.h.s of (3.91), in contrast to its analogue in [80, Eq. (6.1)], contains both η and η', i.e. the error term is of the form $1/\sqrt{N^3\eta^2\eta'} + 1\sqrt{N^3\eta\eta'^2}$.

By (3.91), (3.88) and the local laws in (3.82) we get

$$\frac{1}{N}\sum_{\substack{i,j=N+1\\ j\neq i}}^{N+M}\mathfrak{G}_{ij}\mathfrak{G}'_{ji}=\frac{1}{N}m_2(\zeta)m_2(\zeta')\sum_{\substack{i,j=N+1\\ j\neq i}}^{N+M}\mathbf{E}_j\left(\sum_{\substack{k=2\\ k\neq j}}^{N+M}\mathfrak{G}^{[j]}_{ik}\gamma_{kj}\right)\left(\sum_{\substack{l=2\\ l\neq j}}^{N+M}\gamma_{jl}\mathfrak{G}'^{[j]}_{li}\right) \tag{3.92}$$

$$+\mathcal{O}_\prec\left((\eta+\eta')\Psi\right) \tag{3.93}$$

$$=\frac{1}{N\sqrt{MN}}m_2(\zeta)m_2(\zeta')\sum_{\substack{i,j=N+1\\ j\neq i}}^{N+M}\sum_{k=2}^{N}\mathfrak{G}^{[j]}_{ik}\mathfrak{G}'^{[j]}_{ki}+\mathcal{O}_\prec\left((\eta+\eta')\Psi\right). \tag{3.94}$$

Note that we used (3.83) and (3.85) to estimate the error terms. Using (3.90) the resolvent expansion in (3.88) and fluctuation averaging (3.91) again, (3.93) becomes

$$\begin{aligned}\frac{1}{N}\sum_{\substack{i,j=N+1\\ j\neq i}}^{N+M}\mathfrak{G}_{ij}\mathfrak{G}'_{ji}&=\frac{\sqrt{\phi}}{N}m_2(\zeta)m_2(\zeta')\sum_{i=N+1}^{N+M}\sum_{k=2}^{N}\mathfrak{G}_{ik}\mathfrak{G}'_{ki}+\mathcal{O}_\prec\left((\eta+\eta')\Psi\right)\\ &=\frac{\sqrt{\phi}}{N}m_2(\zeta)m_2(\zeta')\sum_{i=N+1}^{N+M}\sum_{k=2}^{N}\mathbf{E}_k\mathfrak{G}_{ik}\mathfrak{G}'_{ki}+\mathcal{O}_\prec\left((\eta+\eta')\Psi\right)\\ &=\frac{1}{N}m_1(\zeta)m_1(\zeta')m_2(\zeta)m_2(\zeta')\sum_{\substack{i,p=N+1\\ p\neq i}}^{N+M}\mathfrak{G}_{ip}\mathfrak{G}'_{pi}\\ &\quad+\phi m_1(\zeta)m_1(\zeta')m_2(\zeta)^2m_2(\zeta')^2+\mathcal{O}_\prec\left((\eta+\eta')\Psi\right).\end{aligned} \tag{3.95}$$

Solving this equation, we conclude that

$$\frac{1}{N}\sum_{\substack{i,j=N+1\\ i\neq j}}^{N+M}\mathfrak{G}_{ij}\mathfrak{G}'_{ji}=\frac{\phi m_1(\zeta)m_1(\zeta')m_2(\zeta)^2m_2(\zeta')^2}{1-m_1(\zeta)m_1(\zeta')m_2(\zeta)m_2(\zeta')}+\mathcal{O}_\prec(\Psi). \tag{3.96}$$

In estimating the error term we used a lower bound for the denominator. Indeed, using (3.83) and (3.85), we have that

$$|1-m_1(\zeta)m_1(\zeta')m_2(\zeta)m_2(\zeta')|\geq 1-|m_1(\zeta)m_1(\zeta')m_2(\zeta)m_2(\zeta')|\gtrsim(\eta+\eta'). \tag{3.97}$$

Notice that in the right hand side of (3.96) the deterministic term depends only on m_1 and m_2. Moreover, using the notation $\widehat{\mathfrak{G}}(\zeta):=(-\mathcal{H}-\zeta)^{-1}$ and that m_1 and m_2 are Stieltjes transforms of symmetric distributions, by (3.82) we have that

$$\left|\widehat{\mathfrak{G}}_{ij}(\zeta)-\delta_{ij}m_1(\zeta)\right|\prec\frac{1}{\sqrt{N\eta}},\quad i,j=2,\ldots,,N \tag{3.98}$$

$$\left|\widehat{\mathfrak{G}}_{ij}(\zeta)-\delta_{ij}m_2(\zeta)\right|\prec\frac{1}{\sqrt{N\eta}},\quad i,j=N+1,\ldots,N+M. \tag{3.99}$$

In (3.98) and (3.99) we used that $\Im\zeta \gtrsim \eta$. This means that the leading order deterministic term of each term in (3.89) is exactly the same. Hence, combining (3.89), (3.96) and (3.83) we conclude (3.76). The proof of (3.77) is analogous. □

Before proceeding, we recall that $f_{\mathbf{C}}(z)$ is supported in $[a,b]\times[-10,10]$, where $a = \gamma_- - 3$, $b = \gamma_+ + 3$ and γ_-, γ_+ are the spectral edges. Furthermore, we recall that, by (3.44), $w_\phi = -zm_\phi(z)m_{\phi^{-1}}(z)$, where $w_\phi(z)$ is the Stieltjes transform of the Wigner semicircle law centered at $\phi^{\frac{1}{2}} + \phi^{-\frac{1}{2}}$, hence $w_\phi(z)$ is a solution of the self consistent equation (3.43).

We now plug (3.75)–(3.77) into the integral in (3.72). Integrating the error terms in (3.75)–(3.77) and using that $|g(z)| \le C|z|^{\frac{1}{2}}$ (see (3.16) and (3.22)) we get an error term of the magnitude $N^{-\frac{7}{6}}$. The denominators in (3.76) and (3.77) are expanded into geometric series whose convergence follows from (3.83) and (3.97). Hence, using (3.44), we conclude that if $\sigma_2 = 0$ then (3.72) assumes the following form

$$F_N' = \frac{1}{2N}\Re \iint_a^b \iint_{\eta_0}^{10} \Big[g(z)g(\overline{z}')m_{\phi^{-1}}(z)m_{\phi^{-1}}(\overline{z}') \sum_{k\ge 1} \left[w_\phi(z)w_\phi(\overline{z}')\right]^k \tag{3.100}$$

$$- g(z)g(z')m_{\phi^{-1}}(z)m_{\phi^{-1}}(z') \sum_{k\ge 1} \left[w_\phi(z)w_\phi(z')\right]^k \Big] d\eta d\eta' dx dx' \tag{3.101}$$

$$+ \frac{\sigma_4 - 1}{N}\left(\Im \int_a^b \int_{\eta_0}^{10} g(z)m_{\phi^{-1}}(z)\, d\eta dx \right)^2 + \mathcal{O}_\prec\left(N^{-\frac{7}{6}}\right) \tag{3.102}$$

$$= \frac{1}{N}\sum_{k\ge 1}\left(\Im \int_a^b \int_{\eta_0}^{10} g(z)m_{\phi^{-1}}(z)w_\phi(z)^k \, d\eta dx \right)^2 \tag{3.103}$$

$$+ \frac{\sigma_4 - 1}{N}\left(\Im \int_a^b \int_{\eta_0}^{10} g(z)m_{\phi^{-1}}(z)\, d\eta dx \right)^2 + \mathcal{O}_\prec\left(N^{-\frac{7}{6}}\right). \tag{3.104}$$

$$\tag{3.105}$$

Substituting the expression of g (see (3.70)) in (3.103) we have

$$\begin{aligned} F_N' &= \frac{1}{N}\sum_{k\ge 2}\left(\frac{2}{\pi}\Im \int_a^b \int_{\eta_0}^{10} w_\phi(z)^k \partial_\eta \partial_{\overline{z}} f_{\mathbf{C}}(z)\, d\eta dx \right)^2 \\ &\quad + \frac{\sigma_4 - 1}{N}\left(\frac{2}{\pi}\Im \int_a^b \int_{\eta_0}^{10} w_\phi(z) \partial_\eta \partial_{\overline{z}} f_{\mathbf{C}}(z)\, d\eta dx \right)^2 + \mathcal{O}_\prec\left(N^{-\frac{7}{6}}\right). \end{aligned} \tag{3.106}$$

We start computing the last integral in (3.106):

$$\left(\frac{2}{\pi}\Im \int_a^b \int_{\eta_0}^{10} w_\phi(z) \partial_\eta \partial_{\overline{z}} f_{\mathbf{C}}(z)\, d\eta dx \right)^2 = \left(\frac{1}{\pi}\Im \int_a^b w_\phi(x) f'(x)\, dx \right)^2 + \mathcal{O}_\prec\left(N^{-\frac{1}{6}}\right),$$

where we used Lemma 3.4.4 and

$$\frac{\partial_\eta f_{\mathbf{C}}(z_0)}{i} = \partial_x f_{\mathbf{C}}(z_0) + \mathcal{O}(\eta_0) = f'(x) + \mathcal{O}_\prec(\eta_0), \tag{3.107}$$

where $z_0 = x + i\eta_0$. Furthermore, using Lemma 3.4.4 and (3.107) once more, we have

$$\frac{1}{N}\sum_{k\geq 2}\left(\frac{2}{\pi}\Im\int_a^b\int_{\eta_0}^{10} w_\phi(z)^k \partial_\eta\partial_{\overline{z}} f_{\mathbf{C}}(z)\, d\eta dx\right)^2 \tag{3.108}$$

$$= \frac{1}{N}\sum_{k\geq 2}\left(\frac{1}{\pi}\Im\int_a^b w_\phi(z_0)^k \frac{\partial_\eta f_{\mathbf{C}}(z_0)}{i}\, dx\right)^2 + \mathcal{O}_\prec\left(N^{-\frac{7}{6}}\right) \tag{3.109}$$

$$= \frac{1}{N}\sum_{k\geq 0}\left(\frac{1}{\pi}\Im\int_a^b w_\phi(z_0)^k \frac{\partial_\eta f_{\mathbf{C}}(z_0)}{i}\, dx\right)^2 - \frac{1}{N}\left(\frac{1}{\pi}\Im\int_a^b w_\phi(z_0)\frac{\partial_\eta f_{\mathbf{C}}(z_0)}{i}\, dx\right)^2 \tag{3.110}$$

$$- \frac{1}{N}\left(\frac{1}{\pi}\Im\int_a^b \frac{\partial_\eta f_{\mathbf{C}}(z_0)}{i}\, dx\right)^2 + \mathcal{O}_\prec\left(N^{-\frac{7}{6}}\right) \tag{3.111}$$

$$= \frac{1}{N}\sum_{k\geq 0}\left(\frac{1}{\pi}\Im\int_a^b w_\phi(z_0)^k \frac{\partial_\eta f_{\mathbf{C}}(z_0)}{i}\, dx\right)^2 - \left(\frac{1}{\pi}\Im\int_a^b w_\phi(x) f'(x)\, dx\right)^2 + \mathcal{O}_\prec\left(N^{-\frac{7}{6}}\right). \tag{3.112}$$

In the last equality we used that $\Im\frac{\partial_\eta f_{\mathbf{C}}(z_0)}{i} = \mathcal{O}_\prec(\eta_0)$ by (3.107). We want to use the same approximation in the first integral as well. However, the geometric series converges only slowly, so we need to ensure summability. The following lemma prepares us for that (see Lemma 3.7 in [89]).

Lemma 3.4.7. *There exists an N-independent constant $C > 0$ such that for $z_0 = x + i\eta_0$ and $z_0' = x' + i\eta_0$, with $0 < \eta_0 \leq \frac{1}{2}$, it holds*

$$\iint_a^b \frac{dxdx'}{|1 - w_\phi(z_0)w_\phi(\overline{z_0}')|} + \iint_a^b \frac{dxdx'}{|1 - w_\phi(z_0)w_\phi(z_0')|} \leq C|\log\eta_0|. \tag{3.113}$$

Combining (3.106)-(3.108) and Lemma 3.4.7, using (3.71) again, we conclude that

$$F_N' = \frac{1}{2N\pi^2}\Re\iint_a^b\left(\frac{1}{1 - w_\phi(z_0)w_\phi(\overline{z_0}')} - \frac{1}{1 - w_\phi(z_0)w_\phi(z_0')}\right) f'(x)f'(x')\, dxdx' \tag{3.114}$$

$$+ \frac{\sigma_4 - 2}{N}\left(\frac{1}{\pi}\Im\int_a^b w_\phi(x)f'(x)\, dx\right)^2 + \mathcal{O}\left(N^{-\frac{7}{6}}\right). \tag{3.115}$$

After some computations using (3.43) we have that

$$\begin{aligned}&\Re\left(\frac{1}{1 - w_\phi(z_0)w_\phi(\overline{z_0}')} - \frac{1}{1 - w_\phi(z_0)w_\phi(z_0')}\right)\\ &\quad = \Re\left(\frac{2i\Im w_\phi(z_0')}{\phi^{\frac{1}{2}} + \phi^{-\frac{1}{2}} - z_0 - 2\Re w_\phi(z_0') - w_\phi(z_0')(|w_\phi(z_0')|^2 - 1)}\right).\end{aligned} \tag{3.116}$$

For small η_0 and (x, x') outside the square $[\gamma_-, \gamma_+]^2$ the integral of (3.116) is negligible. Indeed, outside $[\gamma_-, \gamma_+]^2$ we have that $1 - |w_\phi(z)|^2 \asymp \sqrt{\kappa_x + \eta}$ by Lemma 3.6 in [31], where $\kappa_x = \min\{|\gamma_+ - x|, |\gamma_- - x|\}$.

For $(x, x') \in [\gamma_-, \gamma_+]^2$ and small η_0 we have

$$\Re\left(\frac{2i\Im w_\phi(z_0')}{\phi^{\frac{1}{2}} + \phi^{-\frac{1}{2}} - z_0 - 2\Re w_\phi(z_0') - w_\phi(z_0')(|w_\phi(z_0')|^2 - 1)}\right) = \frac{\eta_0\sqrt{(x' - \gamma_-)(\gamma_+ - x')}}{(x - x')^2 + \eta_0^2} + \mathcal{O}_\prec(\eta_0). \tag{3.117}$$

The expression $\frac{\eta_0}{(x-x')^2+\eta_0^2}$ acts like $\pi\delta(x' - x)$ for small η_0, hence for each $h \in L^2$

$$\lim_{\eta\to 0}\int_{\mathbf{R}} \frac{\eta}{(x - x')^2 + \eta^2} h(x')\, dx' = \pi h(x)$$

in L^2-sense. Working out an effective error term for $h \in H^1$ and using the explicit expression in (3.117), by (3.115), we conclude that

$$F_N' = \frac{1}{2\pi N}\int_{\gamma_-}^{\gamma^+} f'(x)^2\sqrt{(x - \gamma_-)(\gamma_+ - x)}\, dx + \frac{\sigma_4 - 2}{N}\left(\frac{1}{\pi}\int_{\gamma_-}^{\gamma_+} \frac{1}{2} f'(x)\sqrt{(x - \gamma_-)(\gamma_+ - x)}\, dx\right)^2 + \mathcal{O}\left(N^{-\frac{7}{6}}\right).$$

This computation gives the explicit expression of V_f in (3.10) for $\sigma_2 = 0$.

When $\sigma_2 \neq 0$ we have to consider (3.77) and so, using a similar analysis, we have to add the following term in the expression of F_N' in (3.103)

$$\frac{1}{2N\pi^2}\Re\iint_{\mathbf{R}} f'(x)f'(x')\left(\frac{|\sigma_2|^2 w_\phi(z_0)^2 w_\phi(\overline{z_0}')^2}{1 - |\sigma_2|^2 w_\phi(z_0)w_\phi(\overline{z_0}')} - \frac{|\sigma_2|^2 w_\phi(z_0)^2 w_\phi(z_0')^2}{1 - |\sigma_2|^2 w_\phi(z_0)w_\phi(z_0')}\right) dxdx'. \tag{3.118}$$

For the special case $|\sigma_2| = 1$ the expressions in (3.76) and (3.77) are exactly the same, hence we define $V_{\sigma_2} := V_{f,1}$. This holds true in particular for the case $X \in \mathbf{R}^{M\times(N-1)}$ when $\sigma_2 = 1$ automatically.

If $|\sigma_2| < 1$, instead, we define V_{σ_2} in the following way

$$V_{\sigma_2} := \frac{1}{2\pi^2}\Re\iint_{\mathbf{R}} f'(x)f'(x')\left(\frac{|\sigma_2|^2 w_\phi(x)^2\overline{w_\phi(x')}^2}{1 - |\sigma_2|^2 w_\phi(x)\overline{w_\phi(x')}} - \frac{|\sigma_2|^2 w_\phi(x)^2 w_\phi(x')^2}{1 - |\sigma_2|^2 w_\phi(x)w_\phi(x')}\right) dxdx', \tag{3.119}$$

that is close to (3.118) by an $\mathcal{O}(\eta_0)$ error using that $|w_\phi(z_0) - w_\phi(x)| \lesssim \eta_0[(x - \gamma_-)(\gamma_+ - x)]^{-\frac{1}{2}}$ and $|1 - |\sigma_2|^2 w_\phi(x)\overline{w_\phi(x')}| \geq 1 - |\sigma_2|^2$. Notice that from (3.119) easily follows that $V_{\sigma_2} \geq 0$. Indeed

$$V_{\sigma_2} = \sum_{k\geq 0}\left(\frac{1}{\pi}\Im\int_a^b f'(x)(|\sigma_2| w_\phi(x))^{k+2}\, dx\right)^2.$$

3.5 Computation of the higher order moments of F_N

In this section we compute the higher order moments of

$$F_N = -\frac{1}{\sqrt{N}}\Im\int_{\mathbf{R}}\int_{\eta_0}^{10} g(z)A(z)\, d\eta dx + \mathcal{O}_\prec(\eta_0),$$

where $g(z)$ and $A(z)$ are defined in (3.70). We remark that for the proof of the normality of F_N it would be sufficient to show that the quadratic form $\langle \boldsymbol{x}, G(z)\boldsymbol{x}\rangle$ has a Gaussian fluctuation conditioned on G and then separately show that the quadratic variation of G is negligible. Here we follow a more robust path that gives an effective control on all higher moments as well without essentially no extra effort since the fluctuation averaging mechanism used already in the proof of Lemma 3.4.6 directly extends to higher moments. Thus, using a similar approach to the one we used to compute the variance of F_N, we start computing

$$\mathbf{E}[A(z_1)\dots A(z_k)]$$

for any $k \in \mathbf{N}$ and $z_l \in \mathbf{C}\setminus\mathbf{R}$, with $l = 1,\dots,k$. We recall that $\mathbf{E}_1 := \mathbf{E}(\cdot|X)$ is the conditional expectation conditioned on the matrix X. This leads to products of cyclic expressions of the form $G_{j_1j_2}G_{j_2j_3}\dots G_{j_{k-1}j_k}$.

Notation. A multiple summation with a star $\sum^*_{j_1,\dots,j_k}$ indicates that the sum is performed over distinct indices.

In the following we prove that the leading order term of the k-th moment of F_N is given by cycles of length two, hence cyclic products with at least three terms are actually of lower order:

Lemma 3.5.1. *For closed cycles of length $k > 2$ we have that*

$$N^{-\frac{k}{2}} \sum_{j_1,\dots,j_k=1}^{M\,*} \mathbf{E}_{j_1+N}\left(G^{(1)}_{j_1j_2}\dots G^{(k-1)}_{j_{k-1}j_k}G^{(k)}_{j_kj_1}\right) \prec \frac{|z_1\dots z_k|^{-\frac{1}{2}}}{(\max_a \eta_a)\sqrt{N\eta_1\dots\eta_k}}\sum_{a=1}^{k}\frac{1}{\sqrt{\eta_a}}, \tag{3.120}$$

and for open cycles of any length $k > 1$ we have that

$$N^{-\frac{k+1}{2}} \sum_{j_1,\dots,j_k=1}^{M\,*} \mathbf{E}_{j_1+N}\left(G^{(1)}_{j_1j_2}\dots G^{(k-1)}_{j_{k-1}j_k}\right) \prec \frac{|z_1\dots z_k|^{-\frac{1}{2}}}{\sqrt{N\eta_1\dots\eta_{k-1}}}\sum_{a=1}^{k}\frac{1}{\sqrt{\eta_a}}, \tag{3.121}$$

where $G^{(l)} := G(z_l)$, $z_l \in \mathbf{C}\setminus\mathbf{R}$ with $\eta_l = |\Im z_l|$ for $l = 1,\dots,k$ and $\mathbf{E}_{j_1+N} := \mathbf{E}(\cdot|\mathcal{H}^{[j_1+N]})$, with $\mathcal{H}^{[j_1+N]}$ defined in (3.87). Moreover, the same bounds hold true when any of the $G^{(l)}$ are replaced by their transposes or Hermitian conjugates.

Proof. The proof is similar to the proof of Lemma 4.1 in [89], so we will skip some details. However, an additional step in needed, see (3.129) later.

We start proving (3.120) for the case $X \in \mathbf{R}^{M\times(N-1)}$. We will actually prove that

$$N^{-\frac{k}{2}} \sum_{j_1,\dots,j_k=1}^{M\,*} \mathbf{E}_{j_1+N}(G^{(1)}_{j_1j_2}\dots G^{(k-1)}_{j_{k-1}j_k}G^{(k)}_{j_kj_1}) \lesssim \frac{N^\epsilon|z_1\dots z_k|^{-\frac{1}{2}}}{(\eta_1+\eta_k)\sqrt{N\eta_1,\dots,\eta_k}}\sum_{a=1}^{k}\frac{1}{\sqrt{\eta_a}},$$

for any $\epsilon > 0$, which implies (3.120) by the definition of $\prec$ in Definition 3.4.2.

We use linearization again to express the resolvents $G^{(1)},\dots,G^{(k)}$ of the matrix XX^* in terms of the resolvents $\mathfrak{G}^{(1)},\dots,\mathfrak{G}^{(k)}$ of the linearized matrix $\mathcal{H}$.

$$\sum_{j_1,\dots,j_k=1}^{M\,*} \mathbf{E}_{j_1+N}\left(G^{(1)}_{j_1j_2}\dots G^{(k-1)}_{j_{k-1}j_k}G^{(k)}_{j_kj_1}\right) = \sum_{i_1,\dots,i_k=N+1}^{N+M\,*} \mathbf{E}_{i_1}\left(\mathcal{G}^{(1)}_{i_1i_2}\dots \mathcal{G}^{(k-1)}_{i_{k-1}i_k}\mathcal{G}^{(k)}_{i_ki_1}\right), \tag{3.122}$$

where $\mathcal{G}^{(l)} = (\mathcal{H}^2 - z_l)^{-1}$, $i_m = j_m + N$. We write each $\mathcal{G}^{(l)}$ in the r.h.s. of (3.122) as

$$\mathcal{G}(z_l) = \frac{1}{2\zeta_l} \cdot (\mathfrak{G}(\zeta_l) - \mathfrak{G}(-\zeta_l)), \tag{3.123}$$

with $\zeta_l^2 = z_l$ (see (3.80)). We have to find a self consistent equation for each term in the right-hand side of (3.122) after rewriting it using (3.123). We start with

$$N^{-\frac{k}{2}} \sum_{i_1,\ldots,i_k=N+1}^{N+M\,*} \mathbf{E}_{i_1} \left(\mathfrak{G}^{(1)}_{i_1 i_2} \ldots \mathfrak{G}^{(k-1)}_{i_{k-1} i_k} \mathfrak{G}^{(k)}_{i_k i_1} \right).$$

Using the resolvent identity $\mathfrak{G}^{(1)} = \frac{1}{\zeta_1}[\mathcal{H}^{(1)}\mathfrak{G}^{(1)} - 1]$ we get

$$\begin{aligned} N^{-\frac{k}{2}} &\sum_{i_1,\ldots,i_k=N+1}^{N+M\,*} \mathbf{E}_{i_1} \left(\mathfrak{G}^{(1)}_{i_1 i_2} \ldots \mathfrak{G}^{(k-1)}_{i_{k-1} i_k} \mathfrak{G}^{(k)}_{i_k i_1} \right) \\ &= \frac{1}{N^{\frac{k}{2}}\zeta_1} \sum_{i_1,\ldots,i_k=N+1}^{N+M\,*} \sum_{n=2}^{N+M} \mathbf{E}_{i_1} \left(\gamma_{i_1 n} \mathfrak{G}^{(1)}_{n i_2} \ldots \mathfrak{G}^{(k-1)}_{i_{k-1} i_k} \mathfrak{G}^{(k)}_{i_k i_1} \right), \end{aligned} \tag{3.124}$$

where γ_{ij}, with $i, j \in \{2, \ldots, N+M\}$, are the entries of the big matrix $\mathcal{H}$.

We use the standard cumulant expansion

$$\begin{aligned} \mathbf{E} h f(h) = \mathbf{E} h \mathbf{E} f(h) + \mathbf{E} h^2 \mathbf{E} f'(h) + \mathcal{O}\left(\mathbf{E} \left| h^3 \mathbf{1}(|h| > N^{\tau - \frac{1}{2}}) \right| \|f''\|_\infty \right) \\ + \mathcal{O}\Big(\mathbf{E}|h|^3 \sup_{|x| \le N^{\tau - \frac{1}{2}}} |f''(x)| \Big), \end{aligned} \tag{3.125}$$

where f is any smooth function of a real random variable h, such that the expectations exist and $\tau > 0$ is arbitrary (see [124]). This yields

$$\mathbf{E}_{i_1} \left(\gamma_{i_1 n} \mathfrak{G}^{(1)}_{n i_2} \ldots \mathfrak{G}^{(k-1)}_{i_{k-1} i_k} \mathfrak{G}^{(k)}_{i_k i_1} \right) = \frac{1}{\sqrt{MN}} \mathbf{E}_{i_1} \left(\frac{\partial \mathfrak{G}^{(1)}_{n i_2}}{\partial \gamma_{i_1 n}} \mathfrak{G}^{(2)}_{i_2 i_3} \ldots \mathfrak{G}^{(k)}_{i_k i_1} \right) \tag{3.126}$$

$$+ \frac{1}{\sqrt{MN}} \sum_{a=2}^{k} \mathbf{E}_{i_1} \left(\frac{\partial \mathfrak{G}^{(a)}_{i_a i_{a+1}}}{\partial \gamma_{i_1 n}} \mathfrak{G}^{(1)}_{n i_2} \prod_{a \neq b = 2}^{k} \mathfrak{G}^{(b)}_{i_b i_b + 1} \right) + R, \tag{3.127}$$

where $i_{k+1} = i_1$ and R is the error term resulting from the cumulant expansion.

Using the expression for the derivative of the resolvent

$$\frac{\partial \mathfrak{G}_{ij}}{\partial \gamma_{kl}} = -\frac{\mathfrak{G}_{ik}\mathfrak{G}_{lj} + \mathfrak{G}_{il}\mathfrak{G}_{kj}}{1 + \delta_{kl}}$$

and the local law by (3.82) for the resolvent of the Gram matrix $\mathcal{H}$, summing over n, the first term of the right hand side of (3.126) becomes

$$\begin{aligned} -\frac{1}{\sqrt{MN}} &\sum_{n=2}^{N} \left(\mathfrak{G}^{(1)}_{n i_1} \mathfrak{G}^{(1)}_{n i_2} + \mathfrak{G}^{(1)}_{nn} \mathfrak{G}^{(1)}_{i_1 i_2} \right) \mathfrak{G}^{(2)}_{i_2 i_3} \ldots \mathfrak{G}^{(k)}_{i_k i_1} \\ &= -\phi^{-\frac{1}{2}} m_1(\zeta_1) \mathfrak{G}^{(1)}_{i_1 i_2} \ldots \mathfrak{G}^{(k)}_{i_k i_1} + \mathcal{O}_\prec \left(\frac{1}{N^{\frac{k}{2}+\frac{1}{2}} \sqrt{\eta \eta_1}} \right), \end{aligned} \tag{3.128}$$

with $n \neq i_1, i_2$ and $\eta := \eta_1 \ldots \eta_k$. If n is equal to i_1 or i_2 we use the trivial bound.
Using the same computations of Lemma 4.1 in [89], if $a \neq k$ the second term of the right-hand side of (3.126) can be estimated by

$$-\left(\mathfrak{G}^{(a)}_{i_a i_1}\mathfrak{G}^{(a)}_{n i_{a+1}} + \mathfrak{G}^{(a)}_{i_a n}\mathfrak{G}^{(a)}_{i_1 i_{a+1}}\right)\mathfrak{G}^{(1)}_{n i_2}\prod_{a\neq b=2}^{k}\mathfrak{G}^{(b)}_{i_b i_{b+1}} \prec \frac{1}{N^{\frac{k}{2}}\sqrt{\eta\eta_a}}$$

and if $n \notin \{i_1, \ldots, i_k\}$ this bound can be improved to

$$-\left(\mathfrak{G}^{(a)}_{i_a i_1}\mathfrak{G}^{(a)}_{n i_{a+1}} + \mathfrak{G}^{(a)}_{i_a n}\mathfrak{G}^{(a)}_{i_1 i_{a+1}}\right)\mathfrak{G}^{(1)}_{n i_2}\prod_{a\neq b=2}^{k}\mathfrak{G}^{(b)}_{i_b i_{b+1}} \prec \frac{1}{N^{\frac{k}{2}+\frac{1}{2}}\sqrt{\eta\eta_a}}.$$

Finally, for the case $a = k$ we have

$$-\left(\mathfrak{G}^{(k)}_{i_k i_1}\mathfrak{G}^{(k)}_{n i_1} + \mathfrak{G}^{(k)}_{i_k n}\mathfrak{G}^{(k)}_{i_1 i_1}\right)\mathfrak{G}^{(1)}_{n i_2}\ldots\mathfrak{G}^{(k-1)}_{i_{k-1} i_k}.$$

Here an additional argument is needed compared to [89]. To get a similar expression to (3.128) we need to have that all the indices of the resolvents in the previous expression are in the set $\{N+1, \ldots, N+M\}$, but this is not the case since $n \in \{2, \ldots, N\}$. Hence using a fluctuation averging for $\sum_{n=2}^{N}\mathfrak{G}^{(k)}_{i_k n}\mathfrak{G}^{(1)}_{n i_2}$ and the one side resolvent expansion in (3.88) as in (3.95) in the proof of Lemma 3.4.6 we get

$$-\frac{1}{\sqrt{MN}}\sum_{n=2}^{N} -\left(\mathfrak{G}^{(k)}_{i_k i_1}\mathfrak{G}^{(k)}_{n i_1} + \mathfrak{G}^{(k)}_{i_k n}\mathfrak{G}^{(k)}_{i_1 i_1}\right)\mathfrak{G}^{(1)}_{n i_2}\ldots\mathfrak{G}^{(k-1)}_{i_{k-1} i_k} \tag{3.129}$$

$$= -m_1(\zeta_1)m_1(\zeta_k)m_2(\zeta_k)\sum_{m=N+1}^{N+M}\mathfrak{G}^{(1)}_{m i_2}\ldots\mathfrak{G}^{(k)}_{i_k m} + \mathcal{O}_\prec\left(\frac{1}{N^{\frac{k}{2}+\frac{1}{2}}\sqrt{\eta\eta_k}}\right). \tag{3.130}$$

Furthermore, following the proof of Lemma 4.1 in [89] for the estimate of the error we obtain that

$$R \prec \sum_{a=1}^{k}\frac{N^\epsilon}{\sqrt{N\eta\eta_a}}. \tag{3.131}$$

Hence, using $z_l = \zeta_l^2$ for $l = 1, \ldots, k$, combining (3.122) and (3.126)-(3.131) we conclude

$$\begin{aligned} N^{-\frac{k}{2}}&\sum_{i_1,\ldots,i_k=N+1}^{N+M *}\mathbf{E}_{i_1}\left(\mathfrak{G}^{(1)}_{i_1 i_2}\ldots\mathfrak{G}^{(k-1)}_{i_{k-1} i_k}\mathfrak{G}^{(k)}_{i_k i_1}\right) \\ &= \frac{m_2(\zeta_1)}{m_1(\zeta_1)m_1(\zeta_k)m_2(\zeta_1)m_2(\zeta_k)-1}\cdot\mathcal{O}_\prec\left(\sum_{a=1}^{k}\frac{N^\epsilon}{\sqrt{N\eta\eta_a}}\right) \\ &= \mathcal{O}_\prec\left(\sum_{a=1}^{k}\frac{N^\epsilon}{(\eta_1+\eta_k)\sqrt{N\eta\eta_a}}\right), \end{aligned} \tag{3.132}$$

where in the last equality we used (3.97) and, since (3.16) holds true also substituting ϕ with ϕ^{-1} (see proof of Lemma 3.3.2 in Appendix A), that $|m_2| \leq \phi^{-\frac{1}{4}} \leq 1$ to estimate the error. With these computations we conclude the estimate of the first term in the right-hand side of (3.122). Notice that the estimate of the error in (3.132) depends only on the Stieltjes transforms m_1 and m_2, hence, using a similar argument as in the proof of Lemma 3.4.6, we conclude that all the terms in the right-hand side of (3.122) give the same contribution. This concludes the proof of (3.120).

The proof of (3.121), using the equality in (3.123), is exactly the same of (3.120) using that for the case $a = k-1$ we have the following estimate

$$-\left(\mathfrak{G}^{(k-1)}_{i_{k-1}i_1}\mathfrak{G}^{(k-1)}_{ni_k} + \mathfrak{G}^{(k-1)}_{i_kn}\mathfrak{G}^{(k-1)}_{i_1i_k}\right)\mathfrak{G}^{(1)}_{ni_2}\dots\mathfrak{G}^{(k-2)}_{i_{k-2}i_{k-1}} \prec \frac{1}{N^{\frac{k}{2}}\sqrt{\eta\eta_{k-1}}}.$$

Hence we have that

$$N^{-\frac{k+1}{2}}\sum_{i_1,\dots,i_k=N+1}^{N+M}\mathbf{E}_{i_1}\left(\mathfrak{G}^{(1)}_{i_1i_2}\dots\mathfrak{G}^{(k-1)}_{i_{k-1}i_k}\right) = \mathcal{O}_\prec\left(\sum_{a=1}^{k}\frac{N^\epsilon}{\sqrt{N\eta\eta_a}}\right).$$

The previous expression only depends on m_2 and so using the same argument as before we conclude the proof of (3.121).

The proof for $X \in \mathbf{C}^{M\times(N-1)}$ is omitted since is similar to the real case after replacing the cumulant expansion by its complex variant (Lemma 7.1 in [112]). □

Notice that the estimates of Lemma 3.5.1 hold also without the expectation:

Corollary 3.5.2. *Under the hypotheses of Lemma 3.5.1, we have that for closed cycles of length* $k > 2$

$$N^{-\frac{k}{2}}\sum_{j_1,\dots,j_k=1}^{M\ *}G^{(1)}_{j_1j_2}\dots G^{(k-1)}_{j_{k-1}j_k}G^{(k)}_{j_kj_1} \prec \frac{|z_1\dots z_k|^{-\frac{1}{2}}}{(\max_a\eta_a)\sqrt{N\eta_1\dots\eta_k}}\sum_{a=1}^{k}\frac{1}{\sqrt{\eta_a}}, \tag{3.133}$$

and for open cycles of length $k > 1$

$$N^{-\frac{k+1}{2}}\sum_{j_1,\dots,j_k=1}^{M\ *}G^{(1)}_{j_1j_2}\dots G^{(k-1)}_{j_{k-1}j_k} \prec \frac{|z_1\dots z_k|^{-\frac{1}{2}}}{\sqrt{N\eta_1\dots\eta_{k-1}}}\sum_{a=1}^{k}\frac{1}{\sqrt{\eta_a}} \tag{3.134}$$

Proof. First, we recall that $\mathfrak{G}(z)$, $z \in \mathbf{C}\setminus\mathbf{R}$, is the resolvent of the linearized matrix $\mathcal{H}$. In order to prove the bounds (3.133)–(3.134), we rely on [80, Proposition 6.1] with exactly the same modification as in the proof of (3.91), i.e. the case when different resolvent factors $\mathfrak{G}$ may have different spectral parameters. In particular, for any fixed and distinct $i_2,\dots,i_k$, the quantity

$$\frac{1}{N}\sum_{i_1=N+1}^{N+M\ *}(1-\mathbf{E}_{i_1})\mathfrak{G}^{(1)}_{i_1i_2}\dots\mathfrak{G}^{(k-1)}_{i_{k-1}i_k}\mathfrak{G}^{(k)}_{i_ki_1}, \tag{3.135}$$

is smaller than the bound given by the local law of an additional factor $1/\sqrt{N\eta_1}+\dots+1/\sqrt{N\eta_k}$. Hence, the bounds in (3.133) and (3.134) follow by Lemma 3.5.1, using the relation (3.123) and that $G_{ij} = \mathcal{G}_{N+i,N+j}$ for $i,j = 1,\dots,M$. □

The following lemma shows that the leading order terms of $\mathbf{E}_1A(z_1)\dots A(z_k)$ are the cycles of length two (see the proof of Lemma 4.3 in [89]).

Lemma 3.5.3. *For each $k \geq 2$ and $z_1, \dots, z_k \in \mathbf{C}$ with $|\Im z_l| = \eta_l > 0$ we have that*

$$\begin{aligned}\mathbf{E}_1 A(z_1)\dots A(z_k) = &\sum_{\pi \in P_2([k])} \prod_{\{a,b\}\in\pi} \mathbf{E}_1(A(z_a)A(z_b)) \\ &+ \mathcal{O}_\prec\left(\frac{|z_1\dots z_k|^{-\frac{1}{2}}}{\sqrt{N}\eta_1\dots\eta_k}\sum_{a\neq b}\frac{1}{(\eta_a+\eta_b)\sqrt{\eta_a}}\right),\end{aligned} \tag{3.136}$$

where $[k] := \{1, \dots, k\}$ and $P_2(L)$ is the set of pairings of the set L.

By Lemma 3.5.3 we conclude that

$$\begin{aligned}\mathbf{E}\left[-\Im\int_{\mathbf{R}}\int_{\eta_0}^{10} g(z)A(z)\,d\eta dx\right]^k &= \sum_{\pi\in P_2([k])}(2V_{f,1}+(\sigma_4-1)V_{f,2})^{\frac{k}{2}} + \mathcal{O}_\prec\left((N^{-\frac{7}{6}}\right) \\ &= (k-1)!!(2V_{f,1}+(\sigma_4-1)V_{f,2})^{\frac{k}{2}} + \mathcal{O}_\prec\left(N^{-\frac{7}{6}}\right),\end{aligned} \tag{3.137}$$

if k is even and

$$\mathbf{E}\left[-\Im\int_{\mathbf{R}}\int_{\eta_0}^{10} g(z)A(z)\,d\eta dx\right]^k = \mathcal{O}_\prec\left(N^{-\frac{7}{6}}\right) \tag{3.138}$$

if k is odd. If $X \in \mathbf{C}^{M\times(N-1)}$, following the same argument, we find

$$\mathbf{E}\left[-\Im\int_{\mathbf{R}}\int_{\eta_0}^{10} g(z)A(z)\,d\eta dx\right]^k = (k-1)!!(V_{f,1}+|\sigma_2|^2V_{\sigma_2}+(\sigma_4-1)V_{f,2})^{\frac{k}{2}}+\mathcal{O}_\prec\left(N^{-\frac{7}{6}}\right).$$

In this way we conclude the computations of the moments for each $k \geq 1$ and so with this result we have shown that the random variable $\sqrt{N}(f_N - \Omega_f)$ converges in distribution to a Gaussian random variable Δ_f with mean zero and variance V_f and that any fixed moment of $\sqrt{N}(f_N - \Omega_f)$ converges to the corresponding Gaussian moment with overwhelming probability at least at a rate $\mathcal{O}\left(N^{-\frac{1}{6}+\epsilon}\right)$.

3.A Proof of Lemma 3.3.2.

We recall that $w_\phi(z)$ is the Stieltjes transform of the Wigner semicircle law centered in $\phi^{\frac{1}{2}} + \phi^{-\frac{1}{2}}$ defined as in (3.42). By the proof of Lemma 3.7 in [89] and Lemma 3.6 in [31], for each $z = x + i\eta$ such that $|z - \sqrt{\phi}| \leq 10$, we have that

$$c \leq |w_\phi(z)| \leq 1, \quad |1 - w_\phi(z)^2| \asymp \sqrt{\kappa_x + \eta}, \quad \Im w_\phi(z) \asymp \begin{cases}\sqrt{\kappa_x+\eta} & \text{if } x \in [\gamma_-, \gamma_+] \\ \frac{\eta}{\sqrt{\kappa_x+\eta}} & \text{if } x \notin [\gamma_-, \gamma_+],\end{cases} \tag{3.139}$$

where $\kappa_x = \min\{|\gamma_+ - x|, |\gamma_- - x|\}$, $w_\phi(z) := \sqrt{\phi}(1 + z m_{\phi^{-1}}(z))$ and $c > 0$ is a constant independent of ϕ.

Proof of Lemma 3.3.2. Let $\tilde{z} := z\phi^{-\frac{1}{2}}$, taking the imaginary part of $-\frac{1}{m_\phi} = z + \tilde{z}m_\phi - (\phi^{\frac{1}{2}} - \phi^{-\frac{1}{2}})$ and $-\frac{1}{\tilde{z}m_\phi} = \phi^{\frac{1}{2}} + m_\phi - \frac{1}{\tilde{z}}(\phi^{\frac{1}{2}} - \phi^{-\frac{1}{2}})$ (see (3.7)), we get

$$\frac{\Im m_\phi}{|m_\phi|^2} = \eta + \Im(\tilde{z}m_\phi), \qquad \frac{\Im(\tilde{z}m_\phi)}{|\tilde{z}m_\phi|^2} = \Im m_\phi + \frac{\phi-1}{|z|^2}\eta. \tag{3.140}$$

Combining these equalities we obtain

$$|\tilde{z}|^2|m_\phi|^4 = 1 - \frac{|m_\phi|^2 + \frac{\phi-1}{|z|^2}}{\Im m_\phi + \frac{\eta(\phi-1)}{|z|^2}}\eta.$$

By our hypotheses $|z-\sqrt{\phi}| \leq 10$ and $\phi \geq 1$, we have that $\eta \leq 10$ and that there exists a constant $d > 0$ independent of ϕ such that $|z| \leq d\sqrt{\phi}$. Furthermore, from (3.140) and $\Im(\tilde{z}m_\phi) = \Im w_{\phi^{-1}} \leq 1$ we have $\Im m_\phi \leq C|m_\phi|^2$, with $C > 0$ some constant independent of ϕ. We conclude that

$$|\tilde{z}|^2|m_\phi|^4 = 1 - \frac{|m_\phi|^2 + \frac{\phi-1}{|z|^2}}{\Im m_\phi + \frac{\eta(\phi-1)}{|z|^2}}\eta \leq 1 - 2\tilde{c}\eta,$$

for any $\phi \geq 1$. The above inequality proves the bound in (3.16).

Furthermore, since $w_\phi(z) = -zm_\phi(z)m_{\phi^{-1}}(z)$ by (3.44) and using that, by similar computations substituting ϕ with ϕ^{-1}, we have an upper bound as in (3.16) for $|m_{\phi^{-1}}|$ and that $|w_\phi| \geq c$ from (3.139), we also obtain the lower bound in (3.16). Note that by a direct computation, substituting ϕ with ϕ^{-1}, we get a lower bound as in (3.16) also for $|m_{\phi^{-1}}|$. Finally, since

$$1 - w_\phi^2(z) = 1 - w_\phi(z)w_{\phi^{-1}}(z) = zm_\phi(z) + zm_{\phi^{-1}}(z) + z^2 m_\phi(z)m_{\phi^{-1}}(z),$$

using (3.84) for $zm_{\phi^{-1}}(z)$ in the right-hand side, we get

$$\left|1 - z\phi^{-\frac{1}{2}}m_\phi(z)^2\right| = \frac{|1-w_\phi^2(z)|}{|zm_{\phi^{-1}}(z)|}.$$

Hence, using (3.139) and that $|m_{\phi^{-1}}| \geq c\phi^{-\frac{1}{4}}|z|^{-\frac{1}{2}}$, we conclude

$$\left|1 - z\phi^{-\frac{1}{2}}m_\phi(z)^2\right| \asymp \frac{\phi^{\frac{1}{4}}}{|z|^{\frac{1}{2}}}\sqrt{\kappa_x + \eta}. \tag{3.141}$$

This proves (3.18). Then, using (3.16), (3.141) and the explicit expression

$$m_\phi(z)' = \frac{m_\phi(z)^2 + \frac{m_\phi(z)^3}{\sqrt{\phi}}}{1 - \frac{z}{\sqrt{\phi}}m_\phi(z)^2},$$

obtained differentiating (3.7), we also get the bound in (3.17) for $|m_\phi(z)'|$.

Cusp Universality for Random Matrices II: The Real Symmetric Case 4

We prove that the local eigenvalue statistics of real symmetric Wigner-type matrices near the cusp points of the eigenvalue density are universal. Together with the companion paper [83], which proves the same result for the complex Hermitian symmetry class, this completes the last remaining case of the Wigner-Dyson-Mehta universality conjecture after bulk and edge universalities have been established in the last years. We extend the recent Dyson Brownian motion analysis at the edge [131] to the cusp regime using the optimal local law from [83] and the accurate local shape analysis of the density from [7, 14]. We also present a novel PDE-based method to improve the estimate on eigenvalue rigidity via the maximum principle of the heat flow related to the Dyson Brownian motion.

Published as G. Cipolloni et al., *Cusp universality for random matrices, II: The real symmetric case*, Pure Appl. Anal. **1**, 615–707 (2019), MR4026551.

4.1 Introduction

We consider *Wigner-type* matrices, i.e. $N \times N$ Hermitian random matrices H with independent, not necessarily identically distributed entries above the diagonal; a natural generalization of the standard Wigner ensembles that have i.i.d. entries. The Wigner-Dyson-Mehta (WDM) conjecture asserts that the local eigenvalue statistics are universal, i.e. they are independent of the details of the ensemble and depend only on the *symmetry type*, i.e. on whether H is real symmetric or complex Hermitian. Moreover, different statistics emerge in the bulk of the spectrum and at the spectral edges with a square root vanishing behavior of the eigenvalue density. The WDM conjecture for both symmetry classes has been proven for Wigner matrices, see [90] for complete historical references. Recently it has been extended to more general ensembles including Wigner-type matrices in the bulk and edge regimes; we refer to the companion paper [83] for up to date references.

The key tool for the recent proofs of the WDM conjecture is the Dyson Brownian motion (DBM), a system of coupled stochastic differential equations. The DBM method has evolved during the last years. The original version, presented in the monograph [90], was in the spirit of a high dimensional analysis of a strongly correlated Gibbs measure and its dynamics. Starting in [91] with the analysis of the underlying parabolic equation and its short range approximation, the PDE component of the theory became prominent. With the coupling idea, introduced in [39, 42], the essential part of the proofs became fully deterministic, greatly simplifying the technical aspects. In the current paper we extend this trend and use PDE methods even for the proof of the rigidity bound, a key technical input, that earlier was obtained with direct random matrix methods.

The historical focus on the bulk and edge universalities has been motivated by the Wigner ensemble since, apart from the natural bulk regime, its semicircle density vanishes as a square root near the edges, giving rise to the Tracy-Widom statistics. Beyond the Wigner ensemble, however, the density profile shows a much richer structure. Already Wigner matrices with nonzero expectation on the diagonal, also called *deformed Wigner ensemble*, may have a density supported on several intervals and a cubic root cusp singularity in the density arises whenever two such intervals touch each other as some deformation parameter varies. Since local spectral universality is ultimately determined by the local behavior of the density near its vanishing points, the appearance of the cusp gives rise to a new type of universality. This was first observed in [50] and the local eigenvalue statistics at the cusp can be explicitly described by the Pearcey process in the complex Hermitian case [204]. The corresponding explicit formulas for the real symmetric case have not yet been established.

The key classification theorem [4] for the density of Wigner-type matrices showed that the density may vanish only as a square root (at regular edges) or as a cubic root (at cusps); no other singularity may occur. This result has recently been extended to a large class of matrices with correlated entries [14]. In other words, the cusp universality is the third and last universal spectral statistics for random matrix ensembles arising from natural generalizations of the Wigner matrices. We note that invariant β-ensembles may exhibit further universality classes, see [62].

In the companion paper [83] we established cusp universality for Wigner-type matrices in the complex Hermitian symmetry class. In the present work we extend this result to the real symmetric class and even to certain space-time correlation functions. In fact, we show the appearance of a natural one-parameter family of universal statistics associated to a family of singularities of the eigenvalue density that we call *physical cusps*. In both works we follow the *three step strategy*, a general method developed for proving local spectral universality for random matrices, see [90] for a pedagogical introduction. The first step is the *local law* or *rigidity*, establishing the location of the eigenvalues with a precision slightly above the typical local eigenvalue spacing. The second step is to establish universality for ensembles with a tiny Gaussian component. The third step is a perturbative argument to remove this tiny Gaussian component relying on the optimal local law. The first and third steps are insensitive to the symmetry type, in fact the optimal local law in the cusp regime has been established for both symmetry classes in [83] and it completes also the third step in both cases.

There are two different strategies for the second step. In the complex Hermitian symmetry class, the Brézin-Hikami formula [49] turns the problem into a saddle point analysis for a contour integral. This direct path was followed in [83] relying on the optimal local law. In the real symmetric case, lacking the Brézin-Hikami formula, only the second strategy via the analysis of Dyson Brownian motion (DBM) is feasible. This approach exploits the very fast decay to local equilibrium of DBM. It is the most robust and powerful method up to now to establish local spectral universality. In this paper we present a version of this method adjusted to the cusp situation. We will work in the real symmetric case for definiteness. The proof can easily be modified for the complex Hermitian case as well. The DBM method does not explicitly yield the local correlation kernel. Instead it establishes that the local statistics are universal and therefore can be identified from a reference ensemble that we will choose as the simplest Gaussian ensemble exhibiting a cusp singularity.

In this paper we partly follow the recent DBM analysis at the regular edges [131] and we extend it to the cusp regime, using the optimal local law from the companion paper [83] and the precise control of the density near the cusps [7, 14]. The main conceptual difference between [131] and the current work is that we obtain the necessary local law along the time evolution of DBM via novel DBM methods in Section 4.6. Some other steps, such as the Sobolev inequality, heat kernel estimates from [41] and the finite speed of propagation [39, 91, 131], require only moderate adjustments for the cusp regime, but for completeness we include them in the Appendix. The comparison of the short range approximation of the DBM with the full evolution, Lemma 4.7.2 and Lemma 4.C.1, will be presented in detail in Section 4.7 and in Appendix 4.C since it is more involved in the cusp setup, after the necessary estimates on the semicircular flow near the cusp are proven in Section 4.4.

We now outline the novelties and main difficulties at the cusp compared with the edge analysis in [131]. The basic idea is to interpolate between the time evolution of two DBM's, with initial conditions given by the original ensemble and the reference ensemble, respectively, after their local densities have been matched by shift and scaling. Beyond this common idea there are several differences.

The first difficulty lies in the rigidity analysis of the DBM starting from the interpolated initial conditions. The optimal rigidity from [83], that holds for very general Wigner-type matrices, applies for the flows of both the original and the reference matrices, but it does not directly apply to the interpolating process. The latter starts from a regular initial data but it runs for a very short time, violating the *flatness* (i.e. effective mean-field) assumption of [83]. While it is possible to extend the analysis of [83] to this case, here we chose a technically lighter and conceptually more interesting route. We use the maximum principle of the DBM to transfer rigidity information on the reference process to the interpolating one after an appropriate localization. Similar ideas for proving rigidity of the β-DBM flow has been used in the bulk [114] and at the edge [1].

The second difficulty in the cusp regime is that the shape of the density is highly unstable under the semicircular flow that describes the evolution of the density under the DBM. The regular edge analysed in [131] remains of square root type along its dynamics and it can be simply described by its location and its multiplicative *slope parameter* — both vary regularly with time. In contrast, the evolution of the cusp is a relatively complicated process: it starts with a small gap that shrinks to zero as the cusp forms and then continues developing a small local minimum. Heavily relying on the main results of [14], the density is described by quite involved shape functions, see (4.3c), (4.3e), that have a two-scale structure, given in terms of a total of three parameters, each varying on different time scales. For example, the location of the gap moves linearly with time, the length of gap shrinks as the $3/2$-th power of the time, while the local minimum after the cusp increases as the $1/2$-th power of the time. The scaling behavior of the corresponding quantiles, that approximate the eigenvalues by rigidity, follows the same complicated pattern of the density. All these require a very precise description of the semicircular flow near the cusp as well as the optimal rigidity.

The third difficulty is that we need to run the DBM for a relatively long time in order to exploit the local decay; in fact this time scale, $N^{-1/2+\epsilon}$ is considerably longer than the characteristic time scale $N^{-3/4}$ on which the physical cusp varies under the semicircular flow. We need to tune the initial condition very precisely so that after a relatively long time it develops a cusp exactly at the right location with the right slope.

The fourth difficulty is that, unlike for the regular edge regime, the eigenvalues or quantiles on both sides of the (physical) cusp contribute to the short range approximation of the dynamics, their effect cannot be treated as mean-field. Moreover, there are two scaling regimes for quantiles corresponding to the two-scale structure of the density.

Finally, we note that the analysis of the semicircular flow around the cusp, partly completed already in the companion paper [83], is relatively short and transparent despite its considerably more complex pattern compared to the corresponding analysis around the regular edge. This is mostly due to strong results imported from the general shape analysis [7]. Not only the exact formulas for the density shapes are taken over, but we also heavily rely on the $1/3$-Hölder continuity in space and time of the density and its Stieltjes transform, established in the strongest form in [14].

Notations and conventions. We now introduce some custom notations we use throughout the paper. For integers n we define $[n] := \{1, \dots, n\}$. For positive quantities f, g, we write $f \lesssim g$ and $f \sim g$ if $f \le Cg$ or, respectively, $cg \le f \le Cg$ for some constants c, C that depend only on the *model parameters*, i.e. on the constants appearing in the basic Assumptions (4.A)–(4.C) listed in Section 4.2 below. Similarly, we write $f \ll g$ if $f \le cg$ for some tiny constant $c > 0$ depending on the model parameters. We denote vectors by bold-faced lower case Roman letters $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{C}^N$, and matrices by upper case Roman letters $A, B \in \mathbf{C}^{N\times N}$. We write $\langle A\rangle := N^{-1} \operatorname{Tr} A$ and $\langle \boldsymbol{x}\rangle := N^{-1}\sum_{a\in[N]} x_a$ for the averaged trace and the average of a vector. We often identify diagonal matrices with the vector of its diagonal elements. Accordingly, for any matrix R, we denote by $\operatorname{diag}(R)$ the vector of its diagonal elements, and for any vector $\boldsymbol{r}$ we denote by $\operatorname{diag}(\boldsymbol{r})$ the corresponding diagonal matrix.

We will frequently use the concept of "with very high probability" meaning that for any fixed $D > 0$ the probability of the event is bigger than $1 - N^{-D}$ if $N \ge N_0(D)$.

Acknowledgement. The authors are very grateful to Johannes Alt for his invaluable contribution in helping improve several results of [14] tailored to the needs of this paper.

4.2 Main results

For definiteness we consider the real symmetric case $H \in \mathbf{R}^{N\times N}$. With small modifications the proof presented in this paper works for complex Hermitian case as well, but this case was already considered in [83] with a contour integral analysis. Let $W = W^* \in \mathbf{R}^{N\times N}$ be a symmetric random matrix and $A = \operatorname{diag}(\boldsymbol{a})$ be a deterministic diagonal matrix with entries $\boldsymbol{a} = (a_i)_{i=1}^N \in \mathbf{R}^N$. We say that W is of *Wigner-type* [6] if its entries w_{ij} for $i \le j$ are centred, $\mathbf{E}\, w_{ij} = 0$, independent random variables. We define the *variance matrix* or *self-energy matrix* $S = (s_{ij})_{i,j=1}^N$, $s_{ij} := \mathbf{E}\, w_{ij}^2$. In [6] it was shown that as N tends to infinity, the resolvent $G(z) := (H-z)^{-1}$ of the *deformed Wigner-type matrix* $H = A+W$ entrywise approaches a diagonal matrix $M(z) := \operatorname{diag}(\mathbf{m}(z))$ for $z \in \mathbf{H} := \{z \in \mathbf{C} | \Im z > 0\}$. The entries $\mathbf{m} = (m_1 \dots, m_N) \colon \mathbf{H} \to \mathbf{H}^N$ of M have positive imaginary parts and solve the *Dyson equation*

$$-\frac{1}{m_i(z)} = z - a_i + \sum_{j=1}^{N} s_{ij} m_j(z), \qquad z \in \mathbf{H} := \{z \in \mathbf{C} | \Im z > 0\}, \quad i \in [N]. \quad (4.1)$$

We call M or $\mathbf{m}$ the *self-consistent Green's function*. The normalised trace $\langle M \rangle$ of M is the Stieltjes transform $\langle M(z) \rangle = \int_{\mathbf{R}} (\tau - z)^{-1} \rho(\mathrm{d}\tau)$ of a unique probability measure ρ on $\mathbf{R}$ that approximates the empirical eigenvalue distribution of $A + W$ increasingly well as $N \to \infty$. We call ρ the *self-consistent density of states* (scDOS). Accordingly, its support $\operatorname{supp} \rho$ is called the *self-consistent spectrum*. It was proven in [7] that under very general conditions, $\rho(\mathrm{d}\tau)$ is an absolutely continuous measure with a $1/3$-Hölder continuous density, $\rho(\tau)$. Furthermore, the self-consistent spectrum consists of finitely many intervals with square root growth of ρ at the *edges*, i.e. at the points in $\partial \operatorname{supp} \rho$.

We call a point $\mathfrak{c} \in \mathbf{R}$ a cusp of ρ if $\mathfrak{c} \in (\mathring{\operatorname{supp}} \rho)$ and $\rho(\mathfrak{c}) = 0$. Cusps naturally emerge when we consider a one-parameter family of ensembles and two support intervals of ρ merge as the parameter value changes. The cusp universality phenomenon is not restricted to the exact cusp; it also occurs for situations shortly before and after the merging of two such support intervals, giving rise to a one parameter family of universal statistics. More precisely, universality emerges if ρ has a *physical cusp*. The terminology indicates that all these singularities become indistinguishable from the exact cusp if the density is resolved with a local precision above the typical eigenvalue spacing. We say that ρ exhibits a physical cusp if it has a small gap $(\mathfrak{e}_-, \mathfrak{e}_+) \subset \mathbf{R} \setminus \operatorname{supp} \rho$ with $\mathfrak{e}_+, \mathfrak{e}_- \in \operatorname{supp} \rho$ in its support of size $\mathfrak{e}_+ - \mathfrak{e}_- \lesssim N^{-3/4}$ or a local minimum $\mathfrak{m} \in (\mathring{\operatorname{supp}} \rho)$ of size $\rho(\mathfrak{m}) \lesssim N^{-1/4}$, cf. Figure 4.1. Correspondingly, we call the points $\mathfrak{b} := \frac{1}{2}(\mathfrak{e}_+ + \mathfrak{e}_-)$ and $\mathfrak{b} := \mathfrak{m}$ *physical cusp points*, respectively. One of the simplest models exhibiting a physical cusp point is the deformed Wigner matrix

$$H = \operatorname{diag}(1, \dots, 1, -1, \dots, -1) + \sqrt{1 + t} W \quad (4.2)$$

with equal numbers of ± 1, and where W is a Wigner matrix of variance $E|w_{ij}|^2 = N^{-1}$. The ensemble H from (4.2) exhibits an exact cusp if $t = 0$ and a physical cusp if $|t| \lesssim N^{-1/2}$, with $t > 0$ corresponding to a small non-zero local minimum and $t < 0$ corresponding to a small gap in the support of the self-consistent density. For the proof of universality in the real symmetric symmetry class we will use (4.2) with $W \sim$ GOE as a Gaussian reference ensemble.

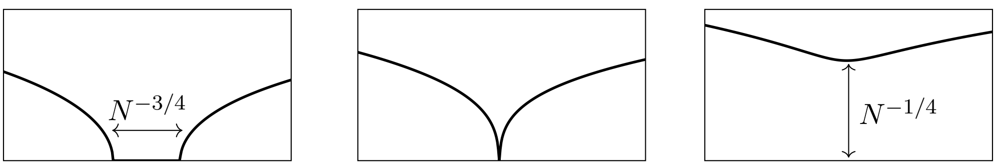


Figure 4.1: The cusp universality class can be observed in a 1-parameter family of *physical cusps*.

Our main result is cusp universality under the real symmetric analogues of the assumptions of [83]. Throughout this paper we make the following three assumptions:

Assumption (4.A) (Bounded moments). *The entries of the matrix $\sqrt{N}W$ have bounded moments and the expectation A is bounded, i.e. there are positive C_k such that*

$$|a_i| \le C_0, \qquad \mathbf{E}|w_{ij}|^k \le C_k N^{-k/2}, \qquad k \in \mathbf{N}.$$

Assumption (4.B) (Flatness). *We assume that the matrix S is flat in the sense $s_{ij} = \mathbf{E}\, w_{ij}^2 \ge c/N$ for some constant $c > 0$.*

Assumption (4.C) (Bounded self-consistent Green's function). *The scDOS ρ has a physical cusp point $\mathfrak{b}$, and in a neighbourhood of the physical cusp point $\mathfrak{b} \in \mathbf{R}$ the self-consistent Green's function is bounded, i.e. for positive C, κ we have*

$$|m_i(z)| \le C, \qquad z \in [\mathfrak{b} - \kappa, \mathfrak{b} + \kappa] + \mathrm{i}\mathbf{R}^+.$$

We call the constants appearing in Assumptions (4.A)–(4.C) *model parameters*. All generic constants in this paper may implicitly depend on these model parameters. Dependence on further parameters, however, will be indicated.

Remark 4.2.1. *The boundedness of $\mathbf{m}$ in Assumption (4.C) can be, for example, ensured by assuming some regularity of the variance matrix S. For more details we refer to [7, Chapter 6].*

According to the extensive analysis in [7, 14] it follows[1] that there exists some small $\delta_* \sim 1$ such that the self-consistent density ρ around the points where it is small exhibits one of the following three types of behaviours.

(i) *Exact cusp.* There is a cusp point $\mathfrak{c} \in \mathbf{R}$ in the sense that $\rho(\mathfrak{c}) = 0$ and $\rho(\mathfrak{c} \pm \delta) > 0$ for $0 \ne \delta \ll 1$. In this case the self-consistent density is locally around $\mathfrak{c}$ given by

$$\rho(\mathfrak{c} + \omega) = \frac{\sqrt{3}\gamma^{4/3}|\omega|^{1/3}}{2\pi}\Big[1 + \mathcal{O}(|\omega|^{1/3})\Big] \tag{4.3a}$$

for $\omega \in [-\delta_*, \delta_*]$ and some $\gamma > 0$.

(ii) *Small gap.* There is a maximal interval $[\mathfrak{e}_-, \mathfrak{e}_+]$ of size $0 < \Delta := \mathfrak{e}_+ - \mathfrak{e}_- \ll 1$ such that $\rho|_{[\mathfrak{e}_-, \mathfrak{e}_+]} \equiv 0$. In this case the density around $\mathfrak{e}_\pm$ is, for some $\gamma > 0$, locally given by

$$\rho(\mathfrak{e}_\pm \pm \omega) = \frac{\sqrt{3}(2\gamma)^{4/3}\Delta^{1/3}}{2\pi}\Psi_{\mathrm{edge}}(\omega/\Delta)\left[1 + \mathcal{O}\Big(\min\Big\{\omega^{1/3}, \frac{\omega^{1/2}}{\Delta^{1/6}}\Big\}\Big)\right] \tag{4.3b}$$

for $\omega \in [0, \delta_*]$, where

$$\Psi_{\mathrm{edge}}(\lambda) := \frac{\sqrt{\lambda(1+\lambda)}}{(1 + 2\lambda + 2\sqrt{\lambda(1+\lambda)})^{2/3} + (1 + 2\lambda - 2\sqrt{\lambda(1+\lambda)})^{2/3} + 1}, \quad \lambda \ge 0. \tag{4.3c}$$

[1] The claimed expansions (4.3a) and (4.3d) follow directly from [14, Theorem 7.2(c), (d)]. The error term in (4.3b) follows from [14, Theorem 7.1(a)], where we define γ according to h therein.

(iii) *Non-zero local minimum.* There is a local minimum at $\mathfrak{m} \in \mathbf{R}$ of ρ such that $0 < \rho(\mathfrak{m}) \ll 1$. In this case there exists some $\gamma > 0$ such that

$$\rho(\mathfrak{m}+\omega) = \rho(\mathfrak{m}) + \rho(\mathfrak{m})\Psi_{\min}\left(\frac{3\sqrt{3}\gamma^4\omega}{2(\pi\rho(\mathfrak{m}))^3}\right) \times \left[1+\mathcal{O}\left(\min\Big\{\rho(\mathfrak{m})^{1/2}, \frac{\rho(\mathfrak{m})^4}{|\omega|}\Big\} + \min\Big\{\frac{\omega^2}{\rho(\mathfrak{m})^5}, |\omega|^{1/3}\Big\}\right)\right] \tag{4.3d}$$

for $\omega \in [-\delta_*, \delta_*]$, where

$$\Psi_{\min}(\lambda) := \frac{\sqrt{1+\lambda^2}}{(\sqrt{1+\lambda^2}+\lambda)^{2/3} + (\sqrt{1+\lambda^2}-\lambda)^{2/3} - 1} - 1, \qquad \lambda \in \mathbf{R}. \tag{4.3e}$$

We note that the choices for the *slope* parameter γ in (4.3b)–(4.3d) are consistent with (4.3a) in the sense that in the regimes $\Delta \ll \omega \ll 1$ and $\rho(\mathfrak{m})^3 \ll |\omega| \ll 1$ the respective formulae asymptotically agree. The precise form of the pre-factors in (4.3) is also chosen such that in the universality statement γ is a linear rescaling parameter.

It is natural to express universality in terms of a rescaled k-point function $p_k^{(N)}$ which we define implicitly by

$$\mathbf{E}\binom{N}{k}^{-1} \sum_{\{i_1,\dots,i_k\}\subset[N]} f(\lambda_{i_1},\dots,\lambda_{i_k}) = \int_{\mathbf{R}^k} f(\boldsymbol{x}) p_k^{(N)}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \tag{4.4}$$

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

Theorem 4.2.2. *Let H be a real symmetric or complex Hermitian deformed Wigner-type matrix whose scDOS ρ has a physical cusp point $\mathfrak{b}$ such that Assumptions (4.A)–(4.C) are satisfied. Let $\gamma > 0$ be the slope parameter at $\mathfrak{b}$, i.e. such that ρ is locally around $\mathfrak{b}$ given by* (4.3)*. Then the local k-point correlation function at $\mathfrak{b}$ is universal, i.e. for any $k \in \mathbf{N}$ there exists a k-point correlation function $p_{k,\alpha}^{\mathrm{GOE/GUE}}$ such that for any test function $F \in C_c^1(\overline{\Omega})$, with $\Omega \subset \mathbf{R}^k$ some bounded open set, it holds that*

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

where the parameter α and the physical cusp $\mathfrak{b}$ are given by

$$\alpha := \begin{cases} 0 & \textit{in case (i)} \\ 3\,(\gamma\Delta/4)^{2/3}\,N^{1/2} & \textit{in case (ii)} \\ -\,(\pi\rho(\mathfrak{m})/\gamma)^2\,N^{1/2} & \textit{in case (iii)}, \end{cases} \qquad \mathfrak{b} := \begin{cases} \mathfrak{c} & \textit{in case (i)} \\ (\mathfrak{e}_- + \mathfrak{e}_+)/2 & \textit{in case (ii)} \\ \mathfrak{m} & \textit{in case (iii)}, \end{cases} \tag{4.5}$$

and $c(k) > 0$ is a small constant only depending on k. The implicit constant in the error term depends on k and the diameter of the set Ω.

Remark 4.2.3. *(i) In the complex Hermitian symmetry class the k-point function is given by*

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

Here the extended Pearcey kernel $K_{\alpha,\beta}$ *is given by*

$$K_{\alpha,\beta}(x,y) = \frac{1}{(2\pi \mathrm{i})^2}\int_\Xi \mathrm{d}z \int_\Phi \mathrm{d}w \frac{\exp(-w^4/4+\beta w^2/2 - yw + z^4/4 - \alpha z^2/2 + xz)}{w-z} - \frac{\mathbf{1}_{\beta>\alpha}}{\sqrt{2\pi(\beta-\alpha)}}\exp\Big(-\frac{(y-x)^2}{2(\beta-\alpha)}\Big), \tag{4.6}$$

where Ξ *is a contour consisting of rays from* $\pm\infty e^{\mathrm{i}\pi/4}$ *to* 0 *and rays from* 0 *to* $\pm\infty e^{-\mathrm{i}\pi/4}$, *and* Φ *is the ray from* $-\mathrm{i}\infty$ *to* $\mathrm{i}\infty$. *For more details we refer to [2, 50, 204] and the references in [83].*

(ii) *The real symmetric k-point function (possibly only a distribution)* $p_{k,\alpha}^{\mathrm{GOE}}$ *is not known explicitly. In fact, it is not even known whether* $p_{k,\alpha}^{\mathrm{GOE}}$ *is Pfaffian. We will nevertheless establish the existence of* $p_{k,\alpha}^{\mathrm{GOE}}$ *as a distribution in the dual of the* C^1 *functions in Section 4.3 as the limit of the correlation functions of a one parameter family of Gaussian comparison models.*

Theorem 4.2.2 is a universality result about the spatial correlations of eigenvalues. Our method also allows us to prove the corresponding statement on space-time universality when we consider the time evolution of eigenvalues $(\lambda_i^t)_{i\in[N]}$ according to the Dyson Brownian motion $\mathrm{d}H^{(t)} = \mathrm{d}\mathfrak{B}_t$ with initial condition $H^{(0)} = H$, where, depending on the symmetry class, $\mathfrak{B}_t$ is a complex Hermitian or real symmetric matrix valued Brownian motion. For any ordered k-tuple $\boldsymbol{\tau} = (\tau_1,\dots,\tau_k)$ with $0 \le \tau_1 \le \dots \le \tau_k \lesssim N^{-1/2}$ we then define the *time-dependent k-point function* as follows. Denote the unique values in the tuple $\boldsymbol{\tau}$ by $\sigma_1 < \dots < \sigma_l$ such that $\{\tau_1,\dots,\tau_k\} = \{\sigma_1,\dots,\sigma_l\}$ and denote the multiplicity of σ_j in $\boldsymbol{\tau}$ by k_j and note that $\sum k_j = k$. We then define $p_{k,\boldsymbol{\tau}}^{(N)}$ implicitly via

$$\mathbf{E}\prod_{j=1}^l \left[\binom{N}{k_j}^{-1} \sum_{\{i_1^j,\dots,i_{k_j}^j\}\subset[N]}\right] f(\lambda_{i_1^1}^{\sigma_1},\dots,\lambda_{i_{k_1}^1}^{\sigma_1},\dots,\lambda_{i_1^l}^{\sigma_l},\dots,\lambda_{i_{k_l}^l}^{\sigma_l}) = \int_{\mathbf{R}^k} f(\boldsymbol{x}) p_{k,\boldsymbol{\tau}}^{(N)}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \tag{4.7}$$

for test functions f and note that (4.7) reduces to (4.4) in the case $\tau_1 = \dots = \tau_k = 0$. We note that in (4.7) coinciding indices are allowed only for eigenvalues at different times. If the scDOS ρ of H has a physical cusp in $\mathfrak{b}$, then for $\tau \lesssim N^{-1/2}$ the scDOS ρ_τ of $H^{(\tau)}$ also has a physical cusp $\mathfrak{b}_\tau$ close to $\mathfrak{b}$ and we can prove space-time universality in the sense of the following theorem, whose proof we defer to Appendix 4.A.

Theorem 4.2.4. *Let* H *be a real symmetric or complex Hermitian deformed Wigner-type matrix whose scDOS* ρ *has a physical cusp point* $\mathfrak{b}$ *such that Assumptions (4.A)–(4.C) are satisfied. Let* $\gamma > 0$ *be the slope parameter at* $\mathfrak{b}$, *i.e. such that* ρ *is locally around* $\mathfrak{b}$ *given by* (4.3). *Then there exists a k-point correlation function* $p_{k,\boldsymbol{\alpha}}^{\mathrm{GOE/GUE}}$ *such that for any* $0 \le \tau_1 \le \dots \le \tau_k \lesssim N^{-1/2}$ *and for any test function* $F \in C_c^1(\overline{\Omega})$, *with* $\Omega \subset \mathbf{R}^k$ *some bounded open set, it holds that*

$$\int_{\mathbf{R}^k} F(\boldsymbol{x}) \left[\frac{N^{k/4}}{\gamma^k} p_{k,\boldsymbol{\tau}/\gamma^2}^{(N)}\left(\mathfrak{b}_{\boldsymbol{\tau}/\gamma^2} + \frac{\boldsymbol{x}}{\gamma N^{3/4}}\right) - p_{k,\boldsymbol{\alpha}}^{\mathrm{GOE/GUE}}(\boldsymbol{x})\right]\mathrm{d}\boldsymbol{x} = \mathcal{O}_{k,\Omega}\left(N^{-c(k)}\|F\|_{C^1}\right),$$

where $\boldsymbol{\tau} = (\tau_1,\dots,\tau_k)$, $\mathfrak{b}_{\boldsymbol{\tau}} = (\mathfrak{b}_{\tau_1},\dots,\mathfrak{b}_{\tau_k})$ *and* $\boldsymbol{\alpha} = \alpha - \boldsymbol{\tau}N^{1/2}$ *with* α *from* (4.5) *and* $c(k) > 0$ *is a small constant only depending on* k. *In the case of the complex Hermitian symmetry* *class the k-point correlation function is known to be determinantal of the form*

$$p^{\mathrm{GUE}}_{\alpha_1,\dots,\alpha_k}(\boldsymbol{x}) = \det\Big(K_{\alpha_i,\alpha_j}(x_i,x_j)\Big)_{i,j=1}^{k}$$

with $K_{\alpha,\beta}$ as in (4.6).

The analogous version of Theorem 4.2.4 for fixed energy bulk multitime universality has been proven in [129, Sec. 2.3.1.].

Remark 4.2.5. *The extended Pearcey kernel $K_{\alpha,\beta}$ in Theorem 4.2.4 has already been observed for the double-scaling limit of non-intersecting Brownian bridges [2, 204]. However, in the random matrix setting our methods also allow us to prove that the space-time universality of Theorem 4.2.4 extends beyond the Gaussian DBM flow. If the times $0 \le \tau_1 \le \cdots \le \tau_k \lesssim N^{-1/2}$ are ordered, then the k-point correlation function of the DBM flow asymptotically agrees with the k-point correlation function of eigenvalues of the matrices*

$$H+\sqrt{\tau_1}W_1, H+\sqrt{\tau_1}W_1+\sqrt{\tau_2-\tau_1}W_2,\dots,H+\sqrt{\tau_1}W_1+\cdots+\sqrt{\tau_k-\tau_{k-1}}W_k$$

for independent standard Wigner matrices $W_1,\dots,W_k$.

4.3 Ornstein-Uhlenbeck flow

Starting from this section we consider a more general framework that allows for random matrix ensembles with certain correlation among the entries. In this way we stress that our proofs regarding the semicircular flow and the Dyson Brownian motion are largely model independent, assuming the optimal local law holds. The independence assumption on the entries of W is made only because we rely on the local law from [83] that was proven for deformed Wigner-type matrices. We therefore present the flow directly in the more general framework of the *matrix Dyson equation* (MDE)

$$1+(z-A+\mathcal{S}[M(z)])M(z)=0, \qquad A:=\mathbf{E}\,H, \qquad \mathcal{S}[R]:=\mathbf{E}\,WRW, \tag{4.8}$$

with spectral parameter in the complex upper half plane, $\Im z>0$, and positive definite imaginary part, $\frac{1}{2\mathrm{i}}(M(z)-M(z)^*)>0$, of the solution M. The MDE generalizes (4.1). Note that in the deformed Wigner-type case the *self-energy operator* $\mathcal{S}\colon \mathbf{C}^{N\times N}\to\mathbf{C}^{N\times N}$ is related to the variance matrix S by $\mathcal{S}[\operatorname{diag}\boldsymbol{r}]=\operatorname{diag}(S\boldsymbol{r})$.

As in [83] we consider the Ornstein-Uhlenbeck flow

$$\mathrm{d}\widetilde{H}_s=-\frac{1}{2}(\widetilde{H}_s-A)\,\mathrm{d}s+\Sigma^{1/2}[\mathrm{d}\mathfrak{B}_s], \qquad \Sigma[R]:=\frac{\beta}{2}\,\mathbf{E}\,W\operatorname{Tr}WR, \qquad \widetilde{H}_0:=H, \tag{4.9}$$

which preserves expectation and self-energy operator $\mathcal{S}$. Since we consider real symmetric H, the parameter β indicating the symmetry class is $\beta=1$. In (4.9) with $\mathfrak{B}_s\in\mathbf{R}^{N\times N}$ we denote a real symmetric matrix valued standard (GOE) Brownian motion, i.e. $(\mathfrak{B}_s)_{ij}$ for $i<j$ and $(\mathfrak{B}_s)_{ii}/\sqrt{2}$ are independent standard Brownian motions and $(\mathfrak{B}_s)_{ji}=(\mathfrak{B}_s)_{ij}$. In case H were complex Hermitian, we would have $\beta=2$ and $\mathrm{d}\mathfrak{B}_s$ would be an infinitesimal GUE matrix. This was the setting in [83]. The OU flow effectively adds a small

Gaussian component of size $\sqrt{s}$ to $\widetilde{H}_s$. More precisely, we can construct a Wigner-type matrix H_s, satisfying Assumptions (4.A)–(4.C), such that, for any fixed s,

$$\widetilde{H}_s = H_s + \sqrt{cs}U, \qquad \mathcal{S}_s = \mathcal{S} - cs\mathcal{S}^{\mathrm{GOE}}, \qquad \mathbf{E}\, H_s = A, \qquad U \sim \mathrm{GOE}, \quad (4.10)$$

where U is independent of H_s. Here $c > 0$ is a small universal constant which depends on the constant in Assumption (4.B), $\mathcal{S}_s$ is the self-energy operator corresponding to H_s and $\mathcal{S}^{\mathrm{GOE}}[R] := \langle R\rangle + R^t/N$, where $\langle\cdot\rangle := N^{-1}\mathrm{Tr}(\cdot)$ and R^t denotes the transpose of R. Since $\mathcal{S}$ is flat in the sense $\mathcal{S}[R] \gtrsim \langle R\rangle$ and s is small it follows that also $\mathcal{S}_s$ is flat.

As a consequence of the well established Green function comparison technique the k-point function of $H = \widetilde{H}_0$ is comparable with the one of $\widetilde{H}_s$ as long as $s \le N^{-1/4-\epsilon}$ for some $\epsilon > 0$. Indeed, from [83, Eq. (116)] for any $F \in C_c^1(\overline{\Omega})$, compactly supported C^1 test function on a bounded open set $\Omega \subset \mathbf{R}^k$, we find

$$\int_{\mathbf{R}^k} F(\boldsymbol{x}) N^{k/4}\Big[p_k^{(N)}\Big(\mathfrak{b}+\frac{\boldsymbol{x}}{\gamma N^{3/4}}\Big) - \widetilde{p}_{k,s}^{(N)}\Big(\mathfrak{b}+\frac{\boldsymbol{x}}{\gamma N^{3/4}}\Big)\Big]\,\mathrm{d}\boldsymbol{x} = \mathcal{O}_{k,\Omega}(N^{-c}\|F\|_{C^1}), \quad (4.11)$$

where $\widetilde{p}_{k,s}^{(N)}$ is the k-point correlation function of $\widetilde{H}_s$, $c = c(k) > 0$ is some constant.

It follows from the flatness assumption that the matrix H_s satisfies the assumptions of the local law from [83, Theorem 2.5] uniformly in $s \ll 1$. Therefore [83, Corollary 2.6] implies that the eigenvalues of H_s are rigid down to the optimal scale. It remains to prove that for long enough times s the local eigenvalue statistics of $H_s + \sqrt{cs}U$ on a scale of $1/\gamma N^{3/4}$ around $\mathfrak{b}$ agree with the local eigenvalue statistics of the Gaussian reference ensemble around 0 at a scale of $1/N^{3/4}$. By a simple rescaling Theorem 4.2.2 then follows from (4.11) together with the following proposition.

Proposition 4.3.1. *Let $t_1 := N^{-1/2+\omega_1}$ with some small $\omega_1 > 0$ and let t_* be such that $|t_* - t_1| \lesssim N^{-1/2}$. Assume that $H^{(\lambda)}$ and $H^{(\mu)}$ [2] are Wigner-type matrices satisfying Assumptions (4.A)–(4.C) such that the scDOSs $\rho_{\lambda,t_*}, \rho_{\mu,t_*}$ of $H^{(\lambda)} + \sqrt{t_*}U^{(\lambda)}$ and $H^{(\mu)} + \sqrt{t_*}U^{(\mu)}$ with independent $U^{(\lambda)}, U^{(\mu)} \sim \mathrm{GOE}$ have cusps in some points $\mathfrak{c}_\lambda$, $\mathfrak{c}_\mu$ such that locally around $\mathfrak{c}_r$, $r = \lambda, \mu$, the densities ρ_{r,t_*} are given by (4.3a) with $\gamma = 1$. Then the local k-point correlation functions $p_{k,t_1}^{(N,r)}$ of $H^{(r)} + \sqrt{t_1}U^{(r)}$ around the respective physical cusps $\mathfrak{b}_{r,t_1}$ of ρ_{r,t_1}, $j = 1, 2$, asymptotically agree in the sense*

$$\int_{\mathbf{R}^k} F(\boldsymbol{x}) N^{k/4}\left[p_{k,t_1}^{(N,\lambda)}\left(\mathfrak{b}_{\lambda,t_1} + \frac{\boldsymbol{x}}{N^{3/4}}\right) - p_{k,t_1}^{(N,\mu)}\left(\mathfrak{b}_{\mu,t_1} + \frac{\boldsymbol{x}}{N^{3/4}}\right)\right]\mathrm{d}\boldsymbol{x} = \mathcal{O}(N^{-c(k)}\|F\|_{C^1})$$

for any $F \in C_c^1(\overline{\Omega})$, with $\Omega \subset \mathbf{R}^k$ a bounded open set. The implicit constant in $\mathcal{O}(\cdot)$ may depend on k and Ω.

Proof of Theorem 4.2.2. Set $s := t_1/c\theta^2$ and $H^{(\lambda)} := \theta H_s$ where c is the constant from (4.10) and $\theta \sim 1$ is yet to be chosen. Note that $H^{(\lambda)} + \sqrt{t}U = \theta(H_s + \sqrt{t/\theta^2}U)$, and in particular $H^{(\lambda)} + \sqrt{t_1}U = \widetilde{H}_s$. Moreover, it follows from the semicircular flow analysis in Section 4.4 that for some t_* with $|t_* - t_1| \lesssim N^{-1/2}$, the scDOS $\theta\rho_{\lambda,t_*}(\lambda\cdot)$ of $H_s + \sqrt{t_*/\theta^2}U$ and thereby also ρ_{λ,t_*}, the one of $H^{(\lambda)} + \sqrt{t_*}U$, have exact cusps in $\mathfrak{c}_\lambda/\theta$ and $\mathfrak{c}_\lambda$, respectively.

[2] We use the notation $H^{(\lambda)}$ and $H^{(\mu)}$ since we denote the eigenvalues of $H^{(\lambda)}$ and $H^{(\mu)}$ by λ_i and μ_i respectively, with $1 \le i \le N$ respectively.

It follows from the $1/3$-Hölder continuity of the slope parameter, cf. [14, Lemma 10.5, Eq. (7.5a)], that locally around $\mathfrak{c}_\lambda/\theta$ the scDOS of $H_s + \sqrt{t_*/\theta^2}U$ is given by

$$\theta\rho_{\lambda,t_*}(\mathfrak{c}_\lambda + \theta\omega) = \theta\rho_{\lambda,t_*}\Big(\theta\Big(\frac{\mathfrak{c}_\lambda}{\theta} + \omega\Big)\Big) = \frac{\sqrt{3}\gamma^{4/3}|\omega|^{1/3}}{2\pi}\Big[1 + \mathcal{O}(|\omega|^{1/3} + |t_* - t_1|^{1/3})\Big].$$

Whence we can choose $\theta = \gamma\big[1 + \mathcal{O}(|t_1 - t_*|^{1/3})\big]$ appropriately such that

$$\rho_{\lambda,t_*}(\mathfrak{c}_\lambda + \omega) = \frac{\sqrt{3}|\omega|^{1/3}}{2\pi}\Big[1 + \mathcal{O}(|\omega|^{1/3})\Big]$$

and it follows that $H^{(\lambda)}$ satisfies the assumptions of Proposition 4.3.1, in particular the slope parameter of $H^{(\lambda)} + \sqrt{t_*}U$ is normalized to 1. Furthermore, the almost cusp $\mathfrak{b}_{\lambda,t_1}$ of $H^{(\lambda)} + \sqrt{t_1}U$ is given by $\mathfrak{b}_{\lambda,t_1} = \theta\mathfrak{b}$ with $\mathfrak{b}$ as in Theorem 4.2.2.

We now choose our Gaussian comparison model. For $\alpha \in \mathbf{R}$ we consider the *reference ensemble*

$$U_\alpha = U_\alpha^{(N)} := \operatorname{diag}(1, \dots, 1, -1, \dots, -1) + \sqrt{1 - \alpha N^{-1/2}}U \in \mathbf{R}^{N\times N}, \qquad U \sim \text{GOE} \tag{4.12}$$

with $\lfloor N/2\rfloor$ and $\lceil N/2\rceil$ times ± 1 in the deterministic diagonal. An elementary computation shows that for even N and $\alpha = 0$, the self-consistent density of U_α has an exact cusp of slope $\gamma = 1$ in $\mathfrak{c} = 0$, i.e. it is given by (4.3a). For odd N the exact cusp is at distance $\lesssim N^{-1}$ away from 0 which is well below the natural scale of order $N^{-3/4}$ of the eigenvalue fluctuation and therefore has no influence on the k-point correlation function. The reference ensemble U_α has for $0 \neq |\alpha| \sim 1$ a small gap of size $N^{-3/4}$ or small local minimum of size $N^{-1/4}$ at the physical cusp point $|\mathfrak{b}| \lesssim \frac{1}{N}$, depending on the sign of α. Using the definition in (4.12), let $H^{(\mu)} := U_{N^{1/2}t_*}$ from which it follows that $H^{(\mu)} + \sqrt{t_*}U \sim U_0$ has an exact cusp in 0 whose slope is 1 by an easy explicit computation in the case of even N. For odd N the cusp emerges at a distance of $\lesssim N^{-1}$ away from 0, which is well below the investigated scale. Thus also $H^{(2)}$ satisfies the assumptions of Proposition 4.3.1. The almost cusp $\mathfrak{b}_{\mu,t_1}$ is given by $\mathfrak{b}_{\mu,t_1} = 0$ by symmetry of the density ρ_{μ,t_1} in the case of even N and at a distance of $|\mathfrak{b}_{\mu,t_1}| \lesssim N^{-1}$ in the case of odd N. This fact follows, for example, from explicitly solving the 2d-quadratic equation. The perturbation of size $1/N$ is not visible on the scale of the k-point correlation functions.

Now Proposition 4.3.1 together with (4.11) and $s \sim N^{-1/2+\omega_1}$ implies that

$$\int_{\mathbf{R}^k} F(\boldsymbol{x})N^{k/4}\left[\frac{1}{\theta^k}p_k^{(N)}\left(\mathfrak{b} + \frac{\boldsymbol{x}}{\theta N^{3/4}}\right) - p_{k,\alpha,\text{GOE}}^{(N)}\left(\frac{\boldsymbol{x}}{N^{3/4}}\right)\right]\mathrm{d}\boldsymbol{x} = \mathcal{O}(N^{-c(k)}\|F\|_{C^1(\Omega)}) \tag{4.13}$$

with $\alpha = N^{1/2}(t_* - t_1)$, where $p_{k,\alpha,\text{GOE}}^{(N)}$ denotes the k-point function of the comparison model U_α. This completes the proof of Theorem 4.2.2 modulo the comparison of $p_{k,\alpha,\text{GOE}}^{(N)}$ with its limit by relating $t_* - t_1$ to the size of the gap and the local minimum of ρ via [83, Lemma 5.1] (or (4.22a)–(4.22c) later) and recalling that $\theta = \gamma\big[1 + \mathcal{O}(|t_1 - t_*|^{1/3})\big]$.

To complete the proof we claim that for any fixed k and α there exists a distribution $p_{k,\alpha}^{\text{GOE}}$ on $\mathbf{R}^k$, locally in the dual of $C_c^1(\overline{\Omega})$ for every open bounded $\Omega \subset \mathbf{R}^k$, such that

$$\int_{\mathbf{R}^k} F(\boldsymbol{x})\left[N^{k/4}p_{k,\alpha,\text{GOE}}^{(N)}\left(\frac{\boldsymbol{x}}{N^{3/4}}\right) - p_{k,\alpha}^{\text{GOE}}(\boldsymbol{x})\right]\mathrm{d}\boldsymbol{x} = \mathcal{O}_{k,\Omega}(N^{-c(k)}\|F\|_{C^1}) \tag{4.14}$$

holds for any $F \in C_c^1(\Omega)$. We now show that (4.14) is a straightforward consequence of (4.13).

First notice that, for notational simplicity, we gave the proof of (4.13) only for the case when H and U_α are of the same dimension, but it works without any modification when their dimensions are only comparable, see Remark 4.5.2. Hence, applying this result to a sequence of GOE ensembles $U_\alpha^{(N_n)}$ with $N_n := (4/3)^n$, for any compactly supported $F \in C_c^1(\overline{\Omega})$ we have

$$\begin{aligned}\int_{\mathbf{R}^k} F(\boldsymbol{x}) \left[N_n^{k/4} p_{k,\alpha,\mathrm{GOE}}^{(N_n)} \left(\frac{\boldsymbol{x}}{N_n^{3/4}} \right) - N_{n+1}^{k/4} p_{k,\alpha,\mathrm{GOE}}^{(N_{n+1})} \left(\frac{\boldsymbol{x}}{N_{n+1}^{3/4}} \right) \right] \mathrm{d}\boldsymbol{x} \\ = \mathcal{O}_{k,\Omega}\Big(\left(\frac{3}{4} \right)^{nc(k)} \|F\|_{C^1} \Big).\end{aligned} \tag{4.15}$$

Fix a bounded open set $\Omega \subset \mathbf{R}^k$ and define the sequence of functionals $\{\mathcal{J}_n\}_{n\in\mathbf{N}}$ in the dual space $C_c^1(\overline{\Omega})^*$ as follows

$$\mathcal{J}_n(F) := \int_{\mathbf{R}^k} F(\boldsymbol{x}) N_n^{k/4} p_{k,\alpha,\mathrm{GOE}}^{(N_n)} \left(\frac{\boldsymbol{x}}{N_n^{3/4}} \right) \mathrm{d}\boldsymbol{x},$$

for any $F \in C_c^1(\overline{\Omega})$. Then, by (4.15) it easily follows that $\{\mathcal{J}_n\}_{n\in\mathbf{N}}$ is a Cauchy sequence on $C_c^1(\overline{\Omega})^*$. Indeed, for any $M > L$ we have by a telescopic sum

$$\begin{aligned}&|(\mathcal{J}_M - \mathcal{J}_L)(F)| \\ &\quad = \left| \sum_{n=L}^{M-1} \int_{\mathbf{R}^k} F(\boldsymbol{x}) \left[N_{n+1}^{k/4} p_{k,\alpha,\mathrm{GOE}}^{(N_{n+1})} \left(\frac{\boldsymbol{x}}{N_{n+1}^{3/4}} \right) - N_n^{k/4} p_{k,\alpha,\mathrm{GOE}}^{(N_n)} \left(\frac{\boldsymbol{x}}{N_n^{3/4}} \right) \right] \mathrm{d}\boldsymbol{x} \right| \\ &\quad \le C_{k,\Omega} \left(\frac{3}{4} \right)^{Lc(k)} \|F\|_{C^1}.\end{aligned} \tag{4.16}$$

Thus, we conclude that there exists a unique $\mathcal{J}_\infty \in C_c^1(\overline{\Omega})^*$ such that $\mathcal{J}_n \to \mathcal{J}_\infty$ as $n \to \infty$ in norm. Then, (4.16) clearly concludes the proof of (4.14), identifying $\mathcal{J}_\infty = \mathcal{J}_\infty^{(\Omega)}$ with $p_{k,\alpha}^{\mathrm{GOE}}$ restricted to $\overline{\Omega}$. Since this holds for any open bounded set $\Omega \subset \mathbf{R}^k$, the distribution $p_{k,\alpha}^{\mathrm{GOE}}$ can be identified with the inductive limit of the consistent family of functionals $\{\mathcal{J}_\infty^{(\Omega_m)}\}_{m\ge1}$, where, say, Ω_m is the ball of radius m. This completes the proof of Theorem 4.2.2.

□

4.4 Semicircular flow analysis

In this section we analyse various properties of the semicircular flow in order to prepare the Dyson Brownian motion argument in Section 4.6 and Section 4.7. If ρ is a probability density on $\mathbf{R}$ with Stieltjes transform m, then the free semicircular evolution $\rho_t^{\mathrm{fc}} = \rho \boxplus \sqrt{t}\rho_{\mathrm{sc}}$ of ρ is defined as the unique probability measure whose Stieltjes transform m_t^{fc} solves the implicit equation

$$m_t^{\mathrm{fc}}(\zeta) = m(\zeta + t m_t^{\mathrm{fc}}(\zeta)), \qquad \zeta \in \mathbf{H}, \quad t \ge 0. \tag{4.17}$$

Here $\sqrt{t}\rho_{\mathrm{sc}}$ is the semicircular distribution of variance t.

We now prepare the Dyson Brownian motion argument in Section 4.7 by providing a detailed analysis of the scDOS along the semicircular flow. As in Proposition 4.3.1 we consider the setting of two densities ρ_λ, ρ_μ whose semicircular evolutions reach a cusp of the same slope at the same time. Within the whole section we shall assume the following setup: Let ρ_λ, ρ_μ be densities associated with solutions M_λ, M_μ to some Dyson equations satisfying Assumptions (4.A)–(4.C) (or their matrix counterparts). We consider the free convolutions $\rho_{\lambda,t} := \rho_\lambda \boxplus \sqrt{t}\rho_{\mathrm{sc}}$, $\rho_{\mu,t} := \rho_\mu \boxplus \sqrt{t}\rho_{\mathrm{sc}}$ of ρ_λ, ρ_μ with semicircular distributions of variance t and assume that after a time $t_* \sim N^{-1/2+\omega_1}$ both densities $\rho_{\lambda,t_*}, \rho_{\mu,t_*}$ have cusps in points $\mathfrak{c}_\lambda, \mathfrak{c}_\mu$ around which they can be approximated by (4.3a) with the same $\gamma = \gamma_\lambda(t_*) = \gamma_\mu(t_*)$. It follows from the semicircular flow analysis in [83, Lemma 5.1] that for $0 \le t \le t_*$ both densities have small gaps $[\mathfrak{e}^-_{r,t}, \mathfrak{e}^+_{r,t}]$, $r = \lambda, \mu$ in their supports, while for $t_* \le t \le 2t_*$ they have non-zero local minima in some points $\mathfrak{m}_{r,t}$, $r = \lambda, \mu$. Instead of comparing the eigenvalue flows corresponding to ρ_λ, ρ_μ directly, we rather consider a continuous interpolation ρ_α for $\alpha \in [0,1]$ of ρ_λ and ρ_μ. For technical reasons we define this interpolated density $\rho_{\alpha,t}$ as an interpolation of $\rho_{\lambda,t}$ and $\rho_{\mu,t}$ separately for each time t, rather than considering the evolution $\rho_{\alpha,0} \boxplus \sqrt{t}\rho_{\mathrm{sc}}$ of the initial interpolation $\rho_{\alpha,0}$. We warn the reader that semicircular evolution and interpolation do not commute, i.e. $\rho_{\alpha,t} \ne \rho_{\alpha,0} \boxplus \sqrt{t}\rho_{\mathrm{sc}}$. We now define the concept of *interpolating densities* following [131, Section 3.1.1].

Definition 4.4.1. *For* $\alpha \in [0,1]$ *define the* α-interpolating density $\rho_{\alpha,t}$ *as follows. For any* $0 \le E \le \delta_*$ *and* $r = \lambda, \mu$ *let*

$$n_{r,t}(E) := \int_{\mathfrak{e}^+_{r,t}}^{\mathfrak{e}^+_{r,t}+E} \rho_{r,t}(\omega)\,\mathrm{d}\omega, \quad 0 \le t \le t_*,$$
$$n_{r,t}(E) := \int_{\mathfrak{m}_{r,t}}^{\mathfrak{m}_{r,t}+E} \rho_{r,t}(\omega)\,\mathrm{d}\omega, \quad t_* \le t \le 2t_*$$

be the counting functions and $\varphi_{\lambda,t}$, $\varphi_{\mu,t}$ *their inverses, i.e.* $n_{r,t}(\varphi_{r,t}(s)) = s$. *Define now*

$$\varphi_{\alpha,t}(s) := \alpha\varphi_{\lambda,t}(s) + (1-\alpha)\varphi_{\mu,t}(s) \tag{4.18}$$

for $s \in [0, \delta_{**}]$ *where* $\delta_{**} \sim 1$ *depends on* δ_* *and is chosen in such a way that* $\varphi_{\alpha,t}$ *is invertible*[3]. *We thus define* $n_{\alpha,t}(E)$ *to be the inverse of* $\varphi_{\alpha,t}(s)$ *near zero. Furthermore, for* $0 \le t \le t_*$ *set*

$$\mathfrak{e}^\pm_{\alpha,t} := \alpha\mathfrak{e}^\pm_{\lambda,t} + (1-\alpha)\mathfrak{e}^\pm_{\mu,t}, \tag{4.19}$$

$$\rho_{\alpha,t}(\mathfrak{e}^+_{\alpha,t} + E) := \frac{\mathrm{d}}{\mathrm{d}E} n_{\alpha,t}(E), \quad E \in [0, \delta_*] \tag{4.20}$$

and for $t \ge t_*$ *set*

$$\mathfrak{m}_{\alpha,t} := \alpha\mathfrak{m}_{\lambda,t} + (1-\alpha)\mathfrak{m}_{\mu,t}, \tag{4.21}$$

$$\rho_{\alpha,t}(\mathfrak{m}_{\alpha,t} + E) := \alpha\rho_{\lambda,t}(\mathfrak{m}_{\lambda,t}) + (1-\alpha)\rho_{\mu,t}(\mathfrak{m}_{\mu,t}) + \frac{\mathrm{d}}{\mathrm{d}E} n_{\alpha,t}(E), \quad E \in [-\delta_*, \delta_*].$$

We define $\rho_{\alpha,t}(E)$ *for* $0 \le t \le t_*$ *and* $E \in [\mathfrak{e}^-_{\alpha,t} - \delta_*, \mathfrak{e}^-_{\alpha,t}]$ *analogously.*

[3] Invertibility in a small neighbourhood follows from the form of the explicit shape functions in (4.3b) and (4.3d)

The motivation for the interpolation mode in Definition 4.4.1 is that (4.18) ensures that the quantiles of $\rho_{\alpha,t}$ are the convex combination of the quantiles of $\rho_{\lambda,t}$ and $\rho_{\mu,t}$, see (4.30c) later. The following two lemmas collect various properties of the interpolating density. Recall that $\rho_{\lambda,t}$ and $\rho_{\mu,t}$ are asymptotically close near the cusp regime, up to a trivial shift, since they develop a cusp with the same slope at the same time. In Lemma 4.4.2 we show that $\rho_{\alpha,t}$ shares this property. Lemma 4.4.3 shows that $\rho_{\alpha,t}$ inherits the regularity properties of $\rho_{\lambda,t}$ and $\rho_{\mu,t}$ from [14].

Lemma 4.4.2 (Size of gaps and minima along the flow). *For $t \le t_*$ and $r = \alpha, \lambda, \mu$ the supports of $\rho_{r,t}$ have small gaps $[\mathfrak{e}^-_{r,t}, \mathfrak{e}^+_{r,t}]$ near $\mathfrak{c}_*$ of size*

$$\Delta_{r,t} := \mathfrak{e}^+_{r,t} - \mathfrak{e}^-_{r,t} = (2\gamma)^2\Big(\frac{t_*-t}{3}\Big)^{3/2}\big[1+\mathcal{O}((t_*-t)^{1/3})\big], \quad \Delta_{r,t} = \Delta_{\mu,t}\big[1+\mathcal{O}((t_*-t)^{1/3})\big] \tag{4.22a}$$

and the densities are close in the sense

$$\rho_{r,t}(\mathfrak{e}^\pm_{r,t} \pm \omega) = \rho_{\mu,t}(\mathfrak{e}^\pm_{\mu,t} \pm \omega)\Big[1 + \mathcal{O}((t_*-t)^{1/3} + \min\Big\{\omega^{1/3}, \frac{\omega^{1/2}}{(t_*-t)^{1/4}}\Big\})\Big] \tag{4.22b}$$

for $0 \le \omega \le \delta_*$. For $t_* < t \le 2t_*$ the densities $\rho_{r,t}$ have small local minima $\mathfrak{m}_{r,t}$ of size

$$\rho_{r,t}(\mathfrak{m}_{r,t}) = \frac{\gamma^2\sqrt{t-t_*}}{\pi}\big[1+\mathcal{O}((t-t_*)^{1/2})\big], \qquad \rho_{r,t}(\mathfrak{m}_{r,t}) = \rho_{\mu,t}(\mathfrak{m}_{\mu,t})\big[1+\mathcal{O}((t-t_*)^{1/2})\big] \tag{4.22c}$$

and the densities are close in the sense

$$\frac{\rho_{r,t}(\mathfrak{m}_{r,t}+\omega)}{\rho_{\mu,t}(\mathfrak{m}_{\mu,t}+\omega)} = 1+\mathcal{O}((t-t_*)^{1/2}+\min\Big\{(t-t_*)^{1/4}, \frac{(t-t_*)^2}{|\omega|}\Big\}+\min\Big\{\frac{\omega^2}{(t-t_*)^{5/2}}, |\omega|^{1/3}\Big\}) \tag{4.22d}$$

for $\omega \in [-\delta_*, \delta_*]$. Here $\delta_*, \delta_{**} \sim 1$ are small constants depending on the model parameters in Assumptions (4.A)–(4.C).

Lemma 4.4.3. *The density $\rho_{\alpha,t}$ from Definition 4.4.1 is well defined and is a $1/3$-Hölder continuous density. More precisely, in the pre-cusp regime, i.e. for $t \le t_*$, we have*

$$|\rho'_{\alpha,t}(\mathfrak{e}^\pm_{\alpha,t} \pm x)| \lesssim \frac{1}{\rho_{\alpha,t}(\mathfrak{e}^\pm_{\alpha,t} \pm x)\Big(\rho_{\alpha,t}(\mathfrak{e}^\pm_{\alpha,t} \pm x) + \Delta^{1/3}_{\alpha,t}\Big)} \tag{4.23a}$$

for $0 \le x \le \delta_*$. Moreover, the Stieltjes transform $m_{\alpha,t}$ satisfies the bounds

$$\begin{aligned} |m_{\alpha,t}(\mathfrak{e}^\pm_{\alpha,t} \pm x)| &\lesssim 1, \\ |m_{\alpha,t}(\mathfrak{e}^\pm_{\alpha,t} \pm (x+y)) - m_{\alpha,t}(\mathfrak{e}^\pm_{\alpha,t} \pm x)| &\lesssim \frac{|y||\log|y||}{\rho_{\alpha,t}(\mathfrak{e}^\pm_{\alpha,t} \pm x)(\rho_{\alpha,t}(\mathfrak{e}^\pm_{\alpha,t} \pm x) + \Delta^{1/3}_{\alpha,t})} \end{aligned} \tag{4.23b}$$

for $|x| \le \delta_*/2$, $|y| \ll x$. In the small minimum case, i.e. for $t \ge t_*$, we similarly have

$$|\rho'_{\alpha,t}(\mathfrak{m}_{\alpha,t} + x)| \lesssim \frac{1}{\rho^2_{\alpha,t}(\mathfrak{m}_{\alpha,t} + x)} \tag{4.24a}$$

for $|x| \leq \delta_*$ *and*

$$|m_{\alpha,t}(\mathfrak{m}_{\alpha,t}+x)| \lesssim 1, \quad |m_{\alpha,t}(\mathfrak{m}_{\alpha,t}+(x+y)) - m_{\alpha,t}(\mathfrak{m}_{\alpha,t}+x)| \lesssim \frac{|y||\log|y||}{\rho^2_{\alpha,t}(\mathfrak{m}_{\alpha,t}+x)} \tag{4.24b}$$

for $|x| \leq \delta_*$ *and* $|y| \ll |x|$.

Proof of Lemma 4.4.2. We first consider the two densities $r = \lambda, \mu$ only. The first claims in (4.22a) and (4.22c) follow directly from [83, Lemma 5.1], while the second claims follow immediately from the first ones. For the proof of (4.22b) and (4.22d) we first note that by elementary calculus

$$\Psi_{\mathrm{edge}}((1+\epsilon)\lambda) = \Psi_{\mathrm{edge}}(\lambda)[1+\mathcal{O}(\epsilon)], \qquad \Psi_{\mathrm{min}}((1+\epsilon)\lambda) = \Psi_{\mathrm{min}}(\lambda)[1+\mathcal{O}(\epsilon)]$$

so that

$$\Delta_{\lambda,t}^{1/3}\Psi_{\mathrm{edge}}(\omega/\Delta_{\lambda,t}) = \Delta_{\mu,t}^{1/3}\Psi_{\mathrm{edge}}(\omega/\Delta_{\mu,t})\Big[1+\mathcal{O}((t_*-t)^{1/3})\Big]$$

and the claimed approximations follow together with (4.3b) and (4.3d). Here the exact cusp case $t = t_*$ is also covered by interpreting $0^{1/3}\Psi_{\mathrm{edge}}(\omega/0) = \omega^{1/3}/2^{4/3}$.

In order to prove the corresponding statements for the interpolating densities $\rho_{\alpha,t}$, we first have to establish a quantitative understanding of the counting function $n_{r,t}$ and its inverse. We claim that for $r = \alpha, \lambda, \mu$ they satisfy for $0 \leq E \leq \delta_*$, $0 \leq s \leq \delta_{**}$ that

$$\begin{aligned} n_{r,t}(E) &\sim \min\Big\{\frac{E^{3/2}}{\Delta_{r,t}^{1/6}}, E^{4/3}\Big\}, \qquad \varphi_{r,t}(s) \sim \max\Big\{s^{3/4}, s^{2/3}\Delta_{r,t}^{1/9}\Big\}, \\ \frac{\varphi_{r,t}(s)}{\varphi_{\lambda,t}(s)} &\sim \min\Big\{\varphi_{\lambda,t}^{1/3}(s), \frac{\varphi_{\lambda,t}^{1/2}(s)}{\Delta_{\lambda,t}^{1/6}}\Big\} \end{aligned} \tag{4.25a}$$

for $t \leq t_*$ and

$$\begin{aligned} n_{r,t}(E) &\sim \max\{E^{4/3}, E\rho_{r,t}(\mathfrak{m}_{r,t})\}, \qquad \varphi_{r,t}(s) \sim \min\Big\{s^{3/4}, \frac{s}{\rho_{r,t}(\mathfrak{m}_{r,t})}\Big\} \\ \frac{\varphi_{r,t}(s)}{\varphi_{\lambda,t}(s)} &\sim \min\Big\{\varphi_{\lambda,t}^{1/3}(s), \frac{\varphi_{\lambda,t}(s)}{\rho^2_{r,t}(\mathfrak{m}_{r,t})}, \frac{\varphi^2_{\lambda,t}(s)}{\rho^{11/2}_{r,t}(\mathfrak{m}_{r,t})}\Big\} \end{aligned} \tag{4.25b}$$

for $t \geq t_*$.

Proof of (4.25). We begin with the proof of (4.25a) for $r = \lambda, \mu$. Recall that the shape function Ψ_{edge} satisfies the scaling $\Delta^{1/3}\Psi_{\mathrm{edge}}(\omega/\Delta) \sim \min\{\omega^{1/3}, \omega^{1/2}/\Delta^{1/6}\}$. We first find by elementary integration that

$$\int_0^q \min\Big\{\omega^{1/3}, \frac{\omega^{1/2}}{\Delta^{1/6}}\Big\}\,\mathrm{d}\omega = \frac{9q^{4/3}\min\{q,\Delta\}^{1/6} - \min\{q,\Delta\}^{3/2}}{12\Delta^{1/6}} \sim \min\Big\{\frac{q^{3/2}}{\Delta^{1/6}}, q^{4/3}\Big\}$$

from which we conclude the first relation in (4.25a), and by inversion also the second relation. Together with the estimate for the error integral for $\rho_{\lambda,t}(\mathfrak{e}^+_{\lambda,t}+\omega) - \rho_{\mu,t}(\mathfrak{e}^+_{\mu,t}+\omega) \lesssim \min\{\omega^{2/3}, \omega/\Delta_{\lambda,t}^{1/3}\}$,

$$\int_0^q \min\Big\{\omega^{2/3}, \frac{\omega}{\Delta^{1/3}}\Big\}\,\mathrm{d}\omega = \frac{6q^{5/3}\min\{q,\Delta\}^{1/3} - \min\{q,\Delta\}^2}{10\Delta^{1/3}} \sim \min\Big\{\frac{q^2}{\Delta^{1/3}}, q^{5/3}\Big\}$$

we can thus conclude also the third relation in (4.25a).

We now turn to the case $t > t_*$ where both densities $\rho_{\lambda,t}, \rho_{\mu,t}$ exhibit a small local minimum. We first record the elementary integral

$$\int_0^q \Big(\rho + \min\Big\{\omega^{1/3}, \frac{\omega^2}{\rho^5}\Big\}\Big)\,\mathrm{d}\omega = \frac{q^{4/3}\min\{\rho^3,q\}^{5/3} + 12q\rho^6 - 5\min\{q,\rho^3\}^3}{12\rho^5} \sim \max\{q^{4/3}, q\rho\}$$

for $q,\rho \geq 0$ and easily conclude the first two relation in (4.25b). For the error integral we obtain

$$\int_0^q \min\Big\{\omega^{1/3}, \frac{\omega^2}{\rho^5}\Big\}\Big[\min\Big\{\rho^{1/2}, \frac{\rho^4}{\omega}\Big\} + \min\Big\{\omega^{1/3}, \frac{\omega^2}{\rho^5}\Big\}\Big]\,\mathrm{d}\omega \sim \min\Big\{q^{5/3}, \frac{q^2}{\rho}, \frac{q^3}{\rho^{9/2}}\Big\}$$

from which the third relation in (4.25b) follows. Finally, the claims (4.25a) and (4.25b) for $r = \alpha$ follow immediately from Definition 4.4.1 and the corresponding statements for $r = \lambda, \mu$. This completes the proof of (4.25). □

We now turn to the density $\rho_{\alpha,t}$ for which the claims (4.22a), (4.22c) follow immediately from Definition 4.4.1 and the corresponding statements for $\rho_{\lambda,t}$ and $\rho_{\mu,t}$. For $t \leq t_*$ we now continue by differentiating $E = \varphi_{r,t}(n_{r,t}(E))$ to obtain

$$\begin{aligned}\rho_{\alpha,t}(\mathfrak{e}^+_{\alpha,t} + \varphi_{\alpha,t}(s)) &= \frac{1}{\varphi'_{\alpha,t}(s)} = \frac{1}{\alpha\varphi'_{\lambda,t}(s) + (1-\alpha)\varphi'_{\mu,t}(s)} \\ &= \Big(\frac{\alpha}{\rho_{\lambda,t}(\mathfrak{e}^+_{\lambda,t} + \varphi_{\lambda,t}(s))} + \frac{1-\alpha}{\rho_{\mu,t}(\mathfrak{e}^+_{\mu,t} + \varphi_{\mu,t}(s))}\Big)^{-1} \\ &= \rho_{\lambda,t}(\mathfrak{e}^+_{\lambda,t} + \varphi_{\lambda,t}(s))\Big(\alpha + (1-\alpha)\frac{\rho_{\lambda,t}(\mathfrak{e}^+_{\lambda,t} + \varphi_{\lambda,t}(s))}{\rho_{\mu,t}(\mathfrak{e}^+_{\mu,t} + \varphi_{\mu,t}(s))}\Big)^{-1}, \qquad (4.26)\end{aligned}$$

from which we can easily conclude (4.22b) for $r = \alpha$ together with (4.22b) for $r = \lambda$ and (4.25a). The proof of (4.22d) for $r = \alpha$ follows by the same argument and replacing $\mathfrak{e}^+_{r,t}$ by $\mathfrak{m}_{r,t}$. This finishes the proof of Lemma 4.4.2 □

Proof of Lemma 4.4.3. By differentiating we find

$$\begin{aligned}\frac{\rho'_{\alpha,t}(\mathfrak{e}^+_{\alpha,t} + \varphi_{\alpha,t}(s))}{\rho_{\alpha,t}(\mathfrak{e}^+_{\alpha,t} + \varphi_{\alpha,t}(s))} &= -\frac{\alpha\varphi''_{\lambda,t}(s) + (1-\alpha)\varphi''_{\mu,t}(s)}{\Big(\alpha\varphi'_{\lambda,t}(s) + (1-\alpha)\varphi'_{\mu,t}(s)\Big)^2} \\ &= \Big[\alpha\frac{\rho'_{\lambda,t}(\mathfrak{e}^+_{\lambda,t} + \varphi_{\lambda,t}(s))}{\rho^3_{\lambda,t}(\mathfrak{e}^+_{\lambda,t} + \varphi_{\lambda,t}(s))} + (1-\alpha)\frac{\rho'_{\mu,t}(\mathfrak{e}^+_{\mu,t} + \varphi_{\mu,t}(s))}{\rho^3_{\mu,t}(\mathfrak{e}^+_{\mu,t} + \varphi_{\mu,t}(s))}\Big] \\ &\quad \times \Big(\frac{\alpha}{\rho_{\lambda,t}(\mathfrak{e}^+_{\lambda,t} + \varphi_{\lambda,t}(s))} + \frac{1-\alpha}{\rho_{\mu,t}(\mathfrak{e}^+_{\mu,t} + \varphi_{\mu,t}(s))}\Big)^{-2},\end{aligned}$$

from which we conclude the claimed bound (4.23a) together with the fact that the densities ρ_λ and ρ_μ fulfil the same bound according to [14, Remark 10.7], and the estimates from Lemma 4.4.2. Similarly, the bound in (4.24a) follows by the same argument by replacing $\mathfrak{e}^\pm_{\alpha,t}$ by $\mathfrak{m}_{\alpha,t}$. The bound $|\rho'| \leq \rho^{-2}$ on the derivative implies $\frac{1}{3}$-Hölder continuity.

We now turn to the claimed bound on the Stieltjes transform and compute

$$m_{\alpha,t}(\mathfrak{e}^+_{\alpha,t}+x)=\int_0^{\delta_*}\frac{\rho_{\alpha,t}(\mathfrak{e}^+_{\alpha,t}+\omega)}{\omega-x}\,\mathrm{d}\omega+\int_{-\delta_*}^0\frac{\rho_{\alpha,t}(\mathfrak{e}^-_{\alpha,t}+\omega)}{\omega-\Delta_{\alpha,t}-x}\,\mathrm{d}\omega,$$

out of which for $x>0$ the first term can be bounded by

$$\begin{aligned}\int_0^{\delta_*}\frac{\rho_{\alpha,t}(\mathfrak{e}^+_{\alpha,t}+\omega)}{\omega-x}\,\mathrm{d}\omega&\lesssim\int_0^{\delta_*}\frac{|\omega-x|^{1/3}}{\omega-x}\,\mathrm{d}\omega+\int_{2x}^{\delta_*}\frac{\rho_{\alpha,t}(\mathfrak{e}^+_{\alpha,t}+x)}{\omega-x}\,\mathrm{d}\omega\\&\lesssim|x|^{1/3}|\log x|+|\delta_*-x|^{1/3},\end{aligned}$$

while the second term can be bounded by

$$|\int_{-\delta_*}^0\frac{\rho_{\alpha,t}(\mathfrak{e}^-_{\alpha,t}+\omega)}{\omega-\Delta_{\alpha,t}-x}\,\mathrm{d}\omega|\lesssim|\delta_*-\Delta_{\alpha,t}-x|^{1/3}+|\Delta_{\alpha,t}+x|^{1/3}|\log(\Delta_{\alpha,t}+x)|,$$

both using the $1/3$-Hölder continuity of $\rho_{\alpha,t}$. The corresponding bounds for $x<0$ are similar, completing the proof of the first bound in (4.23b).

The proof of the first bound in (4.24b) is very similar and follows from

$$|m_{\alpha,t}(\mathfrak{m}_{\alpha,t}+x)|\lesssim|\int_{-\delta_*}^{\delta_*}\frac{|\omega-x|^{1/3}}{\omega-x}\,\mathrm{d}\omega|+|\int_{[-\delta_*,\delta_*]\setminus[x-\delta_*/2,x+\delta_*/2]}\frac{\rho_{\alpha,t}(\mathfrak{m}_{\alpha,t}+x)}{\omega-x}\,\mathrm{d}\omega|\lesssim1.$$

We now turn to the second bound in (4.23b) which is only non-trivial in the case $x>0$. To simplify the following integrals we temporarily use the short-hand notations $m=m_{\alpha,t}$, $\mathfrak{e}^+=\mathfrak{e}^+_{\alpha,t}$, $\rho=\rho_{\alpha,t}$, $\Delta=\Delta_{\alpha,t}$ and compute

$$m(\mathfrak{e}^++x+y)-m(\mathfrak{e}^++x)=\int_{-\Delta-\delta_*}^{\delta_*}\frac{\rho(\mathfrak{e}^++\omega)}{\omega-x-y}\,\mathrm{d}\omega-\int_{-\Delta-\delta_*}^{\delta_*}\frac{\rho(\mathfrak{e}^++\omega)}{\omega-x}\,\mathrm{d}\omega$$

where we now focus on the integration regime $\omega\geq0$ as this is the regime containing the two critical singularities. We first observe that

$$\begin{aligned}\int_{-y}^{\delta_*-y}\frac{\rho(\mathfrak{e}^++\omega+y)}{\omega-x}\,\mathrm{d}\omega-\int_0^{\delta_*}\frac{\rho(\mathfrak{e}^++\omega)}{\omega-x}\,\mathrm{d}\omega&=\int_0^{\delta_*}\frac{\rho(\mathfrak{e}^++\omega+y)-\rho(\mathfrak{e}^++\omega)}{\omega-x}\,\mathrm{d}\omega\\&\quad+\int_{-y}^0\frac{\rho(\mathfrak{e}^++\omega+y)}{\omega-x}\,\mathrm{d}\omega+\mathcal{O}(y),\end{aligned}$$

where the second integral is easily bounded by

$$\int_{-y}^0\frac{\rho(\mathfrak{e}+\omega+y)}{\omega-x}\,\mathrm{d}\omega\lesssim\frac{1}{x}\min\left\{y^{4/3},y^{3/2}\Delta^{-1/6}\right\}\lesssim\frac{y}{\rho(\mathfrak{e}^++x)(\rho(\mathfrak{e}^++x)+\Delta^{1/3})}.$$

We split the remaining integral into three regimes $[0,x/2]$, $[x/2,3x/2]$ and $[3x/2,\delta_*]$. In the first one we use (4.23a) as well as the scaling relation $\rho(\mathfrak{e}^++\omega)\sim\min\{\omega^{1/3},\omega^{1/2}\Delta^{-1/6}\}$ to obtain

$$\begin{aligned}\int_0^{x/2}\frac{\rho(\mathfrak{e}^++\omega+y)-\rho(\mathfrak{e}^++\omega)}{\omega-x}\,\mathrm{d}\omega&\lesssim\frac{y}{x}\int_0^{x/2}\frac{1}{\rho(\mathfrak{e}^++\omega)\Big(\rho(\mathfrak{e}^++\omega)+\Delta^{1/3}\Big)}\,\mathrm{d}\omega\\&\lesssim\frac{y}{x}\min\left\{\frac{x^{1/2}}{\Delta^{1/6}},x^{1/3}\right\}\sim\frac{y}{\max\{x^{2/3},x^{1/2}\Delta^{1/6}\}}\lesssim\frac{y}{\rho(\mathfrak{e}^++x)(\rho(\mathfrak{e}^++x)+\Delta^{1/3})}.\end{aligned}$$

The integral in the regime $[3x/2, \delta_*]$ is completely analogous and contributes the same bound. Finally, we are left with the regime $[x/2, 3x/2]$ which we again subdivide into $[x-y, x+y]$ and $[x/2, 3x/2] \setminus [x-y, x+y]$. In the first of those we have

$$\begin{aligned}
&\int_{x-y}^{x+y} \frac{\rho(\mathfrak{e}^+ + \omega + y) - \rho(\mathfrak{e}^+ + \omega)}{\omega - x} \,\mathrm{d}\omega \\
&\quad = \int_{x-y}^{x+y} \frac{\rho(\mathfrak{e}^+ + \omega + y) - \rho(\mathfrak{e}^+ + x + y) - \rho(\mathfrak{e}^+ + \omega) + \rho(\mathfrak{e}^+ + x)}{\omega - x} \,\mathrm{d}\omega \\
&\quad \lesssim \frac{y}{\rho(\mathfrak{e}^+ + x)(\rho(\mathfrak{e}^+ + x) + \Delta^{1/3})},
\end{aligned}$$

while in the second one we obtain

$$\begin{aligned}
&\int_{[x/2,3x/2]\setminus[x-y,x+y]} \frac{\rho(\mathfrak{e}^+ + \omega + y) - \rho(\mathfrak{e}^+ + x + y) - \rho(\mathfrak{e}^+ + \omega) + \rho(\mathfrak{e}^+ + x)}{\omega - x} \,\mathrm{d}\omega \\
&\lesssim \frac{|y|}{\rho(\mathfrak{e}^+ + x)(\rho(\mathfrak{e}^+ + x) + \Delta^{1/3})} \int_{[x/2,3x/2]\setminus[x-y,x+y]} |\omega - x|^{-1} \,\mathrm{d}\omega \\
&\lesssim \frac{|y||\log y|}{\rho(\mathfrak{e}^+ + x)(\rho(\mathfrak{e}^+ + x) + \Delta^{1/3})}.
\end{aligned}$$

Collecting the various estimates completes the proof of (4.23b).

The second bound in (4.24b) follows by a similar argument and we focus on the most critical term

$$\begin{aligned}
&\int_{-\delta_*/2}^{\delta_*/2} \frac{\rho(\mathfrak{m} + \omega + y) - \rho(\mathfrak{m} + \omega)}{\omega - x} \,\mathrm{d}\omega \\
&\quad = \Big(\int_{-\delta_*/2}^{x-y} + \int_{x-y}^{x+y} + \int_{x+y}^{\delta_*/2} \Big) \frac{\rho(\mathfrak{m} + \omega + y) - \rho(\mathfrak{m} + \omega)}{\omega - x} \,\mathrm{d}\omega.
\end{aligned}$$

Here we can bound the middle integral by

$$\begin{aligned}
&\Big| \int_{x-y}^{x+y} \frac{\rho(\mathfrak{m} + \omega + y) - \rho(\mathfrak{m} + \omega)}{\omega - x} \,\mathrm{d}\omega \Big| \\
&\quad = \Big| \int_{x-y}^{x+y} \frac{\rho(\mathfrak{m} + \omega + y) - \rho(\mathfrak{m} + x + y) - \rho(\mathfrak{m} + \omega) + \rho(\mathfrak{m} + x)}{\omega - x} \,\mathrm{d}\omega \Big| \\
&\quad \lesssim \frac{|y|}{\rho^2(\mathfrak{m} + x)},
\end{aligned}$$

while for the first integral we have

$$\begin{aligned}
&\Big| \int_{-\delta_*/2}^{x-y} \frac{\rho(\mathfrak{m} + \omega + y) - \rho(\mathfrak{m} + x + y) - \rho(\mathfrak{m} + \omega) + \rho(\mathfrak{m} + x)}{\omega - x} \,\mathrm{d}\omega \Big| \\
&\quad \lesssim \frac{|y|}{\rho^2(\mathfrak{m} + x)} \int_{-\delta_*/2}^{x-y} \frac{1}{|\omega - x|} \,\mathrm{d}\omega \lesssim \frac{|y||\log|y||}{\rho^2(\mathfrak{m} + x)}.
\end{aligned}$$

The third integral is completely analogous, completing the proof of (4.24b). □

4.4.1 Quantiles

Finally we consider the locations of quantiles of $\rho_{r,t}$ for $r = \alpha, \lambda, \mu$ and their fluctuation scales. For $0 \leq t \leq t_*$ we define the shifted quantiles $\widehat{\gamma}_{r,i}(t)$, and for $t_* \leq t \leq 2t_*$ the shifted quantiles[4] $\widecheck{\gamma}_{r,i}(t)$ in such a way that

$$\int_0^{\widehat{\gamma}_{r,i}(t)} \rho_{r,t}(\mathfrak{e}_{r,t}^{+}+\omega)\,\mathrm{d}\omega = \frac{i}{N}, \qquad \int_0^{\widecheck{\gamma}_{r,i}(t)} \rho_{r,t}(\mathfrak{m}_{r,t}+\omega)\,\mathrm{d}\omega = \frac{i}{N}, \qquad |i| \ll N. \quad (4.27)$$

Notice that for $i = 0$ we always have $\widehat{\gamma}_{r,0}(t) = \widecheck{\gamma}_{r,0}(t) = 0$. We will also need to define the semiquantiles, distinguished by star from the quantiles, defined as follows:

$$\int_0^{\widehat{\gamma}^*_{r,i}(t)} \rho_{r,t}(\mathfrak{e}_{r,t}^{+}+\omega)\,\mathrm{d}\omega = \frac{i-\frac{1}{2}}{N}, \quad \int_0^{\widecheck{\gamma}^*_{r,i}(t)} \rho_{r,t}(\mathfrak{m}_{r,t}+\omega)\,\mathrm{d}\omega = \frac{i-\frac{1}{2}}{N}, \quad 1 \leq i \ll N \quad (4.28)$$

and

$$\int_0^{\widehat{\gamma}^*_{r,i}(t)} \rho_{r,t}(\mathfrak{e}_{r,t}^{+}+\omega)\,\mathrm{d}\omega = \frac{i+\frac{1}{2}}{N}, \quad \int_0^{\widecheck{\gamma}^*_{r,i}(t)} \rho_{r,t}(\mathfrak{m}_{r,t}+\omega)\,\mathrm{d}\omega = \frac{i+\frac{1}{2}}{N}, \quad -N \ll i \leq -1 \quad (4.29)$$

Note that the definition is slightly different for positive and negative i's, in particular $\widehat{\gamma}_i^* \in [\widehat{\gamma}_{i-1}, \widehat{\gamma}_i]$ for $i \geq 1$ and $\widehat{\gamma}_i^* \in [\widehat{\gamma}_i, \widehat{\gamma}_{i+1}]$ for $i < 0$. The semiquantiles are not defined for $i = 0$.

Lemma 4.4.4. *For $1 \leq |i| \ll N$, $r = \alpha, \lambda, \mu$ and $0 \leq t \leq t_*$ we have*

$$\begin{aligned} \widehat{\gamma}_{r,i}(t) &\sim \operatorname{sgn}(i) \max\Big\{ \Big(\frac{|i|}{N}\Big)^{3/4}, \Big(\frac{|i|}{N}\Big)^{2/3} (t_* - t)^{1/6} \Big\} - \begin{cases} 0, & i > 0 \\ \Delta_{r,t}, & i < 0 \end{cases} \\ \widehat{\gamma}_{r,i}(t) &= \widehat{\gamma}_{\mu,i}(t)\Big[1 + \mathcal{O}\Big((t_* - t)^{1/3} + \min\Big\{ \frac{\widehat{\gamma}_{\mu,i}(t)^{1/2}}{(t_* - t)^{1/4}}, \widehat{\gamma}_{\mu,i}(t)^{1/3} \Big\}\Big)\Big], \end{aligned} \quad (4.30a)$$

while for $t_* \leq t \leq 2t_*$ we have

$$\begin{aligned} \widecheck{\gamma}_{r,i}(t) &\sim \operatorname{sgn}(i) \min\Big\{ \Big(\frac{|i|}{N}\Big)^{3/4}, \frac{|i|}{N} (t_* - t)^{-1/2} \Big\}, \\ \widecheck{\gamma}_{r,i}(t) &= \widecheck{\gamma}_{\mu,i}(t)\Big[1 + \mathcal{O}\Big((t_* - t)^{1/2} + \min\Big\{ \frac{\widecheck{\gamma}_{\mu,i}(t)^2}{(t_* - t)^{11/4}}, \frac{\widecheck{\gamma}_{\mu,i}(t)}{t_* - t}, \widecheck{\gamma}_{\mu,i}(t)^{1/3} \Big\}\Big)\Big]. \end{aligned} \quad (4.30b)$$

Moreover, the quantiles of $\rho_{\alpha,t}$ are the convex combination

$$\widehat{\gamma}_{\alpha,i}(t) = \alpha\widehat{\gamma}_{\lambda,i}(t) + (1-\alpha)\widehat{\gamma}_{\mu,i}(t), \quad \widecheck{\gamma}_{\alpha,i}(t) = \alpha\widecheck{\gamma}_{\lambda,i}(t) + (1-\alpha)\widecheck{\gamma}_{\mu,i}(t). \quad (4.30c)$$

Proof. The proof follows directly from the estimates in (4.25a) and (4.25b). The relation (4.30c) follows directly from (4.18) in the definition of the α-interpolating density. □

[4] We use a separate variable name $\widecheck{\gamma}$ because in Section 4.8 the name $\widehat{\gamma}$ is used for the quantiles with respect to the base point $\widetilde{\mathfrak{m}}$ instead of $\mathfrak{m}$.

4.4.2 Movement of edges, quantiles and minima

For the analysis of the Dyson Brownian motion it is necessary to have a precise understanding of the movement of the reference points $\mathfrak{e}_{r,t}^{\pm}$ and $\mathfrak{m}_{r,t}$, $r = \lambda, \mu$. For technical reasons it is slightly easier to work with an auxiliary quantity $\widetilde{\mathfrak{m}}_{r,t}$ which is very close to $\mathfrak{m}_{r,t}$. According to [83, Lemma 5.1] the minimum $\mathfrak{m}_{r,t}$ can approximately be found by solving the implicit equation

$$\widetilde{\mathfrak{m}}_{r,t} = \mathfrak{c}_r - (t - t_*)\Re m_{r,t}(\widetilde{\mathfrak{m}}_{r,t}), \qquad \widetilde{\mathfrak{m}}_{r,t} \in \mathbf{R}, \quad r = \lambda, \mu. \tag{4.31a}$$

The explicit relation (4.31a) is the main reason why it is more convenient to study the movement of $\widetilde{\mathfrak{m}}_t$ rather than the one of $\mathfrak{m}_t$. We claim that $\widetilde{\mathfrak{m}}_{r,t}$ is indeed a very good approximation for $\mathfrak{m}_{r,t}$ in the sense that

$$|\mathfrak{m}_{r,t} - \widetilde{\mathfrak{m}}_{r,t}| \lesssim (t - t_*)^{3/2+1/4}, \quad \Im m_{r,t}(\widetilde{\mathfrak{m}}_{r,t}) = \gamma^2 (t - t_*)^{1/2} + \mathcal{O}(t - t_*), \quad r = \lambda, \mu. \tag{4.31b}$$

Proof of (4.31b). The first claim in (4.31b) is a direct consequence of [83, Lemma 5.1]. For the second claim we refer to [83, Eq. (89a)] which implies

$$\Im m_{r,t}(\widetilde{\mathfrak{m}}_{r,t}) = (t-t_*)^{1/2}\gamma^2\Big[1+\mathcal{O}((t-t_*)^{1/3}[\Im m_{r,t}(\widetilde{\mathfrak{m}}_{r,t})]^{1/3})\Big] = \gamma^2(t-t_*)^{1/2}+\mathcal{O}(t-t_*).$$

□

For the t-derivative of (semi-)quantiles $\gamma_{r,t}$, i.e. points such that $\int_{-\infty}^{\gamma_{r,t}} \rho_{r,t}(x)\,\mathrm{d}x$ is constant in t, as well as for the minima $\widetilde{\mathfrak{m}}_{r,t}$ we have the explicit relations

$$\frac{\mathrm{d}}{\mathrm{d}t}\gamma_{r,t} = -\Re m_{r,t}(\gamma_{r,t}), \tag{4.31c}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\mathfrak{m}}_{r,t} = -\Re m_{r,t}(\widetilde{\mathfrak{m}}_{r,t}) + \mathcal{O}(t - t_*), \quad t_* \le t \le 2t_*. \tag{4.31d}$$

In particular, for the spectral edges it follows from (4.31c) that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathfrak{e}_{r,t}^{+} = -m_{r,t}(\mathfrak{e}_{r,t}^{+}), \quad 0 \le t \le t_*. \tag{4.31e}$$

Proof of (4.31c)–(4.31e). For the proof of (4.31c) we first recall that from the defining equation (4.17) of the semicircular flow it follows that the Stieltjes transform $m = m_t(\zeta)$ of ρ_t satisfies the Burgers equation

$$\dot m = m m' = \frac{1}{2}(m^2)', \tag{4.32}$$

where prime denotes the $\frac{\mathrm{d}}{\mathrm{d}\zeta}$ derivative and dot denotes the $\frac{\mathrm{d}}{\mathrm{d}t}$ derivative. Thus

$$\begin{aligned}\dot\gamma_{r,t} = -\frac{1}{\rho_{r,t}(\gamma_{r,t})}\Im\int_{-\infty}^{\gamma_{r,t}} \dot m_{r,t}(E)\,\mathrm{d}E \\ &= -\frac{1}{2\rho_{r,t}(\gamma_{r,t})}\Im\int_{-\infty}^{\gamma_{r,t}} (m_{r,t}^2)'(E)\,\mathrm{d}E \\ &= -\frac{\Im m_{r,t}^2(\gamma_{r,t})}{2\Im m_{r,t}(\gamma_{r,t})} = -\Re m_{r,t}(\gamma_{r,t}).\end{aligned}$$

follows directly from differentiating $\int_{-\infty}^{\gamma_{r,t}} \rho_{r,t}(x)\,\mathrm{d}x \equiv \text{const}$.

For (4.31d) we begin by computing the integral

$$m'_{r,t_*}(\mathfrak{c}_r+\mathrm{i}\eta) = \int_{\mathbf{R}} \frac{\rho_{t_*}(\mathfrak{c}_r+x)}{(x-\mathrm{i}\eta)^2}\,\mathrm{d}x = \int_{\mathbf{R}} \frac{\sqrt{3}\gamma^{4/3}|x|^{1/3} + \mathcal{O}(|x|^{2/3})}{2\pi(x-\mathrm{i}\eta)^2}\,\mathrm{d}x = \frac{\gamma^{4/3}}{3\eta^{2/3}} + \mathcal{O}(\eta^{-1/3}), \tag{4.33}$$

so that by definition $m_{r,t}(z) = m_{r,t_*}(z+(t-t_*)m_{r,t}(z))$ of the free semicircular flow,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} m_{r,t}(\widetilde{\mathfrak{m}}_{r,t}) &= m'_{r,t_*}(\widetilde{\mathfrak{m}}_{r,t} + (t-t_*)m_{r,t}(\widetilde{\mathfrak{m}}_{r,t}))\Big[\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\mathfrak{m}}_{r,t} + m_{r,t}(\widetilde{\mathfrak{m}}_{r,t}) + (t-t_*)\frac{\mathrm{d}}{\mathrm{d}t} m_{r,t}(\widetilde{\mathfrak{m}}_{z,t})\Big] \\
&= \Big(\frac{1}{3(t-t_*)} + \mathcal{O}((t-t_*)^{-1/2})\Big)\Big[\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\mathfrak{m}}_{r,t} + m_{r,t}(\widetilde{\mathfrak{m}}_{r,t}) + (t-t_*)\frac{\mathrm{d}}{\mathrm{d}t} m_{r,t}(\widetilde{\mathfrak{m}}_{r,t})\Big] \\
&= \mathrm{i}\Big(\frac{1}{3(t-t_*)} + \mathcal{O}((t-t_*)^{-1/2})\Big)\Big[\Im m_{r,t}(\widetilde{\mathfrak{m}}_{r,t}) + (t-t_*)\frac{\mathrm{d}}{\mathrm{d}t}\Im m_{r,t}(\widetilde{\mathfrak{m}}_{r,t})\Big] \\
&= \Big(\mathrm{i}\frac{\gamma^2}{3(t-t_*)^{1/2}} + \frac{\mathrm{i}}{3}\frac{\mathrm{d}}{\mathrm{d}t}\Im m_{r,t}(\widetilde{\mathfrak{m}}_{r,t})\Big)\Big[1 + \mathcal{O}((t-t_*)^{1/2})\Big].
\end{aligned}$$

Here we used (4.31a), (4.31b) together with (4.33) in the second step. The third step follows from taking the t-derivative of (4.31a). The ultimate inequality is again a consequence of (4.31b). By considering real and imaginary part separately it thus follows that

$$\frac{\mathrm{d}}{\mathrm{d}t}\Im m_{r,t}(\widetilde{\mathfrak{m}}_{r,t}) = \frac{\gamma^2}{2(t-t_*)^{1/2}}\Big[1+\mathcal{O}((t-t_*)^{1/2})\Big], \qquad \frac{\mathrm{d}}{\mathrm{d}t}\Re m_{r,t}(\widetilde{\mathfrak{m}}_{r,t}) = \mathcal{O}(1)$$

and therefore (4.31d) follows by differentiating (4.31a). □

4.4.3 Rigidity scales

In this section we compute, up to leading order, the fluctuations of the eigenvalues around their classical locations, i.e. the quantiles defined in Section 4.4.1. Indeed, the computation of the fluctuation scale for the particles $x_i(t)$, $y_i(t)$, defined in (4.49), (4.51), will be one of the fundamental inputs to prove rigidity for the interpolated process in Section 4.6. The fluctuation scale $\eta_{\mathrm{f}}^{\rho}(\tau)$ of any density function $\rho(\omega)$ around τ is defined via

$$\int_{\tau-\eta_{\mathrm{f}}^{\rho}(\tau)}^{\tau+\eta_{\mathrm{f}}^{\rho}(\tau)} \rho(\omega)\,\mathrm{d}\omega = \frac{1}{N}$$

for $\tau \in \operatorname{supp}\rho$ and by the value $\eta_{\mathrm{f}}(\tau) := \eta_{\mathrm{f}}(\tau')$ where $\tau' \in \operatorname{supp}\rho$ is the edge closest to τ for $\tau \notin \operatorname{supp}\rho$. If this edge is not unique, an arbitrary choice can be made between the two possibilities. From (4.30a) we immediately obtain for $0 \le t \le t_*$ and $1 \le i \le N$, that

$$\eta_{\mathrm{f}}^{\rho_{r,t}}(\mathfrak{e}_{r,t}^{+} + \widehat{\gamma}_{r,\pm i}(t)) \sim \max\Big\{\frac{\Delta_{r,t}^{1/9}}{N^{2/3}i^{1/3}}, \frac{1}{N^{3/4}i^{1/4}}\Big\} \sim \max\Big\{\frac{(t_*-t)^{1/6}}{N^{2/3}i^{1/3}}, \frac{1}{N^{3/4}i^{1/4}}\Big\}, \tag{4.34a}$$

for $r = \alpha, \lambda, \mu$, while for $t_* \le t \le 2t_*$, $1 \le |i| \ll N$ we obtain from (4.30b) that

$$\eta_{\mathrm{f}}^{\rho_{r,t}}(\mathfrak{m}_{r,t}+\check{\gamma}_{r,i}(t)) \sim \min\Big\{\frac{1}{N\rho_{r,t}(\mathfrak{m}_{r,t})}, \frac{1}{N^{3/4}|i|^{1/4}}\Big\} \sim \min\Big\{\frac{1}{N(t-t_*)^{1/2}}, \frac{1}{N^{3/4}|i|^{1/4}}\Big\}, \tag{4.34b}$$

for $r=\alpha,\lambda,\mu$. In the second relations we used (4.22a) and (4.22c). For reference purposes we also list for $0<i,j\ll N$ the bounds

$$|\widehat{\gamma}_{r,i}(t)-\widehat{\gamma}_{r,j}(t)|\sim\max\Big\{\frac{\Delta_{r,t}^{1/9}|i-j|}{N^{2/3}(i+j)^{1/3}},\frac{|i-j|}{N^{3/4}(i+j)^{1/4}}\Big\},\tag{4.35}$$

in case $t\le t_*$ and

$$|\widecheck{\gamma}_{r,i}(t)-\widecheck{\gamma}_{r,j}(t)|\sim\min\Big\{\frac{|i-j|}{\rho_{r,t}(\mathfrak{m}_{r,t})N},\frac{|i-j|}{N^{3/4}(i+j)^{1/4}}\Big\}\tag{4.36}$$

in case $t>t_*$. Furthermore we have

$$\rho_{r,t}(\mathfrak{e}_{r,t}^{+}+\widehat{\gamma}_{r,i}(t))\sim\min\Big\{\frac{i^{1/3}}{N^{1/3}(t_*-t)^{1/6}},\frac{i^{1/4}}{N^{1/4}}\Big\}\tag{4.37}$$

and

$$\rho_{r,t}(\mathfrak{m}_{r,t}+\widecheck{\gamma}_{r,i}(t))\sim\max\Big\{\rho_{r,t}(\mathfrak{m}_{r,t}),\frac{i^{1/4}}{N^{1/4}}\Big\}.\tag{4.38}$$

4.4.4 Stieltjes transform bounds

It follows from (4.22b) and (4.22d) that also the real parts of the Stieltjes transforms $m_{\alpha,t}$, $m_{\lambda,t}$, $m_{\mu,t}$ are close. We claim that for $r=\lambda,\alpha$, $\nu\in[-\delta_*,\delta_*]$ and $0\le t\le t_*$ we have

$$\begin{aligned}\Big|\Re\Big[\Big(m_{r,t}(\mathfrak{e}_{r,t}^{+}+\nu)-m_{r,t}(\mathfrak{e}_{r,t}^{+})\Big)-\Big(m_{\mu,t}(\mathfrak{e}_{\mu,t}^{+}+\nu)-m_{\mu,t}(\mathfrak{e}_{\mu,t}^{+})\Big)\Big]\Big|\\ \lesssim|\nu|^{1/3}\Big[|\nu|^{1/3}+(t_*-t)^{1/3}\Big]|\log|\nu||+(t_*-t)^{11/18}\mathbf{1}(\nu\le-\Delta_{\mu,t}/2),\end{aligned}\tag{4.39a}$$

while for $t_*\le t\le 2t_*$ we have

$$\begin{aligned}\Big|\Re\Big[\Big(m_{r,t}(\mathfrak{m}_{r,t}+\nu)-m_{r,t}(\mathfrak{m}_{r,t})\Big)-\Big(m_{\mu,t}(\mathfrak{m}_{\mu,t}+\nu)-m_{\mu,t}(\mathfrak{m}_{\mu,t})\Big)\Big]\Big|\\ \lesssim\Big[|\nu|^{1/3}(t-t_*)^{1/4}+(t_*-t)^{3/4}+|\nu|^{2/3}\Big]|\log|\nu||.\end{aligned}\tag{4.39b}$$

Proof of (4.39). We first recall from Lemma 4.4.3 that also the density $\rho_{\alpha,t}$ is $1/3$-Hölder continuous which we will use repeatedly in the following proof. We begin with the proof of (4.39a) and compute for $r=\alpha,\lambda,\mu$

$$\begin{aligned}\Re\Big[m_{r,t}(\mathfrak{e}_{r,t}^{+}+\nu)-m_{r,t}(\mathfrak{e}_{r,t}^{+})\Big]&=\int_0^\infty\frac{\nu\rho_{r,t}(\mathfrak{e}_{r,t}^{+}+\omega)}{(\omega-\nu)\omega}\,\mathrm{d}\omega\\&\quad+\int_0^\infty\frac{\nu\rho_{r,t}(\mathfrak{e}_{r,t}^{-}-\omega)}{(\omega+\Delta_{r,t}+\nu)(\omega+\Delta_{r,t})}\,\mathrm{d}\omega.\end{aligned}\tag{4.40}$$

For $\nu>0$ the first of the two terms is the more critical one. Our goal is to obtain a bound on

$$\int_0^\infty\frac{\nu}{(\omega-\nu)\omega}\Big[\rho_{\lambda,t}(\mathfrak{e}_{\lambda,t}^{+}+\omega)-\rho_{\mu,t}(\mathfrak{e}_{\mu,t}^{+}+\omega)\Big]\,\mathrm{d}\omega$$

by using (4.22b). Let $0 < \epsilon < \nu/2$ be a small parameter for which we separately consider the two critical regimes $0 \le \omega \le \epsilon$ and $|\nu - \omega| \le \epsilon$. We use

$$\rho_{r,t}(\mathfrak{e}^+_{r,t} + \omega) \lesssim \omega^{1/3} \quad \text{and} \quad \rho_{r,t}(\mathfrak{e}^+_{r,t} + \omega) = \rho_{r,t}(\mathfrak{e}^+_{r,t} + \nu) + \mathcal{O}(|\omega - \nu|^{1/3}), \quad r = \lambda, \mu, \tag{4.41}$$

from the $1/3$-Hölder continuity of $\rho_{r,t}$ and the fact that the integral over $1/(\omega - \nu)$ from $\nu - \epsilon$ to $\nu + \epsilon$ vanishes by symmetry to estimate, for $r = \lambda, \mu$,

$$\Big| \int_0^\epsilon \frac{\nu}{(\omega - \nu)\omega} \rho_{r,t}(\mathfrak{e}^+_{r,t} + \omega) \, \mathrm{d}\omega \Big| \lesssim \int_0^\epsilon |\omega|^{-2/3} \, \mathrm{d}\omega \lesssim \epsilon^{1/3}$$

and

$$\Big| \int_{\nu-\epsilon}^{\nu+\epsilon} \Big[\frac{\rho_{r,t}(\mathfrak{e}^+_{r,t} + \omega)}{\omega - \nu} - \frac{\rho_{r,t}(\mathfrak{e}^+_{r,t} + \omega)}{\omega} \Big] \mathrm{d}\omega \Big| \lesssim \int_{\nu-\epsilon}^{\nu+\epsilon} |\omega - \nu|^{-2/3} \, \mathrm{d}\omega + \epsilon\nu^{-2/3} \lesssim \epsilon^{1/3} + \epsilon\nu^{-2/3}.$$

Next, we consider the remaining integration regimes where we use (4.22b) and (4.41) to estimate

$$\begin{aligned} &\Big| \int_\epsilon^{\nu-\epsilon} \frac{\nu}{(\omega - \nu)\omega} \Big[\rho_{r,t}(\mathfrak{e}^+_{r,t} + \omega) - \rho_{\mu,t}(\mathfrak{e}^+_{\mu,t} + \omega) \Big] \mathrm{d}\omega \Big| \\ &\quad \lesssim \int_\epsilon^{\nu/2} \frac{\omega^{1/3}(t_* - t)^{1/3} + \omega^{2/3}}{\omega} \mathrm{d}\omega + \int_{\nu/2}^{\nu-\epsilon} \Big(\frac{\nu^{1/3}(t_* - t)^{1/3}}{\omega - \nu} + \frac{\nu^{2/3}}{\omega - \nu} \Big) \mathrm{d}\omega \\ &\quad \lesssim \nu^{1/3} \Big((t_* - t)^{1/3} + \nu^{1/3} \Big) |\log \epsilon| \end{aligned}$$

and similarly

$$\Big| \int_{\nu+\epsilon}^{\infty} \frac{\nu}{(\omega - \nu)\omega} \Big[\rho_{r,t}(\mathfrak{e}^+_{r,t} + \omega) - \rho_{\mu,t}(\mathfrak{e}^+_{\mu,t} + \omega) \Big] \mathrm{d}\omega \Big| \lesssim \nu^{1/3} \Big((t_* - t)^{1/3} + \nu^{1/3} \Big) |\log \epsilon|.$$

We now consider the difference of the first terms in (4.40) for $r = \lambda, \mu$ and for $\nu < 0$ where the bound is simpler because the integration regime close to ν does not have to be singled out. Using (4.22b) we find

$$\Big| \int_0^{\infty} \frac{\nu}{(\omega - \nu)\omega} \Big[\rho_{r,t}(\mathfrak{e}^+_{r,t} + \omega) - \rho_{\mu,t}(\mathfrak{e}^+_{\mu,t} + \omega) \Big] \mathrm{d}\omega \Big| \lesssim |\nu|^{2/3} + (t_* - t)^{1/3} |\nu|^{1/3}.$$

Finally, it remains to consider the difference of the second terms in (4.40). We first treat the regime where $\nu \ge -\frac{3}{4}\Delta_{r,t}$ and split the difference into the sum of two terms

$$\begin{aligned} &\Big| \int_0^\infty \Big(\frac{\nu \rho_{r,t}(\mathfrak{e}^-_{r,t} - \omega)}{(\omega + \Delta_{r,t} + \nu)(\omega + \Delta_{r,t})} - \frac{\nu \rho_{r,t}(\mathfrak{e}^-_{r,t} - \omega)}{(\omega + \Delta_{\mu,t} + \nu)(\omega + \Delta_{\mu,t})} \Big) \mathrm{d}\omega \Big| \\ &\le |\nu| |\Delta_{r,t} - \Delta_{\mu,t}| \int_0^\infty \frac{\rho_{r,t}(\mathfrak{e}^-_{r,t} - \omega) \big[2\Delta_{r,t} + 2\omega + |\nu| \big]}{(\omega + \Delta_{r,t} + \nu)^2 (\omega + \Delta_{r,t})^2} \mathrm{d}\omega \\ &\lesssim \frac{|\Delta_{r,t} - \Delta_{\mu,t}|}{\Delta_{r,t}^{2/3}} - \frac{|\Delta_{r,t} - \Delta_{\mu,t}|}{(\Delta_{r,t} + |\nu|)^{2/3}} \lesssim (t_* - t)^{1/3} |\nu|^{1/3} \end{aligned}$$

and

$$\begin{aligned} &\Big| \int_0^\infty \Big(\frac{\nu \rho_{r,t}(\mathfrak{e}^-_{r,t} - \omega)}{(\omega + \Delta_{\mu,t} + \nu)(\omega + \Delta_{\mu,t})} - \frac{\nu \rho_{\mu,t}(\mathfrak{e}^-_{\mu,t} - \omega)}{(\omega + \Delta_{\mu,t} + \nu)(\omega + \Delta_{\mu,t})} \Big) \mathrm{d}\omega \Big| \\ &\qquad \lesssim |\nu|^{2/3} + (t_* - t)^{1/3} |\nu|^{1/3}. \end{aligned}$$

Here we used $\rho_{r,t}(\mathfrak{e}^-_{r,t} - \omega) \lesssim \omega^{1/3}$ as well as (4.22a) for the first and (4.22a),(4.22b) for the second computation. By collecting the various error terms and choosing $\epsilon = \nu^2$ we conclude (4.39a).

We define $\kappa := -\nu - \Delta_{r,t}$. Then we are left with the regime $\nu < -\frac{3}{4}\Delta_{r,t}$ or equivalently $\kappa > -\frac{1}{4}\Delta_{r,t}$ and use

$$m_{r,t}(\mathfrak{e}^+_{r,t} + \nu) - m_{r,t}(\mathfrak{e}^+_{r,t}) = (m_{r,t}(\mathfrak{e}^-_{r,t} - \kappa) - m_{r,t}(\mathfrak{e}^-_{r,t})) + (m_{r,t}(\mathfrak{e}^-_{r,t}) - m_{r,t}(\mathfrak{e}^+_{r,t}))\,,$$

as well as

$$\begin{aligned} m_{\mu,t}(\mathfrak{e}^+_{\mu,t} + \nu) - m_{\mu,t}(\mathfrak{e}^+_{\mu,t}) =& (m_{\mu,t}(\mathfrak{e}^-_{\mu,t} - \kappa + \Delta_{\mu,t} - \Delta_{r,t}) - m_{\mu,t}(\mathfrak{e}^-_{\mu,t} - \kappa)) \\ &+ (m_{\mu,t}(\mathfrak{e}^-_{\mu,t} - \kappa) - m_{\mu,t}(\mathfrak{e}^-_{\mu,t})) + (m_{\mu,t}(\mathfrak{e}^-_{\mu,t}) - m_{\mu,t}(\mathfrak{e}^+_{\mu,t})) \end{aligned} \tag{4.42}$$

in the left hand side of (4.39a). Thus we have to estimate the three expressions,

$$\Big|\Re\Big[\Big(m_{r,t}(\mathfrak{e}^-_{r,t} - \kappa) - m_{r,t}(\mathfrak{e}^-_{r,t})\Big) - \Big(m_{\mu,t}(\mathfrak{e}^-_{\mu,t} - \kappa) - m_{\mu,t}(\mathfrak{e}^-_{\mu,t})\Big)\Big]\Big|, \tag{4.43a}$$

$$\Big|\Re\Big[\Big(m_{r,t}(\mathfrak{e}^-_{r,t}) - m_{r,t}(\mathfrak{e}^+_{r,t})\Big) - \Big(m_{\mu,t}(\mathfrak{e}^-_{\mu,t}) - m_{\mu,t}(\mathfrak{e}^+_{\mu,t})\Big)\Big]\Big|, \tag{4.43b}$$

$$\Big|\Re\Big[m_{\mu,t}(\mathfrak{e}^-_{\mu,t} - \kappa + \Delta_{\mu,t} - \Delta_{r,t}) - m_{\mu,t}(\mathfrak{e}^-_{\mu,t} - \kappa)\Big]\Big|. \tag{4.43c}$$

In order to bound the first term we use that estimating (4.43a) for $\kappa \geq -\frac{3}{4}\Delta_{r,t}$ is equivalent to estimating the left hand side of (4.39a) for $\nu \geq -\frac{3}{4}\Delta_{r,t}$, i.e. the regime we already considered above. This equivalence follows by using the reflection $A \to -A$ of the expectation (cf. (4.8)) that turns every left edge $\mathfrak{e}^+_{z,t}$ into a right edge $\mathfrak{e}^-_{z,t}$. In particular, by the analysis that we already performed (4.43a) is bounded by $|\kappa|^{1/3}[|\kappa|^{1/3} + (t_* - t)^{1/3}]|\log|\kappa||$. Since $|\kappa| \leq |\nu|$ this is the desired bound.

For the second term (4.43b) we see from (4.40) that we have to estimate the difference between the expressions

$$\int_0^\infty \frac{\Delta_{r,t}\rho_{r,t}(\mathfrak{e}^+_{r,t} + \omega)}{\omega(\omega + \Delta_{r,t})}\,\mathrm{d}\omega + \int_0^\infty \frac{\Delta_{r,t}\rho_{r,t}(\mathfrak{e}^-_{r,t} - \omega)}{\omega(\omega + \Delta_{r,t})}\,\mathrm{d}\omega, \tag{4.44}$$

for $r = \alpha, \lambda, \mu$. The summands in (4.44) are treated analogously, so we focus on the first summand. We split the integrand of the difference between the first summands and estimate

$$\frac{(\Delta_{r,t} - \Delta_{\mu,t})\rho_{r,t}(\mathfrak{e}^+_{r,t} + \omega)}{(\omega + \Delta_{r,t})(\omega + \Delta_{\mu,t})} + \frac{\Delta_{\mu,t}(\rho_{r,t}(\mathfrak{e}^+_{r,t} + \omega) - \rho_{\mu,t}(\mathfrak{e}^+_{\mu,t} + \omega))}{\omega(\omega + \Delta_{\mu,t})} \lesssim \frac{\Delta(\omega^{1/3} + (t_* - t)^{1/3})}{\omega^{2/3}(\omega + \Delta)}$$

where $\Delta := \Delta_{r,t} \sim \Delta_{\mu,t}$ and we used (4.22a), (4.22b) and the first inequality of (4.41). Thus

$$\left|\int_0^\infty \frac{\Delta_{r,t}\rho_{r,t}(\mathfrak{e}^+_{r,t} + \omega)}{\omega(\omega + \Delta_{r,t})}\,\mathrm{d}\omega - \int_0^\infty \frac{\Delta_{\mu,t}\rho_{\mu,t}(\mathfrak{e}^+_{\mu,t} + \omega)}{\omega(\omega + \Delta_{\mu,t})}\,\mathrm{d}\omega\right| \lesssim \Delta^{2/3} + \Delta^{1/3}(t_* - t)^{1/3}.$$

Since $|\nu| \gtrsim \Delta$ this finishes the estimate on (4.43b).

For (4.43c) we use the $1/3$-Hölder regularity of $m_{\mu,t}$ and (4.22a) to get an upper bound $\Delta^{1/3}(t_* - t)^{1/9} \lesssim (t_* - t)^{11/18}$. This finishes the proof of (4.39a).

We now turn to the case of a small local minimum in (4.39b) and compute for $r = \alpha, \lambda, \mu$ and $\nu \neq 0$ that

$$\Re\Big[m_{r,t}(\mathfrak{m}_{r,t} + \nu) - m_{r,t}(\mathfrak{m}_{r,t})\Big] = \int_{\mathbf{R}} \frac{\nu \rho_{r,t}(\mathfrak{m}_{r,t} + \omega)}{(\omega - \nu)\omega} \,\mathrm{d}\omega.$$

Without loss of generality, we consider the case $\nu > 0$ as $\nu < 0$ is completely analogous. As before, we first pick a threshold $\epsilon \leq \nu/2$ and single out the integration over $[-\epsilon, \epsilon]$ and $[\nu - \epsilon, \nu + \epsilon]$. From the $1/3$-Hölder continuity of $\rho_{r,t}$ we have, for $r = \lambda, \mu$,

$$\rho_{r,t}(\mathfrak{m}_{r,t} + \omega) = \rho_{r,t}(\mathfrak{m}_{r,t} + \nu) + \mathcal{O}(|\nu - \omega|^{1/3})$$

and therefore

$$\left|\int_{-\epsilon}^{\epsilon} \frac{\rho_{r,t}(\mathfrak{m}_{r,t} + \omega)}{\omega - \nu} \,\mathrm{d}\omega\right| \lesssim \frac{\epsilon}{\nu}, \quad \left|\int_{-\epsilon}^{\epsilon} \frac{\rho_{r,t}(\mathfrak{m}_{r,t} + \omega)}{\omega} \,\mathrm{d}\omega\right| \lesssim \int_{-\epsilon}^{\epsilon} |\omega|^{-2/3} \,\mathrm{d}\omega \lesssim \epsilon^{1/3}$$

and

$$\left|\int_{\nu-\epsilon}^{\nu+\epsilon} \frac{\rho_{r,t}(\mathfrak{m}_{r,t} + \omega)}{\omega - \nu} \,\mathrm{d}\omega\right| \lesssim \int_{\nu-\epsilon}^{\nu+\epsilon} |\omega - \nu|^{-2/3} \,\mathrm{d}\omega \lesssim \epsilon^{1/3}, \quad \left|\int_{\nu-\epsilon}^{\nu+\epsilon} \frac{\rho_{r,t}(\mathfrak{m}_{r,t} + \omega)}{\omega} \,\mathrm{d}\omega\right| \lesssim \frac{\epsilon}{\nu}.$$

We now consider the difference between $\rho_{r,t}$ and $\rho_{\mu,t}$ for which we have

$$|\rho_{r,t}(\mathfrak{m}_{r,t} + \omega) - \rho_{\mu,t}(\mathfrak{m}_{\mu,t} + \omega)| \lesssim (t - t_*)|\omega|^{1/3}(t - t_*)^{1/4} + (t - t_*)^{3/4} + |\omega|^{2/3}$$

from (4.22d), (4.22c) and the $1/3$-Hölder continuity of $\rho_{r,t}$. Thus we can estimate

$$\begin{aligned}
&\left|\left[\int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\nu-\epsilon} + \int_{\nu+\epsilon}^{\infty}\right] \frac{\nu\big(\rho_{\lambda,t}(\mathfrak{m}_{r,t} + \omega) - \rho_{r,t}(\mathfrak{m}_{r,t} + \omega)\big)}{(\omega - \nu)\omega} \,\mathrm{d}\omega\right| \\
&\quad \lesssim \left[\int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\nu-\epsilon} + \int_{\nu+\epsilon}^{\infty}\right] \frac{\nu\big(|\omega|^{1/3}(t - t_*)^{1/4} + (t - t_*)^{3/4} + |\omega|^{2/3}\big)}{|\omega - \nu|\omega} \,\mathrm{d}\omega \\
&\quad \lesssim |\log \epsilon| \Big[\nu^{1/3}(t - t_*)^{1/4} + (t - t_*)^{3/4} + \nu^{2/3}\Big].
\end{aligned}$$

We again choose $\epsilon = \nu^2$ and by collecting the various error estimates can conclude (4.39b). □

4.5 Index matching for two DBM

For two real symmetric matrix valued standard (GOE) Brownian motions $\mathfrak{B}_t^{(\lambda)}, \mathfrak{B}_t^{(\mu)} \in \mathbf{R}^{N \times N}$ we define the matrix flows

$$H_t^{(\lambda)} := H^{(\lambda)} + \mathfrak{B}_t^{(\lambda)}, \quad H_t^{(\mu)} := H^{(\mu)} + \mathfrak{B}_t^{(\mu)}. \tag{4.45}$$

In particular, by (4.45) it follows that

$$H_t^{(\lambda)} \stackrel{d}{=} H^{(\lambda)} + \sqrt{t}U^{(\lambda)}, \quad H_t^{(\mu)} \stackrel{d}{=} H^{(\mu)} + \sqrt{t}U^{(\mu)}, \tag{4.46}$$

for any fixed $0 \leq t \leq t_1$, where $U^{(\lambda)}$ and $U^{(\mu)}$ are GOE matrices. In (4.46) with $X \stackrel{d}{=} Y$ we denote that the two random variables X and Y are equal in distribution.

We will prove Proposition 4.3.1 by comparing the two Dyson Brownian motions for the eigenvalues of the matrices $H_t^{(\lambda)}$ and $H_t^{(\mu)}$ for $0 \leq t \leq t_1$, see (4.47)–(4.48) below. To do this, we will use the coupling idea of [39] and [42], where the DBMs for the eigenvalues of $H_t^{(\lambda)}$ and $H_t^{(\mu)}$ are coupled in such a way that the difference of the two DBMs obeys a discrete parabolic equation with good decay properties. In order to analyse this equation we consider a *short range approximation* for the DBM, first introduced in [91]. Coupling only the short range approximation of the DBMs leads to a parabolic equation whose heat kernel has a rapid off diagonal decay by *finite speed of propagation* estimates. In this way the kernels of both DBMs are locally determined and thus can be directly compared by optimal rigidity since locally the two densities, hence their quantiles, are close. Technically it is much easier to work with a one parameter interpolation between the two DBM's and consider its derivative with respect to the parameter, as introduced in [39]; the proof of the finite speed propagation for this dynamics does not require to establish level repulsion unlike in several previous works [88, 91, 130]. However, it requires to establish (almost) optimal rigidity for the interpolating dynamics as well. Note that optimal rigidity is known for $H_t^{(\lambda)}$ and $H_t^{(\mu)}$ from [83], see Lemma 4.6.1 later, but not for the interpolation. For a complete picture, we mention that in the works [88, 91, 130] on *bulk gap universality*, beyond heat kernel and Sobolev estimates, a version of De Giorgi-Nash-Moser parabolic regularity estimate, which used level repulsion in a more substantial way than finite speed of propagation, was also necessary. *Fixed energy universality* in the bulk can be proven via homogenisation without De Giorgi-Nash-Moser estimates, hence level repulsion can also be avoided [129]. In a certain sense, the situation at the edge/cusp is easier than the bulk regime since relatively simple heat kernel bounds are sufficient for local relaxation to equilibrium. In another sense, due to singularities in the density, the edge and especially the cusp regime is more difficult.

In Section 4.6 we will establish rigidity for the interpolating process by DBM methods. Armed with this rigidity, in Section 4.7 we prove Proposition 4.3.1 for the small gap and the exact cusp case, i.e. $t_1 \leq t_*$. Some estimates are slightly different for the small minimum case, i.e. $t_* \leq t_1 \leq 2t_*$, the modifications are given in Section 4.8. We recall that t_* is the time at which both $H_{t_*}^{(\lambda)}$ and $H_{t_*}^{(\mu)}$ have an exact cusp. Some technical details on the corresponding Sobolev inequality and heat kernel estimates as well as finite speed of propagation and short range approximation are deferred to the Appendix: these are similar to the corresponding estimates for the edge case, see [41] and [131], respectively.

In the rest of this section we prepare the proof of Proposition 4.3.1 by setting up the appropriate framework. While we are interested only in the eigenvalues near the physical cusp, the DBM is highly non-local, so we need to define the dynamics for all eigenvalues. In the setup of Proposition 4.3.1 we could easily assume that the cusps for the two matrix flows are formed at the same time and their slope parameters coincide – these could be achieved by a rescaling and a trivial time shift. However, the number of eigenvalues to the left of the cusp may macroscopically differ for the two ensembles which would mean that the labels of the ordered eigenvalues near the cusp would not be constant along the interpolation. To resolve this discrepancy, we will pad the system with N fictitious particles in addition to the original flow of N eigenvalues similarly as in [129], giving sufficient freedom to match the labels of the eigenvalues near the cusp. These artificial particles will be placed very far from the cusp regime and from each other so that their effect on the dynamics of the relevant particles is negligible.

With the notation of Section 4.4, we let $\rho_{\lambda,t}$, $\rho_{\mu,t}$ denote the (self-consistent) densities at time $0 \le t \le t_1$ of $H_t^{(\lambda)}$ and $H_t^{(\mu)}$, respectively. In particular, $\rho_{\lambda,0} = \rho_\lambda$ and $\rho_{\mu,0} = \rho_\mu$, where ρ_λ, ρ_μ are the self consistent densities of $H^{(\lambda)}$ and $H^{(\mu)}$ and $\rho_{\lambda,t}$, $\rho_{\mu,t}$ are their semicircular evolutions. For each $0 \le t \le t_*$ both densities $\rho_{\lambda,t}$, $\rho_{\mu,t}$ have a small gap, denoted by $[\mathfrak{e}_{\lambda,t}^-, \mathfrak{e}_{\lambda,t}^+]$ and $[\mathfrak{e}_{\mu,t}^-, \mathfrak{e}_{\mu,t}^+]$ and we let

$$\Delta_{\lambda,t} := \mathfrak{e}_{\lambda,t}^+ - \mathfrak{e}_{\lambda,t}^-, \qquad \Delta_{\mu,t} := \mathfrak{e}_{\mu,t}^+ - \mathfrak{e}_{\mu,t}^-$$

denote the length of these gaps. In case of $t_* \le t \le 2t_*$ the densities $\rho_{\lambda,t}$, $\rho_{\mu,t}$ have a small minimum denoted by $\mathfrak{m}_{\lambda,t}$ and $\mathfrak{m}_{\mu,t}$ respectively. Since we always assume $0 \le t \le t_1 \ll 1$, both $H_t^{(\lambda)}$ and $H_t^{(\mu)}$ will always have exactly one physical cusp near $\mathfrak{c}_\lambda$ and $\mathfrak{c}_\mu$, respectively, using that the Stieltjes transform of the density is a Hölder continuous function of t, see [14, Proposition 10.1].

Let i_λ and i_μ be the indices defined by

$$\int_{-\infty}^{\mathfrak{e}_{\lambda,0}^-} \rho_\lambda = \frac{i_\lambda - 1}{N}, \quad \int_{-\infty}^{\mathfrak{e}_{\mu,0}^-} \rho_\mu = \frac{i_\mu - 1}{N}.$$

By band rigidity (see Remark 2.6 in [15]) i_λ and i_μ are integers. Note that by the explicit expression of the density in (4.3a)-(4.3b) it follows that $cN \le i_\lambda, i_\mu \le (1-c)N$ with some small $c > 0$, because the density on both sides of a physical cusp is macroscopic.

We let $\lambda_i(t)$ and $\mu_i(t)$ denote the eigenvalues of $H_t^{(\lambda)}$ and $H_t^{(\mu)}$, respectively. Let $\{B_i\}_{i \in [-N,N] \setminus \{0\}}$ be a family of independent standard (scalar) Brownian motions. It is well known [74] that the eigenvalues of $H_t^{(\lambda)}$ satisfy the equation for *Dyson Brownian motion*, i.e. the following system of coupled SDE's

$$\mathrm{d}\lambda_i = \sqrt{\frac{2}{N}}\, \mathrm{d}B_{i-i_\lambda+1} + \frac{1}{N} \sum_{j \ne i} \frac{1}{\lambda_i - \lambda_j}\, \mathrm{d}t \tag{4.47}$$

with initial conditions $\lambda_i(0) = \lambda_i(H^{(\lambda)})$. Similarly, for the eigenvalues of $H_t^{(\mu)}$ we have

$$\mathrm{d}\mu_i = \sqrt{\frac{2}{N}}\, \mathrm{d}B_{i-i_\mu+1} + \frac{1}{N} \sum_{j \ne i} \frac{1}{\mu_i - \mu_j}\, \mathrm{d}t \tag{4.48}$$

with initial conditions $\mu_i(0) = \mu_i(H^{(\mu)})$. Note that we chose the Brownian motions for λ_i and $\mu_{i+i_\mu-i_\lambda}$ to be identical. This is the key ingredient for the coupling argument, since in this way the stochastic differentials will cancel when we take the difference of the two DBMs or we differentiate it with respect to an additional parameter.

For convenience of notation, we will shift the indices so that the same index labels the last quantile before the gap in ρ_λ and ρ_μ. This shift was already prepared by choosing the Brownian motions for μ_{i_μ} and λ_{i_λ} to be identical. We achieve this shift by adding N "ghost" particles very far away and relabelling, as in [129]. We thus embed λ_i and μ_i into the enlarged processes $\{x_i\}_{i \in [-N,N] \setminus \{0\}}$ and $\{y_i\}_{i \in [-N,N] \setminus \{0\}}$. Note that the index 0 is always omitted.

More precisely, the processes x_i are defined by the following SDE *(extended Dyson Brownian motion)*

$$\mathrm{d}x_i = \sqrt{\frac{2}{N}}\, \mathrm{d}B_i + \frac{1}{N} \sum_{j \ne i} \frac{1}{x_i - x_j}\, \mathrm{d}t, \qquad 1 \le |i| \le N, \tag{4.49}$$

with initial data

$$x_i(0) = \begin{cases} -N^{200} + iN & \text{if } -N \le i \le -i_\lambda \\ \lambda_{i+i_\lambda}(0) & \text{if } 1 - i_\lambda \le i \le -1 \\ \lambda_{i+i_\lambda-1}(0) & \text{if } 1 \le i \le N+1-i_\lambda \\ N^{200} + iN & \text{if } N+2-i_\lambda \le i \le N, \end{cases} \tag{4.50}$$

and the y_i are defined by

$$\mathrm{d}y_i = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \frac{1}{N}\sum_{j\neq i}\frac{1}{y_i - y_j}\,\mathrm{d}t, \qquad 1 \le |i| \le N, \tag{4.51}$$

with initial data

$$y_i(0) = \begin{cases} -N^{200} + iN & \text{if } -N \le i \le -i_\mu \\ \mu_{i+i_\mu}(0) & \text{if } 1 - i_\mu \le i \le -1 \\ \mu_{i+i_\mu-1}(0) & \text{if } 1 \le i \le N+1-i_\mu \\ N^{200} + iN & \text{if } N+2-i_\mu \le i \le N. \end{cases} \tag{4.52}$$

The summations in (4.49) and (4.51) extend to all j with $1 \le |j| \le N$ except $j = i$.

The following lemma shows that the additional particles at distance N^{200} have negligible effect on the dynamics of the re-indexed eigenvalues, thus we can study the processes x_i and y_i instead of the eigenvalues λ_i, μ_i. The proof of this lemma follows by Appendix C of [129].

Lemma 4.5.1. *With very high probability the following estimates hold:*

$$\sup_{0\le t\le 1}\sup_{1\le i\le N+1-i_\lambda}|x_i(t) - \lambda_{i+i_\lambda-1}(t)| \le N^{-100},$$

$$\sup_{0\le t\le 1}\sup_{1-i_\lambda\le i\le N+1-i_\lambda}|x_i(t) - \lambda_{i+i_\lambda}(t)| \le N^{-100},$$

$$\sup_{0\le t\le 1}\sup_{1\le i\le N+1-i_\mu}|y_i(t) - \mu_{i+i_\mu-1}(t)| \le N^{-100},$$

$$\sup_{0\le t\le 1}\sup_{1-i_\mu\le i\le N+1-i_\mu}|y_i(t) - \mu_{i+i_\mu}(t)| \le N^{-100},$$

$$\sup_{0\le t\le 1} x_{-i_\lambda}(t) \lesssim -N^{200}, \qquad \sup_{0\le t\le 1} x_{N+2-i_\lambda}(t) \gtrsim N^{200},$$

$$\sup_{0\le t\le 1} y_{-i_\mu}(t) \lesssim -N^{200}, \qquad \sup_{0\le t\le 1} y_{N+2-i_\mu}(t) \gtrsim N^{200}.$$

Remark 4.5.2. *For notational simplicity we assumed that $H^{(\lambda)}$ and $H^{(\mu)}$ have the same dimensions, but our proof works as long as the corresponding dimensions N_λ and N_μ are merely comparable, say $\frac{2}{3}N_\lambda \le N_\mu \le \frac{3}{2}N_\lambda$. The only modification is that the times in* (4.45) *need to be scaled differently in order to keep the strength of the stochastic differential terms in* (4.47)–(4.48) *identical. In particular, we rescale the time in the process* (4.47) *as $t' = (N_\mu/N_\lambda)t$, in such a way the N-scaling in front of the stochastic differential and in front of the potential term are exactly the same in both the processes* (4.47) *and* (4.48)*; namely we may replace N with N_μ in both* (4.47) *and* (4.48).

Furthermore, the number of additional "ghost" particles in the extended Dyson Brownian motion *(see* (4.49) *and* (4.51)*) will be different to ensure that we have the same total number of particles, i.e. the total number of x and y particles will be $2N := 2\max\{N_\mu, N_\lambda\}$, after the extension. Hence, assuming that $N_\mu \geq N_\lambda$, there will be $N = N_\mu$ particles added to the DBM of the eigenvalues of $H^{(\mu)}$ and $2N_\mu - N_\lambda$ particles added to the DBM of $H^{(\lambda)}$. In particular, under the assumption $N_\mu \geq N_\lambda$, we may replace* (4.50) *and* (4.52) *by*

$$x_i(0) = \begin{cases} -N_\mu^{200} + iN_\mu & \text{if } -N_\mu \leq i \leq -i_\lambda \\ \lambda_{i+i_\lambda}(0) & \text{if } 1 - i_\lambda \leq i \leq -1 \\ \lambda_{i+i_\lambda-1}(0) & \text{if } 1 \leq i \leq N_\lambda + 1 - i_\lambda \\ N_\mu^{200} + iN_\mu & \text{if } N_\lambda + 2 - i_\lambda \leq i \leq N_\mu, \end{cases}$$

and

$$y_i(0) = \begin{cases} -N_\mu^{200} + iN_\mu & \text{if } -N_\mu \leq i \leq -i_\mu \\ \mu_{i+i_\mu}(0) & \text{if } 1 - i_\mu \leq i \leq -1 \\ \mu_{i+i_\mu-1}(0) & \text{if } 1 \leq i \leq N_\mu + 1 - i_\mu \\ N_\mu^{200} + iN_\mu & \text{if } N_\mu + 2 - i_\mu \leq i \leq N_\mu. \end{cases}$$

Then, all the proofs of Section 4.5 and Section 4.6 are exactly the same of the case $N := N_\mu = N_\lambda$, since all the analysis of the latter sections is done in a small, order one neighborhood of the physical cusp. In particular, only the particles $x_i(t)$, $y_i(t)$ with $1 \leq |i| \leq \epsilon \min\{N_\mu, N_\lambda\}$, for some small fixed $\epsilon > 0$, will matter for our analysis. The far away particles in the case will be treated exactly as in (4.53)–(4.57) *replacing N by N_μ.*

We now construct the analogues of the self-consistent densities $\rho_{\lambda,t}$, $\rho_{\mu,t}$ for the $x(t)$ and $y(t)$ processes as well as for their α-interpolations. We start with $\rho_{x,t}$. Recall $\rho_{\lambda,t}$ from Section 4.4, and set

$$\rho_{x,t}(E) := \rho_{\lambda,t}(E) + \frac{1}{N}\sum_{i=-N}^{-i_\lambda} \psi(E - x_i(t)) + \frac{1}{N}\sum_{i=N+2-i_\lambda}^{N} \psi(E - x_i(t)), \qquad E \in \mathbf{R}, \tag{4.53}$$

where ψ is a non-negative symmetric approximate delta-function on scale N^{-1}, i.e. it is supported in an N^{-1} neighbourhood of zero, $\int \psi = 1$, $\|\psi\|_\infty \lesssim N$ and $\|\psi'\|_\infty \lesssim N^2$. Note that the total mass is $\int_{\mathbf{R}} \rho_{x,t} = 2$. For the Stieltjes transform $m_{x,t}$ of $\rho_{x,t}$, we have $\sup_{z\in\mathbf{C}^+}|m_{x,t}(z)| \leq C$ since the same bound holds for $\rho_{\lambda,t}$ by the shape analysis. Note that $\rho_{\lambda,t}$ is the semicircular flow with initial condition $\rho_{\lambda,t=0} = \rho_\lambda$ by definition, but $\rho_{x,t}$ is not exactly the semicircular evolution of $\rho_{x,0}$. We will not need this information, but in fact, the effect of the far away padding particles on the density near the cusp is very tiny.

Since $\rho_{x,t}$ coincides with $\rho_{\lambda,t}$ in a big finite interval, their edges and local minima near the cusp regime coincide, i.e we can identify

$$\mathfrak{e}_{x,t}^{\pm} = \mathfrak{e}_{\lambda,t}^{\pm}, \qquad \mathfrak{m}_{x,t} = \mathfrak{m}_{\lambda,t}.$$

The shifted quantiles and semiquantiles $\widehat{\gamma}_{x,i}(t)$, $\check{\gamma}_{x,i}(t)$ and $\widehat{\gamma}^*_{x,i}(t)$, $\check{\gamma}^*_{x,i}(t)$ of $\rho_{x,t}$ are defined by the obvious analogues of the formulas (4.27)–(4.29) except that r subscript is replaced with x and the indices run over the entire range $1 \leq |i| \leq N$. As before, $\gamma_{x,0}(t) = \mathfrak{e}_{x,t}^+$. The unshifted quantiles are defined by

$$\gamma_{x,i}(t) = \widehat{\gamma}_{x,i}(t) + \mathfrak{e}_{x,t}^+, \quad 0 \leq t \leq t_*, \qquad \gamma_{x,i}(t) = \check{\gamma}_{x,i}(t) + \mathfrak{m}_{x,t}, \quad t_* \leq t \leq 2t_*$$

and similarly for the semiquantiles.

So far we explained how to construct $\rho_{x,t}$ and its quantiles from $\rho_{\lambda,t}$, exactly in the same way we obtain $\rho_{y,t}$ from $\rho_{\mu,t}$ with straightforward notations.

Now for any $\alpha \in [0,1]$ we construct the α-interpolation of $\rho_{x,t}$ and $\rho_{y,t}$ that we will denote by $\overline{\rho}_t$. The bar will indicate quantities related to α-interpolation that implicitly depend on α; a dependence that we often omit from the notation. The interpolating measure will be constructed via its quantiles, i.e. we define

$$\overline{\gamma}_i(t) := \alpha\widehat{\gamma}_{x,i}(t) + (1-\alpha)\widehat{\gamma}_{y,i}(t), \qquad \overline{\gamma}_i^*(t) := \alpha\widehat{\gamma}^*_{x,i}(t) + (1-\alpha)\widehat{\gamma}^*_{y,i}(t), \tag{4.54}$$

for $1 \le |i| \le N$ and $0 \le t \le t_*$, and similarly for $t_* \le t \le 2t_*$ involving $\check{\gamma}$'s. We also set the interpolating edges

$$\overline{\mathfrak{e}}_t^{\pm} = \alpha\mathfrak{e}^{\pm}_{x,t} + (1-\alpha)\mathfrak{e}^{\pm}_{y,t}. \tag{4.55}$$

Recall the parameter δ_* describing the size of a neighbourhood around the physical cusp where the shape analysis for ρ_λ and ρ_μ in Section 4.2 holds. Choose $i(\delta_*) \sim N$ such that $|\overline{\gamma}_{x,-i(\delta_*)}(t)| \le \delta_*$ as well as $|\overline{\gamma}_{x,i(\delta_*)}(t)| \le \delta_*$ hold for all $0 \le t \le 2t_*$. Then define, for any $E \in \mathbf{R}$, the function

$$\overline{\rho}_t(E) := \rho_{\alpha,t}(E)\cdot\mathbf{1}\big(\overline{\gamma}_{-i(\delta_*)}(t)+\overline{\mathfrak{e}}_t^+ \le E \le \overline{\gamma}_{i(\delta_*)}(t)+\overline{\mathfrak{e}}_t^+\big)+\frac{1}{N}\sum_{i(\delta_*)<|i|\le N}\psi(E-\overline{\mathfrak{e}}_t^+-\overline{\gamma}_i^*(t)), \tag{4.56}$$

where $\rho_{\alpha,t}$ is the α-interpolation, constructed in Definition 4.4.1, between $\rho_{\lambda,t}(E) = \rho_{x,t}(E)$ and $\rho_{\mu,t}(E) = \rho_{y,t}(E)$ for $|E| \le \delta_*$. By this construction (using also the symmetry of ψ) we know that all shifted semiquantiles of $\overline{\rho}_t$ are exactly $\overline{\gamma}_i^*(t)$. The same holds for all shifted quantiles $\overline{\gamma}_i(t)$ at least in the interval $[-\delta_*, \delta_*]$ since here $\overline{\rho}_t \equiv \rho_{\alpha,t}$ and the latter was constructed exactly by the requirement of linearity of the quantiles (4.54), see (4.30c).

We also record $\int \overline{\rho}_t = 2$ and that for the Stieltjes transform $\overline{m}_t(z)$ of $\overline{\rho}_t$ we have

$$\max_{|\Re z-\overline{\mathfrak{e}}_t^+|\le\frac{1}{2}\delta_*} |\overline{m}_t(z)| \le C \tag{4.57}$$

for all $0 \le t \le 2t_*$. The first bound follows easily from the same boundedness of the Stieltjes transform of $\rho_{\alpha,t}$. Moreover, $\overline{m}_t(z)$ is $\frac{1}{3}$-Hölder continuous in the regime $|\Re z - \overline{\mathfrak{e}}_t^+| \le \frac{1}{2}\delta_*$ since in this regime $\overline{\rho}_t = \rho_{\alpha,t}$ and $\rho_{\alpha,t}$ is $\frac{1}{3}$-Hölder continuous by Lemma 4.4.3.

4.6 Rigidity for the short range approximation

In this section we consider Dyson Brownian Motion (DBM), i.e. a system of $2N$ coupled stochastic differential equations for $z(t) = \{z_i(t)\}_{[-N,N]\setminus\{0\}}$ of the form

$$\mathrm{d}z_i = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \frac{1}{N}\sum_j \frac{1}{z_i - z_j}\,\mathrm{d}t, \qquad 1 \le |i| \le N, \tag{4.58}$$

with some initial condition $z_i(t=0) = z_i(0)$, where

$$B(s) = (B_{-N}(s), \dots, B_{-1}(s), B_1(s), \dots, B_N(s))$$

is the vector of $2N$ independent standard Brownian motions. We use the indexing convention that all indices i, j, etc., run from $-N$ to N but zero index is excluded.

We will assume that $z_i(0)$ is an α-linear interpolation of $x_i(0), y_i(0)$ for some $\alpha \in [0,1]$:

$$z_i(0) = z_i(0,\alpha) := \alpha x_i(0) + (1-\alpha) y_i(0). \tag{4.59}$$

In the following of this section we will refer to the process defined by (4.58) using $z(t,\alpha)$ in order to underline the α dependence of the process. Clearly for $\alpha = 0, 1$ we recover the original $y(t)$ and $x(t)$ processes, $z(t, \alpha = 0) = y(t)$, $z(t, \alpha = 1) = x(t)$. For these processes we have the following optimal rigidity estimate that immediately follows from [83, Corollary 2.6] and Lemma 4.5.1:

Lemma 4.6.1. *Let $r_i(t) = x_i(t)$ or $r_i(t) = y_i(t)$ and $r = x, y$. Then, there exists a fixed small $\epsilon > 0$, depending only on the model parameters, such that for each $1 \le |i| \le \epsilon N$, we have*

$$\sup_{0 \le t \le 2t_*} |r_i(t) - \gamma_{r,i}(t)| \le N^{\xi} \eta_{\mathrm{f}}^{\rho_{r,t}}(\gamma_{r,i}(t)), \tag{4.60}$$

for any $\xi > 0$ with very high probability, where we recall that the behavior of $\eta_{\mathrm{f}}^{\rho_{r,t}}(\gamma_{r,i}(t))$, with $r = x, y$, is given by (4.34a).

Note that, by (4.22a), (4.22c) and (4.34), for all $1 \le |i| \le \epsilon N$ and for all $0 \le t \le t_*$ we have that

$$\eta_{\mathrm{f}}^{\rho_{r,t}}(\gamma_{r,i}(t)) \lesssim \frac{N^{\frac{\omega_1}{6}}}{|i|^{\frac{1}{4}} N^{\frac{3}{4}}}, \tag{4.61}$$

with $r = x, y$.

In particular, we know that $z(0,\alpha)$ lie close to the quantiles (4.54) of an α-*interpolating density* $\rho_z = \overline{\rho}_0$, see the definition in (4.56). This means that ρ_z has a small gap $[\mathfrak{e}_z^-, \mathfrak{e}_z^+]$ of size $\Delta_z \sim t_*^{3/2}$ (i.e. it will develop a physical cusp in a time of order t_*) and it is an α-interpolation between $\rho_{x,0}$ and $\rho_{y,0}$. Here interpolation refers to the process introduced in Section 4.5 that guarantees that the corresponding quantiles are convex linear combinations of the two initial densities with weights α and $1-\alpha$, i.e.

$$\gamma_{z,i} = \alpha \gamma_{x,i} + (1-\alpha)\gamma_{y,i}.$$

In this section we will prove rigidity results for $z(t,\alpha)$ and for its appropriate short range approximation.

Remark 4.6.2. *Before we go into the details, we point out that we will prove rigidity* dynamically, *i.e. using the DBM. The route chosen here is very different from the one in [131, Sec. 6], where the authors prove a local law for short times in order to get rigidity for the short range approximation of the interpolated process. While it would be possible to follow the latter strategy in the cusp regime as well, the technical difficulties are overwhelming, in fact already in the much simpler edge regime a large part of [131] was devoted to this task. The current proof of the optimal law at the cusp regime [83] heavily use an effective mean-field condition (called* flatness*) that corresponds to large time in the DBM. Relaxing this condition would require to adjust not only [83] but also the necessary deterministic analysis from [14] to the short time case. Similar complications would have arisen if we had followed the strategy of [1, 114] where rigidity is proven by analysing the characteristics of the McKean-Vlasov equation. The route chosen here is shorter and more interesting.*

Since the group velocity of the entire cusp regime is different for $\rho_{x,t}$ and $\rho_{y,t}$, the interpolated process will have an intermediate group velocity. Since we have to follow the process for time scales $t \sim N^{-\frac{1}{2}+\omega_1}$, much bigger than the relevant rigidity scale $N^{-\frac{3}{4}}$ we have to determine the group velocity quite precisely. Technically, we will encode this information by defining an appropriately shifted process $\widetilde{z}(t,\alpha) = z(t,\alpha) - \mathrm{Shift}(t,\alpha)$. It is essential that the shift function is independent of the indices i to preserve the local statistics of the process. In the next section we explain how to choose the shift.

4.6.1 Choice of the shifted process $\widetilde{z}$

The remainder of Section 4.6 is formulated for the small gap regime, i.e. for $0 \le t \le t_*$. We will comment on the modifications in the small minimum regime in Section 4.8. To match the location of the gap, the natural guess would be to study the shifted process $z_i(t,\alpha) - \mathfrak{e}^+_{z,t}$ where $[\mathfrak{e}^-_{z,t}, \mathfrak{e}^+_{z,t}]$ is the gap of the semicircular evolution $\rho_{z,t}$ of ρ_z near the physical cusp, and approximate $z_i(t,\alpha) - \mathfrak{e}^+_{z,t}$ by the shifted semiquantiles $\widehat{\gamma}^*_{z,i}(t)$ of $\rho_{z,t}$. However, the evolution of the semicircular flow $t \to \rho_{z,t}$ near the cusp is not sufficiently well understood. We circumvent this technical problem by considering the quantiles of another approximating density $\overline{\rho}_t$ defined by the requirement that its quantiles are exactly the α-linear combinations of the quantiles of $\rho_{x,t}$ and $\rho_{y,t}$ as described in Section 4.5. The necessary regularity properties of $\overline{\rho}_t$ follow directly from its construction. The precise description below assumes that $0 \le t \le 2t_*$, i.e. we are in the small gap situation. For $t_* \le t \le t_*$ an identical construction works but the reference point $\mathfrak{e}^+_{r,t}$ is replaced with the approximate minimum $\widetilde{\mathfrak{m}}_{r,t}$, for $r = x, y$. For simplicity we present all formulas for $0 \le t \le t^*$ and we will comment on the other case in Section 4.8.

More concretely, for any fixed $\alpha \in [0,1]$ recall the (semi)quantiles from (4.54). These are the (semi)quantiles of the interpolating density $\overline{\rho} = \overline{\rho}_t$ defined in (4.56) and let its Stieltjes transform be denoted by $\overline{m} = \overline{m}_t$. Bar will refer to quantities related to this interpolation; implicitly all quantities marked by bar depend on the interpolation parameter α, which dependence will be omitted from the notation. Notice that $\overline{\rho}_t$ has a gap $[\overline{\mathfrak{e}}^-_t, \overline{\mathfrak{e}}^+_t]$ near the cusp satisfying (4.55). Initially at $t = 0$ we have $\overline{\rho}_{t=0} = \rho_z$, in particular $\overline{\gamma}_i(t=0) = \widehat{\gamma}_{z,i}(t=0)$ and $\overline{\mathfrak{e}}^\pm_0 = \mathfrak{e}^\pm_z$. We will choose the shift in the definition of the $\widetilde{z}_i(t,\alpha)$ process so that we could use $\overline{\gamma}^*_i(t)$ to trail it.

The semicircular flow and the α-interpolation do not commute hence $\overline{\gamma}_i(t)$ are not the same as the quantiles $\widehat{\gamma}_{z,i}(t)$ of the semicircular evolution $\rho_{z,t}$ of the initial density ρ_z. We will, however, show that they are sufficiently close near the cusp and up to times relevant for us, modulo an irrelevant time dependent shift. Notice that the evolution of $\widehat{\gamma}_{z,i}(t)$ is hard to control since analysing $\frac{\mathrm{d}}{\mathrm{d}t}\widehat{\gamma}_{z,i}(t) = -\Re m_{z,t}(\gamma_{z,i}(t)) + \Re m_{z,t}(\mathfrak{e}^+_{z,t})$ would involve knowing the evolved density $\rho_{z,t}$ quite precisely in the critical cusp regime. While this necessary information is in principle accessible from the explicit expression for the semicircular flow and the precise shape analysis of ρ_z obtained from that of ρ_x and ρ_y, here we chose a different, technically lighter path by using $\overline{\gamma}_i(t)$. Note that unlike $\widehat{\gamma}_{z,i}(t)$, the derivative of $\overline{\gamma}_i(t)$ involves only the Stieltjes transform of the densities $\rho_{x,t}$ and $\rho_{y,t}$ for which shape analysis is available.

However, the global group velocities of $\overline{\gamma}(t)$ and $\widehat{\gamma}_z(t)$ are not the same near the cusp. We thus need to define $\widetilde{z}(t,\alpha)$ not as $z(t,\alpha) - \overline{\mathfrak{e}}^+_t$ but with a modified time dependent shift to make up for this velocity difference so that $\overline{\gamma}(t)$ indeed correctly follows $\widetilde{z}(t,\alpha)$. To determine this shift, we first define the function

$$h^*(t,\alpha) := \Re\Big[-\overline{m}_t(\overline{\mathfrak{e}}_t^+) + (1-\alpha)m_{y,t}(\mathfrak{e}_{y,t}^+) + \alpha m_{x,t}(\mathfrak{e}_{x,t}^+)\Big], \tag{4.62}$$

where recall that $\overline{m}_t$ is the Stieltjes transform of the measure $\overline{\rho}_t$. Note that $h^*(t) = O(1)$ following from the boundedness of the Stieltjes transforms $m_{x,t}$, $m_{y,t}$ and $\overline{m}_t(\overline{\mathfrak{e}}_t^+)$. The boundedness of $m_{x,t}$ and $m_{y,t}$ follows by (4.17) and $|\overline{m}_t(\overline{e}_t^+)| \le C$ by (4.57).

We note that

$$h^*(t,\alpha=0) = m_{y,t}(\mathfrak{e}_{y,t}^+) - \overline{m}_t(\overline{\mathfrak{e}}_t^+) = m_{y,t}(\mathfrak{e}_{y,t}^+) - \overline{m}_t(\mathfrak{e}_{y,t}^+)$$

since for $\alpha = 0$ we have $\mathfrak{e}_{y,t}^+ = \overline{\mathfrak{e}}_t^+$ by construction. At $\alpha = 0$ the measure $\overline{\rho}_t$ is given exactly by the density $\rho_{y,t}$ in an $\mathcal{O}(1)$ neighbourhood of the cusp. Away from the cusp, depending on the precise construction in the analogue of (4.56), the continuous $\rho_{y,t}$ is replaced by locally smoothed out Dirac measures at the quantiles. A similar statement holds at $\alpha = 1$, i.e. for the density $\rho_{x,t}$. It is easy to see that the difference of the corresponding Stieltjes transforms evaluated at the cusp regime is of order N^{-1}, i.e.

$$|h^*(t,\alpha=0)| + |h^*(t,\alpha=1)| = \mathcal{O}(N^{-1}). \tag{4.63}$$

Since later in (4.167) we will need to give some very crude estimate on the α-derivative of $h^*(t,\alpha)$, but it actually blows up since $\overline{m}_t'$ is singular at the edge, we introduce a tiny regularization of h^*, i.e. we define the function

$$h^{**}(t,\alpha) := \Re\Big[-\overline{m}_t(\overline{\mathfrak{e}}_t^+ + \mathrm{i}N^{-100}) + (1-\alpha)m_{y,t}(\mathfrak{e}_{y,t}^+) + \alpha m_{x,t}(\mathfrak{e}_{x,t}^+)\Big]. \tag{4.64}$$

Note that by the $\frac{1}{3}$-Hölder continuity of $\overline{m}_t$ in the cusp regime, i.e. for $z \in \mathbf{H}$ such that $|\Re z - \overline{\mathfrak{e}}_t^+| \le \frac{\delta_*}{2}$, it follows that

$$h^{**}(t,\alpha) = h^*(t,\alpha) + \mathcal{O}(N^{-30}). \tag{4.65}$$

Then, we define

$$h(t) = h(t,\alpha) := h^{**}(t,\alpha) - \alpha h^{**}(t,1) - (1-\alpha)h^{**}(t,0) = \mathcal{O}(1) \tag{4.66}$$

to ensure that

$$h(t,\alpha=0) = h(t,\alpha=1) = 0. \tag{4.67}$$

In particular, we have

$$h(t,\alpha) = \Re\Big[-\overline{m}_t(\overline{\mathfrak{e}}_t^+) + (1-\alpha)m_{y,t}(\mathfrak{e}_{y,t}^+) + \alpha m_{x,t}(\mathfrak{e}_{x,t}^+)\Big] + \mathcal{O}(N^{-1}). \tag{4.68}$$

Define its antiderivative

$$H(t,\alpha) := \int_0^t h(s,\alpha)\,\mathrm{d}s, \qquad H(0,\alpha) = 0, \qquad \max_{0\le t\le t_*} |H(t,\alpha)| \lesssim N^{-1/2+\omega_1}. \tag{4.69}$$

Now we are ready to define the correctly shifted process

$$\widetilde{z}_i(t) = \widetilde{z}_i(t,\alpha) := z_i(t) - \big[\alpha\mathfrak{e}_{x,t}^+ + (1-\alpha)\mathfrak{e}_{y,t}^+\big] - H(t,\alpha), \tag{4.70}$$

that will be trailed by $\overline{\gamma}_i(t)$. It satisfies the shifted DBM

$$\mathrm{d}\widetilde{z}_i = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \left[\frac{1}{N}\sum_{j\neq i}\frac{1}{\widetilde{z}_i-\widetilde{z}_j} + \Phi_\alpha(t)\right]\mathrm{d}t \tag{4.71}$$

with

$$\Phi(t) := \Phi_\alpha(t) = \alpha\Re m_{x,t}(\mathfrak{e}^+_{x,t}) + (1-\alpha)\Re m_{y,t}(\mathfrak{e}^+_{y,t}) - h(t,\alpha), \tag{4.72}$$

and with initial conditions $\widetilde{z}(0) := z(0) - \mathfrak{e}^+_z$ by (4.55) and $H(0,\alpha)=0$. The shift function satisfies

$$\Phi_\alpha(t) = \Re[\overline{m}_t(\overline{\mathfrak{e}}^+_t)] + \mathcal{O}(N^{-1}). \tag{4.73}$$

Notice that for $\alpha = 0, 1$ this definition gives back the naturally shifted $x(t)$ and $y(t)$ processes since we clearly have

$$\widetilde{z}(t,\alpha=1) = \widetilde{x}(t) := x(t) - \mathfrak{e}^+_{x,t}, \qquad \widetilde{z}(t,\alpha=0) = \widetilde{y}(t) := y(t) - \mathfrak{e}^+_{y,t}, \tag{4.74}$$

that are trailed by the shifted semiquantiles

$$\overline{\gamma}^*_i(t,\alpha=1) = \widehat{\gamma}^*_{x,i}(t) := \gamma^*_{x,i}(t) - \mathfrak{e}^+_{x,t}, \qquad \overline{\gamma}^*_i(t,\alpha=0) = \widehat{\gamma}^*_{y,i}(t) := \gamma^*_{y,i}(t) - \mathfrak{e}^+_{y,t}. \tag{4.75}$$

As we explained, the time dependent shift $H(t,\alpha)$ in (4.70) makes up for the difference between the true edge velocity of the semicircular flow (which we do not compute directly) and the naive guess which is $\frac{\mathrm{d}}{\mathrm{d}t}[\alpha\mathfrak{e}^+_{x,t} + (1-\alpha)\mathfrak{e}^+_{y,t}]$ hinted by the linear combination procedure. The precise expression (4.62) will come out of the proof. The key point is that this adjustment is global, i.e. it is only time dependent but independent of i since this expresses a group velocity of the entire cusp regime.

4.6.2 Plan of the proof.

In the following three subsections we prove an almost optimal rigidity not directly for $\widetilde{z}_i(t)$ but for its appropriate short range approximation $\widehat{z}_i(t)$. This will be sufficient for the proof of the universality. The proof of the rigidity will be divided into three phases, which we first explain informally, as follows.

Phase 1. (Subsection 4.6.3) The main result is a rigidity for $\widetilde{z}_i(t) - \overline{\gamma}_i(t)$ for $1 \leq |i| \lesssim \sqrt{N}$ on scale $N^{-\frac{3}{4}+C\omega_1}$ without i-dependence in the error term. First we prove a crude rigidity on scale $N^{-1/2+C\omega_1}$ for all indices i. Using this rigidity, we can define a short range approximation $\mathring{z}$ of the original dynamics $\widetilde{z}$ and show that $\widetilde{z}_i$ and $\mathring{z}_i$ are close by $N^{-\frac{3}{4}+C\omega_1}$ for $1 \leq |i| \lesssim \sqrt{N}$. Then we analyse the short range process $\mathring{z}$ that has a finite speed of propagation, so we can localize the dynamics. Finally, we can directly compare $\mathring{z}$ with a deterministic particle dynamics because the effect of the stochastic term $\sqrt{2/N}\,\mathrm{d}B_i$, i.e. $\sqrt{t_*/N} = N^{-3/4+\omega_1/2} \ll N^{-3/4+C\omega_1}$, remains below the rigidity scale of interest in this Phase 1.

However, to understand this deterministic particle dynamics we need to compare it with the corresponding continuum evolution; this boils down to estimating the difference of a Stieltjes transform and its Riemann sum approximation at the semiquantiles. Since the Stieltjes transform is given by a singular integral, this approximation relies on quite delicate cancellations which require some strong regularity properties of the density. We can easily guarantee this regularity by considering the density $\overline{\rho}_t$ of the linear interpolation between the quantiles of $\rho_{x,t}$ and $\rho_{y,t}$.

Phase 2. (Subsection 4.6.4) In this section we improve the rigidity from scale $N^{-\frac{3}{4}+C\omega_1}$ to scale $N^{-\frac{3}{4}+\frac{1}{6}\omega_1}$, for a smaller range of indices, but we can achieve this not for $\widetilde{z}$ directly, but for its short range approximation $\widehat{z}$. Unlike $\mathring{z}$ in Phase 1, this time we choose a very short scale approximation $\widehat{z}$ on scale $N^{4\omega_\ell}$ with $\omega_1 \ll \omega_\ell \ll 1$. As an input, we need the rigidity of $\widetilde{z}_i$ on scale $N^{-\frac{3}{4}+C\omega_1}$ for $1 \le |i| \lesssim \sqrt{N}$ obtained in Phase 1. We use heat kernel contraction for a direct comparison with the $y_i(t)$ dynamics for which we know optimal rigidity by [83], with the precise matching of the indices (*band rigidity*). In particular, when the gap is large, this guarantees that band rigidity is transferred to the $\widehat{z}$ process from the $\widehat{y}$ process.

Phase 3. (Subsection 4.6.5) Finally, we establish the optimal i-dependence in the rigidity estimate for $\widehat{z}_i$ from Phase 2, i.e. we get a precision $N^{-\frac{3}{4}+\frac{1}{6}\omega_1}|i|^{-1/4}$. The main method we use in Phase 3 is maximum principle. We compare $\widehat{z}_i$ with $\widehat{y}_{i-K}$, a slightly shifted element of the $\widehat{y}$ process, where $K = N^{\xi}$ with some tiny ξ. This method allows us to prove the optimal i-dependent rigidity (with a factor $N^{\frac{1}{6}\omega_1}$) but only for indices $|i| \gg K$ because otherwise $\widehat{z}_i$ and $\widehat{y}_{i-K}$ may be on different sides of the gap for small i. For very small indices, therefore, we need to rely on band rigidity for $\widehat{z}$ from Phase 2.

The optimal i-dependence allows us to replace the random particles $\widehat{z}$ by appropriate quantiles with a precision so that

$$|\widehat{z}_i - \widehat{z}_j| \lesssim N^{\frac{\omega_1}{6}}|\overline{\gamma}_i - \overline{\gamma}_j| \sim N^{-\frac{3}{4}+\frac{\omega_1}{6}}||i|^{\frac{3}{4}} - |j|^{\frac{3}{4}}|.$$

Such upper bound on $|\widehat{z}_i - \widehat{z}_j|$, hence a lower bound on the interaction kernel $\mathcal{B}_{ij} = |\widehat{z}_i - \widehat{z}_j|^{-2}$ of the differentiated DBM (see (4.163) later) with the correct dependence on the indices i, j, is essential since this gives the heat kernel contraction which eventually drives the precision below the rigidity scale in order to prove universality. On a time scale $t_* = N^{-\frac{1}{2}+\omega_1}$ the $\ell^p \to \ell^\infty$ contraction of the heat kernel gains a factor $N^{-\frac{4}{15}\omega_1}$ with the convenient choice of $p = 5$. Notice that $\frac{4}{15} > \frac{1}{6}$, so the contraction wins over the imprecision in the rigidity $N^{\frac{1}{6}\omega_1}$ from Phase 3, but not over $N^{C\omega_1}$ from Phase 1, showing that both Phase 2 and Phase 3 are indeed necessary.

4.6.3 Phase 1: Rigidity for $\widetilde{z}$ on scale $N^{-3/4+C\omega_1}$.

The main result of this section is the following proposition:

Proposition 4.6.3. *Fix* $\alpha \in [0,1]$. *Let* $\widetilde{z}(t,\alpha)$ *solve* (4.71) *with initial condition* $\widetilde{z}_i(0,\alpha)$ *satisfying the crude rigidity bound for all indices*

$$\max_{1 \le |i| \le N} |\widetilde{z}_i(0,\alpha) - \overline{\gamma}_i^*(0)| \lesssim N^{-1/2+2\omega_1}. \tag{4.76}$$

We also assume that

$$\|m_{x,0}\|_\infty + \|m_{y,0}\|_\infty + |\overline{m}_t(\mathfrak{e}_t^{\pm})| \le C. \tag{4.77}$$

Then we have a weak but uniform rigidity

$$\sup_{0\le t\le t_*}\max_{1\le|i|\le N}|\widetilde{z}_i(t,\alpha)-\overline{\gamma}_i^*(t)|\lesssim N^{-1/2+2\omega_1}, \tag{4.78}$$

with very high probability. Moreover, for small $|i|$, *i.e.* $1\le|i|\le i_*$, *with* $i_*:=N^{1/2+C_*\omega_1}$ *for some large* $C_*>100$, *we have a stronger rigidity:*

$$\sup_{0\le t\le t_*}\max_{1\le|i|\le i_*}|\widetilde{z}_i(t,\alpha)-\overline{\gamma}_i^*(t)|\lesssim\max_{1\le|i|\le 2i_*}|\widetilde{z}_i(0,\alpha)-\overline{\gamma}_i^*(0)|+\frac{N^{C\omega_1}}{N^{3/4}} \tag{4.79}$$

with very high probability.

In our application, (4.76) is satisfied and the right hand side of (4.79) is simply $N^{-\frac{3}{4}+C\omega_1}$ since

$$\widetilde{z}_i(0,\alpha)-\overline{\gamma}_i^*(0)=\alpha(x_i(0)-\gamma_{x,i}(0))+(1-\alpha)(y_i(0)-\gamma_{y,i}(0))=O\left(\frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}}|i|^{\frac{1}{4}}}\right), \tag{4.80}$$

for any $\xi>0$ with very high probability, by optimal rigidity for $x_i(0)$ and $y_i(0)$ from [83]. Similarly, the assumption (4.77) is trivially satisfied by (4.57). However, we stated Proposition 4.6.3 under the slightly weaker conditions (4.76), (4.77) to highlight what is really needed for its proof.

Before starting the proof, we recall the formula

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{\gamma}_{i,r}^*(t)=-\Re m_{r,t}(\gamma_{r,i}^*(t))+\Re m_{r,t}(\mathfrak{e}_{r,t}^+),\qquad r=x,y. \tag{4.81}$$

on the derivative of the (shifted) semiquantiles of a density which evolves by the semicircular flow and follows directly from (4.31c) and (4.31e).

Proof of Proposition 4.6.3. We start with the proof of the crude rigidity (4.78), then we introduce a short range approximation and finally, with its help, we prove the refined rigidity (4.79). The main technical input of the last step is a refined estimate on the forcing term. These four steps will be presented in the next four subsections.

4.6.3.1 Proof of the crude rigidity:

For the proof of (4.78), using (4.81) twice in (4.54), we notice that

$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{\gamma}_i^*(t)=\alpha[-\Re m_{x,t}(\gamma_{x,i}^*(t))+\Re m_{x,t}(\mathfrak{e}_{x,t}^+)]+(1-\alpha)[-\Re m_{y,t}(\gamma_{y,i}^*(t))+\Re m_{y,t}(\mathfrak{e}_{y,t}^+)]=O(1)$$

since $m_{x,t}$ and $m_{y,t}$ are bounded recalling that the semicircular flow preserves (or reduces) the ℓ^∞ norm of the Stieltjes transform by (4.17), so $\|m_{x,t}\|_\infty\le\|m_{x,0}\|_\infty\le C$, similarly for $m_{y,t}$. This gives

$$|\overline{\gamma}_i^*(t)-\overline{\gamma}_i^*(0)|\lesssim N^{-1/2+\omega_1}. \tag{4.82}$$

Thus in order to prove (4.78) it is sufficient to prove

$$\|\widetilde{z}(t,\alpha)-\widetilde{z}(0,\alpha)\|_\infty\le N^{-1/2+2\omega_1}, \tag{4.83}$$

for any fixed $\alpha \in [0,1]$. To do that, we compare the dynamics of (4.71) with the dynamics of the y-semiquantiles, i.e. set

$$u_i := u_i(t,\alpha) = \widetilde{z}_i(t) - \widehat{\gamma}^*_{y,i}(t),$$

for all $0 \le t \le t_*$.

Compute

$$\mathrm{d}u_i = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + (\widetilde{\mathcal{B}}u)_i\,\mathrm{d}t + \widetilde{F}_i(t)\,\mathrm{d}t \tag{4.84}$$

with

$$(\widetilde{\mathcal{B}}f)_i := \frac{1}{N}\sum_{j\neq i}\frac{f_j - f_i}{(\widetilde{z}_i - \widetilde{z}_j)(\widehat{\gamma}^*_{y,i} - \widehat{\gamma}^*_{y,j})} \tag{4.85}$$

and

$$\widetilde{F}_i(t) := \frac{1}{N}\sum_{j\neq i}\frac{1}{\widehat{\gamma}^*_{y,i} - \widehat{\gamma}^*_{y,j}} + \Re m_{y,t}(\gamma^*_{y,i}(t)) + \alpha\big[\Re m_{x,t}(\mathfrak{e}^+_{x,t}) - \Re m_{y,t}(\mathfrak{e}^+_{y,t})\big] - h(t).$$

The operator $\widetilde{\mathcal{B}}$ is defined on $\mathbf{C}^{2N}$ and we label the vectors $f \in \mathbf{C}^{2N}$ as

$$f = (f_{-N}, f_{-N+1}, \dots, f_{-1}, f_1, \dots, f_N),$$

i.e. we omit the $i = 0$ index. Accordingly, in the summations the $j = 0$ term is always omitted since $\widetilde{z}_j$, $\widehat{z}_j$ and $\widehat{\gamma}^*_{y,j}$ are defined for $1 \le |j| \le N$. Furthermore in the summation of the interaction terms, the $j = i$ term is always omitted.

We now show that

$$\|\widetilde{F}(t)\|_\infty \lesssim \log N, \qquad 0 \le t \le t^*. \tag{4.86}$$

By the boundedness of $m_{x,t}, m_{y,t}$ and the $1/3$-Hölder continuity of $\overline{m}_t$ in the cusp regime, it remains to control

$$\frac{1}{N}\sum_{j\neq i}\frac{1}{\widehat{\gamma}^*_{y,i}(t) - \widehat{\gamma}^*_{y,j}(t)} \lesssim \sum_{1\le|j-i|\le N}\frac{1}{|i-j|} \lesssim \log N$$

since $|\widehat{\gamma}^*_{y,j} - \widehat{\gamma}^*_{y,i}| \ge c|i-j|/N$ as the density $\rho_{y,t}$ is bounded.

Let $\widetilde{\mathcal{U}}(s,t)$ be the fundamental solution of the heat evolution with kernel $\widetilde{\mathcal{B}}$ from (4.85), i.e, for any $0 \le s \le t$

$$\partial_t\widetilde{\mathcal{U}}(s,t) = \widetilde{\mathcal{B}}(t)\widetilde{\mathcal{U}}(s,t), \qquad \widetilde{\mathcal{U}}(s,s) = I. \tag{4.87}$$

Note that $\widetilde{\mathcal{U}}$ is a contraction on every ℓ^p space and the same is true for its adjoint $\widetilde{\mathcal{U}}^*(s,t)$. In particular, for any indices a, b and times s, t we have

$$\widetilde{\mathcal{U}}_{ab}(s,t) \le 1, \quad \widetilde{\mathcal{U}}^*_{ab}(s,t) \le 1. \tag{4.88}$$

By Duhamel principle, the solution to the SDE (4.84) is given by

$$u(t) = \widetilde{\mathcal{U}}(0,t)u(0) + \sqrt{\frac{2}{N}}\int_0^t \widetilde{\mathcal{U}}(s,t)\,\mathrm{d}B(s) + \int_0^t \widetilde{\mathcal{U}}(s,t)\widetilde{F}(s)\,\mathrm{d}s, \tag{4.89}$$

where $B(s) = (B_{-N}(s), \dots, B_{-1}(s), B_1(s) \dots, B_N(s))$ are the $2N$ independent Brownian motions from (4.58).

For the second term in (4.89) we fix an index i and consider the martingale

$$M_t := \sqrt{\frac{2}{N}} \int_0^t \sum_j \widetilde{\mathcal{U}}_{ij}(s,t)\,\mathrm{d}B_j(s)$$

with its quadratic variation process

$$[M]_t := \frac{2}{N}\int_0^t \sum_j \left(\widetilde{\mathcal{U}}_{ij}(s,t)\right)^2 \mathrm{d}s = \frac{2}{N}\int_0^t \|\widetilde{\mathcal{U}}^*(s,t)\delta_i\|_2^2\,\mathrm{d}s \le \frac{2t}{N}.$$

By the Burkholder maximal inequality for martingales, for any $p > 1$ we have that

$$\mathbf{E} \sup_{0\le t\le T} |M_t|^{2p} \le C_p\,\mathbf{E}[M]_T^p \le C_p \frac{T^p}{N^p}.$$

By Markov inequality we obtain that

$$\sup_{0\le t\le T} |M_t| \le N^\xi \sqrt{\frac{T}{N}} \tag{4.90}$$

with probability more than $1 - N^{-D}$, for any (large) $D > 0$ and (small) $\xi > 0$.

The last term in (4.89) is estimated, using (4.86), by

$$|\int_0^t \widetilde{\mathcal{U}}(s,t)\widetilde{F}(s)\,\mathrm{d}s| \le t \max_{s\le t} \|\widetilde{F}(s)\|_\infty \lesssim t\log N. \tag{4.91}$$

This, together with (4.90) and the contraction property of $\widetilde{\mathcal{B}}$ implies from (4.89) that

$$\|u(t) - u(0)\|_\infty \lesssim N^{-3/4+\omega_1} + t\log N \lesssim N^{-1/2+2\omega_1}$$

with very high probability. Recalling the definition of u and (4.82), we get (4.83) since

$$\|\widetilde{z}(t) - \widetilde{z}(0)\|_\infty \le \|u(t) - u(0)\|_\infty + \|\widehat{\gamma}_y^*(t) - \widehat{\gamma}_y^*(0)\|_\infty \lesssim N^{-1/2+2\omega_1}.$$

This completes the proof of the crude rigidity bound (4.78).

4.6.3.2 Crude short range approximation.

Now we turn to the proof of (4.79) by introducing a short range approximation of the dynamics (4.71). Fix an integer L. Let $\mathring{z}_i = \mathring{z}_i(t)$ solve the *L-localized short scale DBM*

$$\mathrm{d}\mathring{z}_i = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \frac{1}{N}\sum_{j:|j-i|\le L} \frac{1}{\mathring{z}_i - \mathring{z}_j}\,\mathrm{d}t + \left[\frac{1}{N}\sum_{j:|j-i|>L} \frac{1}{\overline{\gamma}_i^* - \overline{\gamma}_j^*} + \Phi(t)\right]\mathrm{d}t \tag{4.92}$$

for each $1 \le |i| \le N$ and with initial data $\mathring{z}_i(0) := \widetilde{z}_i(0)$, where we recall that Φ was defined in (4.72). Then, we have the following comparison:

Lemma 4.6.4. *Fix $\alpha \in [0,1]$. Assume that*

$$\max_{1\le|i|\le N} |\widetilde{z}_i(0,\alpha) - \overline{\gamma}_i^*(0)| \lesssim N^{-1/2+2\omega_1}. \tag{4.93}$$

Consider the short scale DBM (4.92) with a range $L = N^{1/2+C_1\omega_1}$ with a constant $10 \le C_1 \ll C_*$, in particular L is much smaller than i_*. Then we have a weak uniform comparison

$$\sup_{0\le t\le t_*} \max_{1\le|i|\le N} |\mathring{z}_i(t,\alpha) - \widetilde{z}_i(t,\alpha)| \lesssim N^{-1/2+2\omega_1}, \tag{4.94}$$

and a stronger comparison for small i:

$$\sup_{0\le t\le t_*} \max_{1\le|i|\le i_*} |\mathring{z}_i(t,\alpha) - \widetilde{z}_i(t,\alpha)| \lesssim N^{-3/4+C\omega_1}, \tag{4.95}$$

both with very high probability.

Proof. For any fixed $\alpha \in [0,1]$ and for all $0 \le t \le t_*$, set $w := w(t,\alpha) = \mathring{z}(t,\alpha) - \widetilde{z}(t,\alpha)$ and subtract (4.92) and (4.71) to get

$$\partial_t w = \mathring{\mathcal{B}}_1 w + \mathring{F},$$

where

$$(\mathring{\mathcal{B}}_1 f)_i := \frac{1}{N}\sum_{j:|j-i|\le L} \frac{f_j - f_i}{(\mathring{z}_i - \mathring{z}_j)(\widetilde{z}_i - \widetilde{z}_j)}, \qquad \mathring{F}_i := \frac{1}{N}\sum_{j:|j-i|>L}\Big[\frac{1}{\overline{\gamma}_i^* - \overline{\gamma}_j^*} - \frac{1}{\widetilde{z}_i - \widetilde{z}_j}\Big].$$

We estimate

$$|\mathring{F}_i| \le \frac{1}{N}\sum_{j:|j-i|>L} \frac{|\widetilde{z}_i - \overline{\gamma}_i^*| + |\widetilde{z}_j - \overline{\gamma}_j^*|}{(\overline{\gamma}_i^* - \overline{\gamma}_j^*)(\widetilde{z}_i - \widetilde{z}_j)} \lesssim \frac{N^{-1/2+2\omega_1}}{N}\sum_{j:|j-i|>L}\frac{1}{(\overline{\gamma}_i^* - \overline{\gamma}_j^*)(\widetilde{z}_i - \widetilde{z}_j)},$$

where we used the crude rigidity (4.78) (applicable by (4.93)), and we chose C_1 in $L = N^{1/2+C_1\omega_1}$ large enough so that $|\overline{\gamma}_i^* - \overline{\gamma}_j^*|$ for any $|i-j| \ge L$ be much bigger than the rigidity scale $N^{-1/2+2\omega_1}$ in (4.78). This is guaranteed since

$$|\overline{\gamma}_i^* - \overline{\gamma}_j^*| = \alpha|\widehat{\gamma}_{x,i}^* - \widehat{\gamma}_{x,j}^*| + (1-\alpha)|\widehat{\gamma}_{y,i}^* - \widehat{\gamma}_{y,j}^*| \gtrsim \frac{|i-j|}{N} \gtrsim N^{-1/2+C_1\omega}$$

with very high probability. By this choice of L we have $|\widetilde{z}_i - \widetilde{z}_j| \sim |\overline{\gamma}_i^* - \overline{\gamma}_j^*|$ and therefore

$$|\mathring{F}_i| \lesssim \frac{N^{-\frac{1}{2}+2\omega_1}}{N}\sum_{j:|j-i|>L}\frac{1}{(\overline{\gamma}_i^* - \overline{\gamma}_j^*)^2} \lesssim N^{1/2+2\omega_1}\sum_{j:|j-i|>L}\frac{1}{|i-j|^2} \lesssim N^{-(\frac{1}{2}C_1-2)\omega_1} \le 1, \tag{4.96}$$

for all $|i| \le N$. Since $\mathcal{B}_1$ is positivity preserving, its evolution is a contraction, so by Duhamel formula, similarly to (4.89), we get

$$\|\mathring{z}(t) - \widetilde{z}(t)\|_\infty = \|w(t)\|_\infty \le \|w(0)\|_\infty + t\max_{s\le t}\|\mathring{F}(s)\|_\infty \lesssim N^{-1/2+\omega_1}$$

with very high probability.

Next, we proceed with the proof of (4.95).

In fact, for $1 \le |i| \le 2i_*$, with i_* much bigger than L, we have a better bound:

$$|\mathring{F}_i| \lesssim \frac{N^{-\frac{1}{2}+2\omega_1}}{N} \sum_{j:|j-i|>L} \frac{1}{(\overline{\gamma}_i^* - \overline{\gamma}_j^*)^2} \lesssim \sum_{j:|j-i|>L} \frac{N^{2\omega_1}}{||i|^{3/4} - |j|^{3/4}|^2} \lesssim N^{-\frac{1}{4}-(\frac{1}{2}C_1-2)\omega_1} \le N^{-\frac{1}{4}}, \tag{4.97}$$

for $|i| \le 2i_*$, which we can use to get the better bound (4.95). To do so, we define a continuous interpolation $v(t,\beta)$ between $\widetilde{z}$ and $\mathring{z}$. More precisely, for any fixed $\beta \in [0,1]$ we set $v(t,\beta) = \{v(t,\beta)_i\}_{i=-N}^N$ as the solution to the SDE

$$\begin{aligned} \mathrm{d}v_i =& \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \frac{1}{N}\sum_{j:|j-i|\le L}\frac{1}{v_i - v_j}\,\mathrm{d}t + \Phi_\alpha(t)\,\mathrm{d}t \\ &+ \frac{1-\beta}{N}\sum_{j:|j-i|>L}\frac{1}{\widetilde{z}_i - \widetilde{z}_j}\,\mathrm{d}t + \frac{\beta}{N}\sum_{j:|j-i|>L}\frac{1}{\overline{\gamma}_i^* - \overline{\gamma}_j^*}\,\mathrm{d}t \end{aligned} \tag{4.98}$$

with initial condition $v(t=0,\beta) = (1-\beta)\widetilde{z}_i(0) + \beta\mathring{z}_i(0)$. Clearly $v(t,\beta=0) = \widetilde{z}(t)$ and $v(t,\beta=1) = \mathring{z}(t)$.

Differentiating in β, for $u := u(t,\beta) = \partial_\beta v(t,\beta)$ we obtain the SDE

$$\mathrm{d}u_i = (\mathcal{B}^v u)_i\,\mathrm{d}t + \mathring{F}_i\,\mathrm{d}t, \qquad \text{with} \quad (\mathcal{B}^v f)_i := \frac{1}{N}\sum_{j:|j-i|\le L}\frac{f_j - f_i}{(v_i - v_j)^2}, \tag{4.99}$$

with initial condition $u(t=0,\beta) = \mathring{z}(0) - \widetilde{z}(0) = 0$. By the contraction property of the heat evolution kernel $\mathcal{U}^v$ of $\mathcal{B}^v$, with a simple Duhamel formula, we have for any fixed β

$$\sup_{0\le t\le t_*} \|u(t,\beta)\|_\infty \le t_*\|\mathring{F}\|_\infty \le N^{-1/2+\frac{3}{2}\omega_1}, \tag{4.100}$$

with very high probability, where we used (4.96). After integration in β we get

$$\|v(t,\beta) - \overline{\gamma}^*(t)\|_\infty \le \|v(t,0) - \overline{\gamma}^*(t)\|_\infty + \|\int_0^\beta u(t,\beta')\,\mathrm{d}\beta'\|_\infty, \qquad 0 \le t \le t_*, \quad \beta \in [0,1]. \tag{4.101}$$

From (4.100) we have

$$\mathbf{E}\|\int_0^\beta u(t,\beta')\,\mathrm{d}\beta'\|_\infty^p \le \int_0^\beta \mathbf{E}\|u(t,\beta')\|^p\,\mathrm{d}\beta' \lesssim (N^{-1/2+\frac{3}{2}\omega_1})^p \tag{4.102}$$

for any exponent p. Hence, using a high moment Markov inequality, we have

$$\mathbf{P}\left(\|\int_0^\beta u(t,\beta')\,\mathrm{d}\beta'\|_\infty \ge N^{-1/2+\frac{3}{2}\omega_1+\xi}\right) \le N^{-D} \tag{4.103}$$

for any (large) $D > 0$ and (small) $\xi > 0$ by choosing p large enough. Since $v(t,0) = \widetilde{z}(t)$, for which we have rigidity in (4.78), by (4.101) and (4.103) we conclude that

$$\sup_{0\le t\le t_*} \|v(t,\beta) - \overline{\gamma}^*(t)\|_\infty \lesssim N^{-\frac{1}{2}+2\omega_1} \tag{4.104}$$

with very high probability for any $\beta \in [0,1]$.

In particular L is much larger than the rigidity scale of $v = v(t,\beta)$. This means that

$$||v_i - v_j| - |\overline{\gamma}_i^* - \overline{\gamma}_j^*|| \lesssim N^{-\frac{1}{2}+2\omega_1}$$

and $|\overline{\gamma}_i^* - \overline{\gamma}_j^*| \gtrsim \frac{|i-j|}{N} \geq N^{-\frac{1}{2}+C_1\omega_1} \gg N^{-\frac{1}{2}+2\omega_1}$ whenever $|i-j| \geq L$, so we have

$$|v_i - v_j| \sim |\overline{\gamma}_i^* - \overline{\gamma}_j^*|, \qquad |i-j| \geq L. \tag{4.105}$$

Since i_* is much bigger than L and L is much larger than the rigidity scale of $v_i(t,\beta)$ in the sense of (4.105), the heat evolution kernel $\mathcal{U}^v$ satisfies the following finite speed of propagation estimate (the proof is given in Appendix 4.B):

Lemma 4.6.5. *With the notations above we have*

$$\sup_{0\leq s\leq t\leq t_*} [\mathcal{U}_{pi}^v + \mathcal{U}_{ip}^v] \leq N^{-D}, \qquad 1 \leq |i| \leq i_*, \quad |p| \geq 2i_* \tag{4.106}$$

for any D if N is sufficiently large.

Using a Duhamel formula again, for any fixed β, we have

$$u_i(t) = \sum_p \mathcal{U}_{ip}^v u_p(0) + \int_0^t \sum_p \mathcal{U}_{ip}^v(s,t) \mathring{F}_p(s)\, \mathrm{d}s.$$

We can split the summation and estimate

$$|u_i(t)| \leq \Big[\sum_{|p|\leq 2i_*} + \sum_{|p|>2i_*} \Big] \mathcal{U}_{ip}^v |u_p(0)| + \int_0^t \Big[\sum_{|p|\leq 2i_*} + \sum_{|p|>2i_*} \Big] \mathcal{U}_{ip}^v(s,t) |\mathring{F}_p(s)|\, \mathrm{d}s.$$

For $|i| \leq i_*$, the terms with $|p| > 2i_*$ are negligible by (4.106) and the trivial bounds (4.96) and (4.100). For $1 \leq |p| \leq 2i_*$ we use the improved bound (4.97). This gives

$$|u_i(t,\beta)| \leq \max_{1\leq|j|\leq 2i_*} |u_j(0,\beta)| + N^{-3/4+\omega_1} = N^{-3/4+\omega_1}, \qquad |i| \leq i_*,$$

since $u(t=0,\beta) = 0$. Integrating from $\beta = 0$ to $\beta = 1$, and recalling that $v(\beta = 0) = \widetilde{z}$ and $v(\beta = 1) = \mathring{z}$, by high moment Markov inequality, we conclude

$$|\widetilde{z}_i(t) - \mathring{z}_i(t)| \lesssim N^{-\frac{3}{4}+\omega_1}, \qquad 1 \leq |i| \leq i_*,$$

with very high probability. This yields (4.95) and completes the proof of Lemma 4.6.4.

We remark that it would have been sufficient to require that $|\widetilde{z}_j(0) - \mathring{z}_j(0)| \leq N^{-\frac{3}{4}+\omega_1}$ for all $1 \leq |j| \leq 2i_*$ instead of setting $\mathring{z}(0) := \widetilde{z}(0)$ initially. Later in Section 4.6.4 we will use a similar finite speed of propagation mechanism to show that changing the initial condition for large indices has negligible effect. □

4.6.3.3 Refined rigidity for small $|i|$.

Finally, in the last but main step of the proof of (4.79) in Proposition 4.6.3 we compare $\mathring{z}_i$ with $\overline{\gamma}_i^*$ for small $|i|$ with a much higher precision than the crude bound $N^{-\frac{1}{2}+C\omega_1}$ which directly follows from (4.94) and (4.78). Notice that we use the semiquantiles for comparison since $\overline{\gamma}_i^* \in [\overline{\gamma}_{i-1}, \overline{\gamma}_i]$ and $\overline{\gamma}_i^*$ is typically close to the midpoint of this interval. In particular, $\overline{\rho}_t(\overline{\gamma}_i^*(t))$ is never zero, in fact we have $\overline{\rho}_t(\overline{\gamma}_i^*(t)) \geq cN^{-1/3}$, because by band rigidity quantiles may fall exactly at spectral edges, but semiquantiles cannot. This lower bound makes the semiquantiles much more convenient reference points than the quantiles.

Proposition 4.6.6. *Fix* $\alpha \in [0,1]$, *then with the notations above for the localized DBM* $\mathring{z}(t,\alpha)$ *on short scale* $L = N^{1/2+C_1\omega_1}$ *with* $10 \leq C_1 \leq \frac{1}{10}C_*$, *defined in* (4.92), *we have*

$$|(\mathring{z}_i(t,\alpha) - \overline{\gamma}_i^*(t)) - (\mathring{z}_i(0,\alpha) - \overline{\gamma}_i^*(0))| \leq N^{-3/4+C\omega_1}, \qquad 1 \leq |i| \leq i_* = N^{\frac{1}{2}+C_*\omega_1} \tag{4.107}$$

with very high probability.

Combining (4.107) with (4.95) and noticing that

$$\mathring{z}_i(0,\alpha) - \overline{\gamma}_i^*(0) = \widetilde{z}_i(0,\alpha) - \overline{\gamma}_i^*(0) = O\left(\frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}}|i|^{\frac{1}{4}}}\right)$$

for any $\xi > 0$ with very high probability by (4.80), we obtain (4.79) and complete the proof of Proposition 4.6.3. □

Proof of Proposition 4.6.6. We recall from (4.81) that

$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{\gamma}_i^*(t) = \alpha\big[-\Re m_{x,t}(\gamma_{x,i}^*(t)) + \Re m_{x,t}(\mathfrak{e}_{x,t}^+)\big] + (1-\alpha)\big[-\Re m_{y,t}(\gamma_{y,i}^*(t)) + \Re m_{y,t}(\mathfrak{e}_{y,t}^+)\big]. \tag{4.108}$$

Next, we define a dynamics that interpolates between $\mathring{z}_i(t,\alpha)$ and $\overline{\gamma}_i^*(t)$, i.e. between (4.92) and (4.108). Let $\beta \in [0,1]$ and for any fixed β define the process $v = v(t,\beta) = \{v_i(t,\beta)\}_{i=-N}^N$ as the solution of the following interpolating DBM

$$\begin{aligned} \mathrm{d}v_i =& \beta\sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \frac{1}{N}\sum_{j:|j-i|\leq L}\frac{1}{v_i - v_j}\,\mathrm{d}t + \beta\left[\frac{1}{N}\sum_{j:|j-i|>L}\frac{1}{\overline{\gamma}_i^* - \overline{\gamma}_j^*}\,\mathrm{d}t + \Phi(t)\right]\mathrm{d}t \\ &+ (1-\beta)\left[\frac{\mathrm{d}}{\mathrm{d}t}\overline{\gamma}_i^*(t) - \frac{1}{N}\sum_{j:|j-i|\leq L}\frac{1}{\overline{\gamma}_i^* - \overline{\gamma}_j^*}\right]\mathrm{d}t, \qquad 1 \leq |i| \leq N, \end{aligned} \tag{4.109}$$

with initial condition $v_i(0,\beta) := \beta\mathring{z}_i(0) + (1-\beta)\overline{\gamma}_i^*(0)$. Notice that

$$v_i(t,\beta=0) = \overline{\gamma}_i^*(t), \qquad v_i(t,\beta=1) = \mathring{z}_i(t). \tag{4.110}$$

Here we use the same letter v as in (4.98) within the proof of Lemma 4.6.4, but this is now a new interpolation. Since both appearances of the letter v are used only within the proofs of separate lemmas, this should not cause any confusion. The same remark applies to the letter u below.

Let $u := u(t,\beta) = \partial_\beta v(t,\beta)$, then it satisfies the equation

$$\mathrm{d}u_i = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \sum_{j\neq i} \mathcal{B}_{ij}(u_i - u_j)\,\mathrm{d}t + F_i\,\mathrm{d}t, \qquad 1 \le |i| \le N, \tag{4.111}$$

with a time dependent short range kernel (omitting the time argument and the β parameter)

$$\mathcal{B}_{ij}(t) = \mathcal{B}_{ij} := -\frac{1}{N}\frac{\mathbf{1}(|i-j|\le L)}{(v_i - v_j)^2} \tag{4.112}$$

and external force

$$F_i = F_i(t) := -\sum_j \frac{N^{-1}}{\overline{\gamma}_j^*(t) - \overline{\gamma}_i^*(t)} + \alpha \Re m_{x,t}(\gamma_{x,i}^*(t)) + (1-\alpha)\Re m_{y,t}(\gamma_{y,i}^*(t)) - h(t,\alpha), \tag{4.113}$$

for $1 \le |i| \le N$. Since the density $\overline{\rho}$ is regular, at least near the cusp regime, we can replace the sum over j with an integral with very high precision for small i; this integral is $\Re\overline{m}(\overline{\mathfrak{e}}^+ + \overline{\gamma}_i^*)$. A simple rearrangement of various terms yields

$$F_i = \left[\Re\overline{m}(\overline{\mathfrak{e}}^+ + \overline{\gamma}_i^*) - \frac{1}{N}\sum_j \frac{1}{\overline{\gamma}_{j\neq i}^* - \overline{\gamma}_i^*}\right] - (1-\alpha)D_{y,i} - \alpha D_{x,i} + O(N^{-1}), \tag{4.114}$$

with

$$D_{r,i} := \Re\Big[\big(\overline{m}(\overline{\mathfrak{e}}^+ + \overline{\gamma}_i^*) - \overline{m}(\overline{\mathfrak{e}}^+)\big) - \big(m_r(\gamma_{r,i}^*) - m_r(\mathfrak{e}_r^+)\big)\Big], \qquad r = x, y,$$

where we used the formula for h from (4.68) and the definition of Φ from (4.72). The choice of the shift h was governed by the idea to replace the last three terms in (4.113) by $\Re\overline{m}(\overline{\mathfrak{e}}^+ + \overline{\gamma}_i^*)$. However, the shift cannot be i dependent as it would result in an i-dependent shift in the definition of $\widetilde{z}_i$, see (4.70), which would mean that the differences (gaps) of the processes z_i and $\widetilde{z}_i$ are not the same. Therefore, we defined the shift $h(t)$ by the similar formula evaluated at the edge, justifying the choice (4.68). The discrepancy is expressed by $D_{x,i}$ and $D_{y,i}$ which are small. Indeed we have, for $r = x, y$ and $1 \le |i| \le 2i_*$ that

$$\begin{aligned}
|D_{r,i}| \le& \left|\Re\Big[\big(\overline{m}(\overline{\mathfrak{e}}^+ + \widehat{\gamma}_{r,i}^*) - \overline{m}(\overline{\mathfrak{e}}^+)\big) - \big(m_r(\mathfrak{e}_r^+ + \widehat{\gamma}_{r,i}^*) - m_r(\mathfrak{e}_r^+)\big)\Big]\right| \\
&+ |\overline{m}(\overline{\mathfrak{e}}^+ + \widehat{\gamma}_{r,i}^*) - \overline{m}(\overline{\mathfrak{e}}^+ + \overline{\gamma}_i^*)| \\
\lesssim& |\widehat{\gamma}_{r,i}^*|^{1/3}\Big[|\widehat{\gamma}_{r,i}^*|^{1/3} + N^{-\frac{1}{6}+\frac{\omega_1}{3}}\Big]|\log|\widehat{\gamma}_{r,i}^*|| + N^{-\frac{11}{36}+\omega_1} + \frac{|\widehat{\gamma}_{r,i}^* - \overline{\gamma}_i^*|}{\overline{\rho}(\overline{\gamma}_i^*)^2} \\
\lesssim& \left[\left(\frac{|i|}{N}\right)^{1/2} + \left(\frac{|i|}{N}\right)^{1/4} N^{-\frac{1}{6}+\frac{\omega_1}{3}}\right](\log N) + N^{-\frac{11}{36}+\omega_1} + \frac{\left(\frac{|i|}{N}\right) + \left(\frac{|i|}{N}\right)^{3/4} N^{-\frac{1}{6}+\omega_1}}{\left(\frac{|i|}{N}\right)^{1/2}} \\
\lesssim& N^{-\frac{1}{4}+C\omega_1},
\end{aligned} \tag{4.115}$$

where from the first to the second line we used (4.39a) and the bound on the derivative of $\overline{m}$, see (4.23b). In the last inequality we used (4.30a) to estimate $|\widehat{\gamma}_{r,i}^*| \lesssim (|i|/N)^{3/4} N^{C\omega_1}$

and similarly $|\widehat{\gamma}^*_{r,i} - \overline{\gamma}^*_i|$ in the regime $|i| \le i_* = N^{\frac{1}{2}+C_*\omega_1}$, furthermore we used that $\overline{\rho}(\overline{\gamma}^*_i) \ge (|i|/N)^{1/4}$ and also $|\overline{\gamma}^*_i| \ge c/N$, since a semiquantile is always away from the edge.

Let $\mathcal{U}(s,t)$ be the fundamental solution of the heat evolution with kernel $\mathcal{B}$ from (4.112). Similarly to (4.89), the solution to the SDE (4.111) is given by

$$u(t) = \mathcal{U}(0,t)u + \sqrt{\frac{2}{N}}\int_0^t \mathcal{U}(s,t)\,\mathrm{d}B(s) + \int_0^t \mathcal{U}(s,t)F(s)\,\mathrm{d}s. \tag{4.116}$$

The middle martingale term can be estimated as in (4.90). The last term in (4.116) is estimated by

$$|\int_0^t \mathcal{U}(s,t)F(s)\,\mathrm{d}s| \le t \max_{0\le s\le t}\|F(s)\|_\infty. \tag{4.117}$$

First we use these simple Duhamel bounds to obtain a crude rigidity bound on $v_i(t,\beta)$ by integrating the bound on u

$$|v_i(t,\beta) - v_i(t,\beta=0)| \le \beta \max_{\beta'\in[0,\beta]}|u_i(t,\beta')| \le \max_{\beta'\in[0,1]}\|u(0,\beta')\|_\infty + N^{-1/2+\omega_1+\xi}, \tag{4.118}$$

with $1 \le |i| \le N$, and for any $\xi > 0$ with very high probability, using (4.90), (4.116), (4.117) and that $\mathcal{U}$ is a contraction. Note that in the first inequality of (4.118) we used that it holds with very high probability by Markov inequality as in (4.102)-(4.103). We also used the trivial bound

$$\max_{0\le s\le t_*}\|F(s)\|_\infty \lesssim \log L \sim \log N, \tag{4.119}$$

which easily follows from (4.113),(4.115) and the fact that $|\overline{\gamma}^*_j(t) - \overline{\gamma}^*_i(t)| \gtrsim |i-j|/N$.

Recalling that $v_i(t,\beta=0) = \overline{\gamma}^*_i(t)$ and $u_i(0,\beta') = \mathring{z}_i(0) - \overline{\gamma}^*_i(0)$, together with (4.94) and (4.78), by (4.118), we obtain the crude rigidity

$$|v_i(t,\beta) - \overline{\gamma}^*_i(t)| \le N^{-\frac{1}{2}+2\omega_1}, \qquad 1 \le |i| \le N, \tag{4.120}$$

with very high probability.

The main technical result is a considerable improvement of the bound (4.120) at least for i near the cusp regime. This is the content of the following proposition whose proof is postponed:

Proposition 4.6.7. *The vector F defined in* (4.113) *satisfies the bound*

$$\max_{s\le t_*}|F_i(s)| \le N^{-\frac{1}{4}+C\omega_1}, \qquad 1 \le |i| \le 2i_*. \tag{4.121}$$

Since i_* is much bigger than $L = N^{\frac{1}{2}+C_1\omega_1}$ with a large C_1, and we have the rigidity (4.120) on scale much smaller than L, similarly to Lemma 4.6.5, we have the following finite speed of propagation result. The proof is identical to that of Lemma 4.6.5.

Proposition 4.6.8. *For the short range dynamics $\mathcal{U} = \mathcal{U}^{\mathcal{B}}$ defined by the operator* (4.112)*:*

$$\sup_{0\le s\le t\le t_*}\Big[\mathcal{U}_{pi}(s,t) + \mathcal{U}_{ip}(s,t)\Big] \le N^{-D}, \qquad 1 \le |i| \le i_*, \quad |p| \ge 2i_*. \tag{4.122}$$

for any D if N is sufficiently large. □

Armed with these two propositions, we can easily complete the proof of Proposition 4.6.6. For any $1 \leq |i| \leq i_*$ we have from (4.89), using (4.88), (4.90), (4.122) and that $\mathcal{U}$ is a contraction on ℓ^∞ that

$$
\begin{aligned}
|u_i(t)| \leq & N^{-3/4+\omega_1+\xi} + \sum_p \mathcal{U}_{ip}|u_p(0)| + \int_0^t \sum_p \mathcal{U}_{ip}(s,t)|F_p(s)| \,\mathrm{d}s \\
\leq & N^{-3/4+\omega_1+\xi} + \max_{|p|\leq 2i_*} |u_p(0)| + t \max_{0\leq s\leq t_*} \max_{|p|\leq 2i_*} |F_p(s)| + N^{-D} \max_{0\leq s\leq t} \|F(s)\|_\infty .
\end{aligned}
\tag{4.123}
$$

The trivial bound (4.119) together with (4.121) completes the proof of (4.107) by integrating back the bound (4.123) for $u = \partial_\beta v$ in β, using a high moment Markov inequality similar to (4.102)-(4.103), and recalling (4.110). This completes the proof of Proposition 4.6.6. □

4.6.3.4 Estimate of the forcing term.

Proof of Proposition 4.6.7. Within this proof we will use $\gamma_i := \overline{\gamma}_i(t)$, $\gamma_i^* := \overline{\gamma}_i^*(t)$, $\rho = \overline{\rho}_t$, $m = \overline{m}_t$ and $\mathfrak{e}^+ = \overline{\mathfrak{e}}_t^+$ for brevity. For notational simplicity we may assume within this proof that $\mathfrak{e}^+ = 0$ by a simple shift. The key input is the following bound on the derivative of the density, proven in [14] for self-consistent densities of Wigner type matrices

$$
|\rho'(x)| \leq \frac{C}{\rho(x)[\rho(x) + \Delta^{1/3}]}, \qquad |x| \leq \delta_* \tag{4.124}
$$

where $\Delta = \overline{\Delta}_t$ is the length of the unique gap in the support of $\rho = \overline{\rho}_t$ in a small neighbourhood of size $\delta_* \sim 1$ around $\mathfrak{e}^+ = 0$. If there is no such gap, then we set $\Delta = 0$ in (4.124). By the definition of the interpolated density $\overline{\rho}_t$ in (4.56) clearly follows that it satisfies (4.124) by (4.4.3). Notice that (4.124) implies local Hölder continuity, i.e.

$$
|\rho(x) - \rho(y)| \leq \min\left\{|x-y|^{1/3}, |x-y|^{1/2}\Delta^{-1/6}\right\} \tag{4.125}
$$

for any x, y in a small neighbourhood of the gap or the local minimum.

Throughout the entire proof we fix an i with $1 \leq |i| \leq 2i_*$. For simplicity, we assume $i > 0$, the case $i < 0$ is analogous. We rewrite F_i from (4.114) as follows

$$
F_i = G_1 + G_2 + G_3 + G_4 \tag{4.126}
$$

with

$$
G_1 := \sum_{1\leq|j-i|\leq L} \int_{\gamma_{j-1}}^{\gamma_j} \left[\frac{1}{x-\gamma_i^*} - \frac{1}{\gamma_j^*-\gamma_i^*}\right] \rho(x)\,\mathrm{d}x, \qquad G_2 := \int_{\gamma_{i-1}}^{\gamma_i} \frac{\rho(x)\,\mathrm{d}x}{x-\gamma_i^*},
$$

$$
G_3 := \sum_{|j-i|>L} \int_{\gamma_{j-1}}^{\gamma_j} \left[\frac{1}{x-\gamma_i^*} - \frac{1}{\gamma_j^*-\gamma_i^*}\right] \rho(x)\,\mathrm{d}x, \;\; G_4 := -(1-\alpha)D_{y,i} - \alpha D_{x,i} + O(N^{-1}).
$$

The term G_4 was already estimated in (4.115). In the following we will show separately that $|G_a| \lesssim N^{-1/4}$, $a = 1, 2, 3$.

Estimate of G_3. By elementary computations, using the crude rigidity (4.78), it follows that

$$
|G_3| \lesssim \frac{N^{-\frac{1}{2}+2\omega_1}}{N} \sum_{j:|j-i|>L} \frac{1}{(\gamma_i^*-\gamma_j^*)^2}.
$$

Then, the estimate $|G_3| \lesssim N^{-\frac{1}{4}}$ follows using the same computations as in (4.97).

Estimate of G_2. We write

$$G_2 = \int_{\gamma_{i-1}}^{\gamma_i} \frac{\rho(x)\,\mathrm{d}x}{x-\gamma_i^*} = \int_{\gamma_{i-1}}^{\gamma_i} \frac{\rho(x)-\rho(\gamma_i^*)}{x-\gamma_i^*}\,\mathrm{d}x + \rho(\gamma_i^*)\int_{\gamma_{i-1}}^{\gamma_i} \frac{\mathrm{d}x}{x-\gamma_i^*} \tag{4.127}$$

and we will show that both summands are bounded by $CN^{-1/4}$. We make the convention that if γ_{i-1} is exactly at the left edge of a gap, then for the purpose of this proof we redefine it to be the right edge of the same gap and similarly, if γ_i is exactly at the right edge of the gap, then we set it to be left edge. This is just to make sure that $[\gamma_{i-1}, \gamma_i]$ is always included in the support of ρ.

In the first integral we use (4.125) to get

$$\left|\int_{\gamma_{i-1}}^{\gamma_i} \frac{\rho(x)-\rho(\gamma_i^*)}{x-\gamma_i^*}\,\mathrm{d}x\right| \lesssim \min\left\{(\gamma_i-\gamma_{i-1})^{1/3}, (\gamma_i-\gamma_{i-1})^{1/2}\Delta^{-1/6}\right\} = O(N^{-1/4}). \tag{4.128}$$

Here we used that the local eigenvalue spacing (with the convention above) is bounded by

$$\gamma_i - \gamma_{i-1} \lesssim \max\left\{\frac{\Delta^{1/9}}{N^{2/3}}, \frac{1}{N^{3/4}}\right\}. \tag{4.129}$$

For the second integral in (4.127) is an explicit calculation

$$\rho(\gamma_i^*)\int_{\gamma_{i-1}}^{\gamma_i} \frac{\mathrm{d}x}{x-\gamma_i^*} = \rho(\gamma_i^*)\log\frac{\gamma_i-\gamma_i^*}{\gamma_i^*-\gamma_{i-1}}. \tag{4.130}$$

Using the definition of the quantiles and (4.125), we have

$$\frac{1}{2N} = \int_{\gamma_{i-1}}^{\gamma_i^*} \rho(x)\,\mathrm{d}x = \rho(\gamma_i^*)(\gamma_i^*-\gamma_{i-1}) + O\Big(\min\left\{|\gamma_i^*-\gamma_{i-1}|^{4/3}, |\gamma_i^*-\gamma_{i-1}|^{3/2}\Delta^{-1/6}\right\}\Big),$$

and similarly

$$\frac{1}{2N} = \int_{\gamma_i^*}^{\gamma_i} \rho(x)\,\mathrm{d}x = \rho(\gamma_i^*)(\gamma_i-\gamma_i^*) + O\Big(\min\left\{|\gamma_i^*-\gamma_i|^{4/3}, |\gamma_i^*-\gamma_i|^{3/2}\Delta^{-1/6}\right\}\Big).$$

The error terms are comparable and they are $O(N^{-1})$ using (4.129), thus, subtracting these two equations, we have

$$|(\gamma_i-\gamma_i^*)-(\gamma_i^*-\gamma_{i-1})| \lesssim \frac{\min\left\{|\gamma_i^*-\gamma_i|^{4/3}, |\gamma_i^*-\gamma_i|^{3/2}\Delta^{-1/6}\right\}}{\rho(\gamma_i^*)}.$$

Expanding the logarithm in (4.130), we have

$$\begin{aligned}\left|\rho(\gamma_i^*)\int_{\gamma_{i-1}}^{\gamma_i} \frac{\mathrm{d}x}{x-\gamma_i^*}\right| &\lesssim \rho(\gamma_i^*)\frac{|(\gamma_i-\gamma_i^*)-(\gamma_i^*-\gamma_{i-1})|}{\gamma_i^*-\gamma_{i-1}} \\ &\lesssim \min\left\{|\gamma_i^*-\gamma_i|^{1/3}, |\gamma_i^*-\gamma_i|^{1/2}\Delta^{-1/6}\right\} \lesssim N^{-1/4}\end{aligned}$$

as in (4.128). This completes the estimate

$$|G_2| \lesssim N^{-1/4}. \tag{4.131}$$

Estimate of G_1. Fix $i > 0$ and set $n = n(i)$ as follows

$$n(i) := \min\Big\{n \in \mathbf{N} \ : \ \min\left\{|\gamma_{i-n-1} - \gamma_i^*|, |\gamma_{i+n} - \gamma_i^*|\right\} \geq cN^{-3/4}\Big\} \tag{4.132}$$

with some small constant $c > 0$.

Next, we estimate $n(i)$. Notice that for $i = 1$ we have $n(i) = 0$. If $i \geq 2$, then we notice that one can choose c sufficiently small depending only on the model parameters, such that

$$\frac{1}{2} \leq \frac{\rho(x)}{\rho(\gamma_i^*)} \leq 2 \ : \ \forall x \in [\gamma_{i-n(i)-1}, \gamma_{i+n(i)}], \quad i \geq 2. \tag{4.133}$$

Let

$$m(i) := \max\Big\{m \in \mathbf{N} \ : \ \frac{1}{2} \leq \frac{\rho(x)}{\rho(\gamma_i^*)} \leq 2 \ : \ \forall x \in [\gamma_{i-m-1}, \gamma_{i+m}]\Big\},$$

then, in order to verify (4.133), we need to prove that $m(i) \geq n(i)$.

Then by a case by case calculation it follows that

$$m(i) \geq c_1|i|, \tag{4.134}$$

and thus

$$\min\Big\{|\gamma_{i-m(i)-1} - \gamma_i^*|, |\gamma_{i+m(i)} - \gamma_i^*|\Big\} \gtrsim \max\Big\{\Big(\frac{i}{N}\Big)^{2/3}\Delta^{1/9}, \Big(\frac{i}{N}\Big)^{3/4}\Big\} \geq c_2 N^{-3/4}. \tag{4.135}$$

with some c_1, c_2. Hence (4.133) will hold if $c \leq c_2$ is chosen in the definition (4.132). Notice that in these estimates it is important that the semiquantiles are always at a certain distance away from the quantiles.

Now we give an upper bound on $n(i)$ when γ_i^* is near a (possible small) gap as in the proof above. The local eigenvalue spacing is

$$\gamma_i - \gamma_i^* \sim \max\Big\{\frac{\Delta^{1/9}}{N^{2/3}(i)^{1/3}}, \frac{1}{N^{3/4}(i)^{1/4}}\Big\}, \tag{4.136}$$

which is bigger than $cN^{-3/4}$ if $i \leq \Delta^{1/3}N^{1/4}$. So in this case $n(i) = 0$ and we may now assume that $i \geq \Delta^{1/3}N^{1/4}$ and still $i \geq 2$.

Consider first the so-called *cusp case* when $i \geq N\Delta^{4/3}$, in this case, as long as $n \leq \frac{1}{2}i$, we have

$$\gamma_{i+n} - \gamma_i^* \sim \frac{n}{N^{3/4}(i+1)^{1/4}}.$$

This is bigger than $cN^{-3/4}$ if $n \geq i^{1/4}$, thus we have $n(i) \leq i^{1/4}$ in this case.

In the opposite case, the so-called *edge case*, $i \leq N\Delta^{4/3}$, which together with the above assumption $i \geq \Delta^{1/3}N^{1/4}$ also implies that $\Delta \geq N^{-3/4}$. In this case, as long as $n \leq \frac{1}{2}i$, we have

$$\gamma_{i+n} - \gamma_i^* \sim \frac{n\Delta^{1/9}}{N^{2/3}i^{1/3}}.$$

This is bigger than $cN^{-3/4}$ if $n \geq \Delta^{-1/9}N^{-1/12}i^{1/3}$. So we have $n(i) \leq \Delta^{-1/9}N^{-1/12}i^{1/3} \leq i^{1/3}$ in this case.

We split the sum in the definition of G_1, see (4.126), as follows:

$$\begin{aligned} G_1 &= \sum_{1\le|j-i|\le L} \int_{\gamma_{j-1}}^{\gamma_j} \frac{x-\gamma_j^*}{(\gamma_i^*-\gamma_j^*)(x-\gamma_i^*)}\rho(x)\,\mathrm{d}x \\ &= \Big(\sum_{n(i)<|j-i|\le L} + \sum_{1\le|j-i|\le n(i)}\Big) =: S_1+S_2. \end{aligned} \tag{4.137}$$

For the first sum we use $|x-\gamma_j^*|\le\gamma_{j+1}^*-\gamma_j^*$, $|\gamma_i^*-x|\sim|\gamma_i^*-\gamma_j^*|$. Moreover, we have

$$\rho(\gamma_i^*)(\gamma_i-\gamma_{i-1})\sim\frac{1}{N} \tag{4.138}$$

from the definition of the semiquantiles. Thus we restore the integration in the first sum S_1 and estimate

$$\begin{aligned} |S_1| &\lesssim \frac{1}{N}\left[\int_{-\infty}^{\gamma_{i-n(i)-1}} + \int_{\gamma_{i+n(i)}}^{\infty}\right]\frac{\mathrm{d}x}{|x-\gamma_i^*|^2} \\ &\lesssim \frac{1}{N}\Big[\frac{1}{|\gamma_{i-n(i)-1}-\gamma_i^*|}+\frac{1}{|\gamma_{i+n(i)}-\gamma_i^*|}\Big]\le CN^{-1/4}. \end{aligned} \tag{4.139}$$

In the last step we used the definition of $n(i)$.

Now we consider S_2. Notice that this sum is non-empty only if $n(i)\neq 0$ In this case to estimate S_2 we have to symmetrize. Fix $1\le n\le n(i)$, assume $i>n$ and consider together

$$\int_{\gamma_{i-n-1}}^{\gamma_{i-n}} \frac{x-\gamma_{i-n}^*}{(\gamma_i^*-\gamma_{i-n}^*)(x-\gamma_i^*)}\rho(x)\,\mathrm{d}x + \int_{\gamma_{i+n-1}}^{\gamma_{i+n}} \frac{x-\gamma_{i+n}^*}{(\gamma_i^*-\gamma_{i+n}^*)(x-\gamma_i^*)}\rho(x)\,\mathrm{d}x$$

$$= \frac{1}{\gamma_i^*-\gamma_{i-n}^*}\int_{\gamma_{i-n-1}}^{\gamma_{i-n}}\frac{x-\gamma_{i-n}^*}{x-\gamma_i^*}\rho(x)\,\mathrm{d}x + \frac{1}{\gamma_i^*-\gamma_{i+n}^*}\int_{\gamma_{i+n-1}}^{\gamma_{i+n}}\frac{x-\gamma_{i+n}^*}{x-\gamma_i^*}\rho(x)\,\mathrm{d}x \tag{4.140}$$

$$= \frac{1}{N}\Big[\frac{1}{\gamma_i^*-\gamma_{i-n}^*}+\frac{1}{\gamma_i^*-\gamma_{i+n}^*}\Big]+\left[\int_{\gamma_{i-n-1}}^{\gamma_{i-n}}\frac{\rho(x)\,\mathrm{d}y}{x-\gamma_i^*}+\int_{\gamma_{i+n-1}}^{\gamma_{i+n}}\frac{\rho(x)\,\mathrm{d}x}{x-\gamma_i^*}\right] =: B_1(n)+B_2(n).$$

We now use $\frac{1}{3}$-Hölder regularity

$$\rho(x)=\rho(\gamma_i^*)+O\Big(|x-\gamma_i^*|^{1/3}\Big).$$

We thus have

$$\sum_{n\le n(i)}\int_{\gamma_{i-n-1}}^{\gamma_{i-n}}\frac{\rho(x)\,\mathrm{d}y}{x-\gamma_i^*} = \sum_{n\le n(i)}\rho(\gamma_i^*)\log\frac{\gamma_{i-n-1}-\gamma_i^*}{\gamma_{i-n}-\gamma_i^*}+O\Big(\int_{\gamma_{i-n(i)-1}}^{\gamma_{i+n(i)}}\frac{\mathrm{d}x}{|x-\gamma_i^*|^{2/3}}\Big) \tag{4.141}$$

and similarly

$$\sum_{n\le n(i)}\int_{\gamma_{i+n-1}}^{\gamma_{i+n}}\frac{\rho(x)\,\mathrm{d}y}{x-\gamma_i^*} = \sum_{n\le n(i)}\rho(\gamma_i^*)\log\frac{\gamma_{i+n-1}-\gamma_i^*}{\gamma_{i+n}-\gamma_i^*}+O\Big(\int_{\gamma_{i-n(i)-1}}^{\gamma_{i+n(i)}}\frac{\mathrm{d}x}{|x-\gamma_i^*|^{2/3}}\Big). \tag{4.142}$$

The error terms are bounded by $CN^{-1/4}$ using (4.132) and therefore we have

$$\sum_{n\le n(i)} B_2(n) = \sum_{n\le n(i)} \rho(\gamma_i^*)\Big[\log\frac{\gamma_i^*-\gamma_{i-n-1}}{\gamma_i^*-\gamma_{i-n}} - \log\frac{\gamma_{i+n}-\gamma_i^*}{\gamma_{i+n-1}-\gamma_i^*}\Big] + O(N^{-1/4})$$

$$= \sum_{n\le n(i)} \rho(\gamma_i^*)\Big[\log\frac{\gamma_i^*-\gamma_{i-n-1}}{\gamma_{i+n}-\gamma_i^*} + \log\frac{\gamma_{i+n-1}-\gamma_i^*}{\gamma_i^*-\gamma_{i-n}}\Big] + O(N^{-1/4}).$$

We now use the bound

$$|\rho(x)-\rho(\gamma_i^*)| \lesssim \frac{|x-\gamma_i^*|}{\rho(\gamma_i^*)^2+\rho(\gamma_i^*)\Delta^{1/3}}, \qquad x\in[\gamma_{i-n(i)-1},\gamma_{i+n(i)}], \tag{4.143}$$

which follows from the derivative bound (4.124) if ϵ in the definition of $i_* = \epsilon N$ is chosen sufficiently small, depending on δ since throughout the proof $1\le |i|\le 2i_*$ and $n(i)\ll i_*$.

Note that

$$\frac{n}{N} = \int_{\gamma_{i-n}}^{\gamma_i} \rho(x)\,\mathrm{d}x = \rho(\gamma_i^*)[\gamma_i-\gamma_{i-n}] + O\Big(\frac{|\gamma_{i-n}-\gamma_i^*|^2}{\rho(\gamma_i^*)^2+\rho(\gamma_i^*)\Delta^{1/3}}\Big) \tag{4.144}$$

Thus, using (4.144) also for $\gamma_{i+n}-\gamma_i$, equating the two equations and dividing by $\rho(\gamma_i^*)$, we have

$$\gamma_i-\gamma_{i-n} = \gamma_{i+n}-\gamma_i + O\Big(\frac{|\gamma_{i-n}-\gamma_i^*|^2}{\rho(\gamma_i^*)^3+\rho(\gamma_i^*)^2\Delta^{1/3}}\Big). \tag{4.145}$$

Similar relation holds for the semiquantiles:

$$\gamma_i^*-\gamma_{i-n}^* = \gamma_{i+n}^*-\gamma_i^* + O\Big(\frac{|\gamma_{i-n}^*-\gamma_i^*|^2}{\rho(\gamma_i^*)^3+\rho(\gamma_i^*)^2\Delta^{1/3}}\Big) \tag{4.146}$$

and for the mixed relations among quantiles and semiquantiles:

$$\gamma_i^*-\gamma_{i-n} = \gamma_{i+n-1}-\gamma_i^* + O\Big(\frac{|\gamma_{i-n}-\gamma_i^*|^2}{\rho(\gamma_i^*)^3+\rho(\gamma_i^*)^2\Delta^{1/3}}\Big)$$

$$\gamma_i^*-\gamma_{i-n-1} = \gamma_{i+n}-\gamma_i^* + O\Big(\frac{|\gamma_{i-n}-\gamma_i^*|^2}{\rho(\gamma_i^*)^3+\rho(\gamma_i^*)^2\Delta^{1/3}}\Big).$$

Thus, using $\gamma_i^*-\gamma_{i-n-1}\sim\gamma_{i+n}-\gamma_i^*$, we have

$$\rho(\gamma_i^*)|\log\frac{\gamma_i^*-\gamma_{i-n-1}}{\gamma_{i+n}-\gamma_i^*}| \lesssim \frac{\rho(\gamma_i^*)}{\gamma_{i+n}-\gamma_i^*}O\Big(\frac{|\gamma_{i-n-1}-\gamma_i^*|^2}{\rho(\gamma_i^*)^3+\rho(\gamma_i^*)^2\Delta^{1/3}}\Big) \lesssim \frac{|\gamma_{i-n-1}-\gamma_i^*|}{\rho(\gamma_i^*)^2+\rho(\gamma_i^*)\Delta^{1/3}}. \tag{4.147}$$

Using $n\le n(i)$ and (4.132), we have $|\gamma_{i-n-1}-\gamma_i^*|\lesssim N^{-3/4}$. The contribution of this term to $\sum_n B_2(n)$ is thus

$$N^{-3/4}\sum_{n\le n(i)}\frac{1}{\rho(\gamma_i^*)^2+\rho(\gamma_i^*)\Delta^{1/3}} \le \frac{n(i)N^{-3/4}}{\rho(\gamma_i^*)^2+\rho(\gamma_i^*)\Delta^{1/3}}. \tag{4.148}$$

In the bulk regime we have $\rho(\gamma_i^*)\sim 1$ and $n(i)\sim N^{1/4}$, so this contribution is much smaller than $N^{-1/4}$.

In the cusp regime, i.e. when $\Delta \le (i/N)^{3/4}$, then we have $\gamma_i^* \sim (i/N)^{3/4}$ and $\rho(\gamma_i^*) \sim (i/N)^{1/4}$, thus we get

$$(4.148) \le \frac{n(i)N^{-3/4}}{\rho(\gamma_i^*)^2 + \rho(\gamma_i^*)\Delta^{1/3}} \le \frac{n(i)N^{-3/4}}{\rho(\gamma_i^*)^2} \lesssim N^{-1/4}n(i)i^{-1/2} \lesssim N^{-1/4}$$

since $n(i) \le i^{1/4}$.

In the edge regime, i.e. when $\Delta \ge (i/N)^{3/4}$, then we have $\gamma_i^* \sim \Delta^{1/9}(i/N)^{2/3}$ and $\rho(\gamma_i^*) \sim \Delta^{-1/9}(i/N)^{1/3}$, thus we get

$$(4.148) \le \frac{n(i)N^{-3/4}}{\rho(\gamma_i^*)^2 + \rho(\gamma_i^*)\Delta^{1/3}} \le \frac{n(i)N^{-3/4}}{\rho(\gamma_i^*)\Delta^{1/3}} \lesssim \frac{n(i)N^{-5/12}}{\Delta^{2/9}i^{1/3}} \le \frac{N^{-5/12}}{\Delta^{2/9}} \le N^{-1/4}$$

since $n(i) \le i^{1/3}$ and $\Delta \ge N^{-3/4}$. This completes the proof of $\sum_n B_2(n) \lesssim N^{-1/4}$.

Finally the $\sum_n B_1(n)$ term from (4.140) is estimated as follows by using (4.146):

$$\begin{aligned}\sum_n \frac{1}{N}\Big[\frac{1}{\gamma_i^* - \gamma_{i-n-1}^*} + \frac{1}{\gamma_i^* - \gamma_{i+n-1}^*}\Big] &= \sum_n \frac{1}{N}\frac{1}{(\gamma_i^* - \gamma_{i-n}^*)^2}\mathcal{O}\Big(\frac{(\gamma_i - \gamma_{i-n-1})^2}{\rho(\gamma_i^*)^2[\rho(\gamma_i^*) + \Delta^{1/3}]}\Big) \\ &\lesssim \frac{n(i)}{N\rho(\gamma_i^*)^2[\rho(\gamma_i^*) + \Delta^{1/3}]}.\end{aligned} \tag{4.149}$$

In the bulk regime this is trivially bounded by $CN^{-3/4}$. In the cusp regime, $\Delta \le (i/N)^{3/4}$, we have

$$\frac{n(i)}{N\rho(\gamma_i^*)^2[\rho(\gamma_i^*) + \Delta^{1/3}]} \le \frac{n(i)}{N\rho(\gamma_i^*)^3} \lesssim \frac{n(i)}{N^{1/4}i^{3/4}} \lesssim N^{-1/4}$$

since $n(i) \le i^{1/4}$.

Finally, in the edge regime, $\Delta \ge (i/N)^{3/4}$, we just use

$$\frac{n(i)}{N\rho(\gamma_i^*)^2[\rho(\gamma_i^*) + \Delta^{1/3}]} \le \frac{n(i)}{N\rho(\gamma_i^*)^2\Delta^{1/3}} \lesssim \frac{n(i)}{N^{1/4}i^{3/4}} \lesssim N^{-1/4}$$

since $n(i) \le i^{1/3}$. This gives $\sum_n B_1(n) \lesssim N^{-1/4}$. Together with the estimate on $\sum_n B_2(n)$ we get $|S_2| \lesssim N^{-1/4}$, see (4.137) and (4.140). This completes the estimate of G_1 in (4.126), which, together with (4.131) and (4.115) finishes the proof of Proposition 4.6.7. □

4.6.4 Phase 2: Rigidity of $\widehat{z}$ on scale $N^{-3/4+\omega_1/6}$, without i dependence

For any fixed $\alpha \in [0,1]$ recall the definition of the shifted process $\widetilde{z}(t,\alpha)$ (4.71) and the shifted α-interpolating semiquantiles $\overline{\gamma}_i^*(t)$ from (4.54) that trail $\widetilde{z}$. Furthermore, for all $0 \le t \le t_*$ we consider the interpolated density $\overline{\rho}_t$ with a small gap $[\overline{\mathfrak{e}}_t^-, \overline{\mathfrak{e}}_t^+]$, and its Stieltjes transform $\overline{m}_t$. In particular,

$$\overline{\mathfrak{e}}_t^\pm = \alpha\mathfrak{e}_{x,t}^\pm + (1-\alpha)\mathfrak{e}_{y,t}^\pm.$$

We recall that by Proposition 4.6.3 and (4.80) we have that

$$\sup_{0\le t\le t_*}\max_{1\le|i|\le i_*}|\widetilde{z}_i(t,\alpha) - \overline{\gamma}_i^*(t)| \le N^{-\frac{3}{4}+C\omega_1}, \tag{4.150}$$

holds with very high probability for some $i_* = N^{\frac{1}{2}+C_*\omega_1}$.

In this section we improve the rigidity (4.150) from scale $N^{-\frac{3}{4}+C\omega_1}$ to the almost optimal, but still i-independent rigidity of order $N^{-\frac{3}{4}+\frac{\omega_1}{6}+\xi}$ but only for a new short range approximation $\widehat{z}_i(t,\alpha)$ of $\widetilde{z}_i(t,\alpha)$. The range of this new approximation $\ell^4 = N^{4\omega_\ell}$ with some $\omega_\ell \ll 1$ is much shorter than that of $\mathring{z}_i(t,\alpha)$ in Section 4.6.3. Furthermore, the result will hold only for $1 \le |i| \le N^{4\omega_\ell+\delta_1}$, for some small $\delta_1 > 0$. The rigorous statement is in Proposition 4.6.10 below, after we give the definition of the short range approximation.

Short range approximation on fine scale.

Adapting the idea of [131] to the cusp regime, we now introduce a new short range approximation process $\widehat{z}(t,\alpha)$ for the solution to (4.71). The short range approximation in this section will always be denoted by hat, $\widehat{z}$, in distinction to the other short range approximation $\mathring{z}$ used in Section 4.6.3, see (4.92). Not only the length scale is shorter for $\widehat{z}$, but the definition of $\widehat{z}$ is more subtle than in (4.92)

The new short scale approximation is characterized by two exponents ω_ℓ and ω_A. In particular, we will always assume that $\omega_1 \ll \omega_\ell \ll \omega_A \ll 1$, where recall that $t_* \sim N^{-\frac{1}{2}+\omega_1}$ is defined in such a way $\overline{\rho}_{t_*}$ has an exact cusp. The key quantity is $\ell := N^{\omega_\ell}$ that determines the scale on which the interaction term in (4.71) will be cut off and replaced by its mean-field value. This scale is not constant, it increases away from the cusp at a certain rate. The cutoff will be effective only near the cusp, for indices beyond $\frac{i_*}{2}$, with $i_* = N^{\frac{1}{2}+C_*\omega_1}$, no cutoff is made. Finally, the intermediate scale N^{ω_A} is used for a technical reason: closer to the cusp, for indices less than N^{ω_A}, we always use the density $\rho_{y,t}$ of the reference process $y(t)$ to define the mean field approximation of the cutoff long range terms. Beyond this scale we use the actual density $\overline{\rho}_t$. In this way we can exploit the closeness of the density $\overline{\rho}_t$ to the reference density $\rho_{y,t}$ near the cusp and simplify the estimate. This choice will guarantee that the error term ζ_0 in (4.162) below is non zero only for $|i| > N^{\omega_A}$.

Now we define the $\widehat{z}$ process precisely. Let

$$\mathcal{A} := \left\{(i,j) : |i-j| \le \ell(10\ell^3 + |i|^{\frac{3}{4}} + |j|^{\frac{3}{4}})\right\} \cup \left\{(i,j) : |i|,|j| > \frac{i_*}{2}\right\}. \tag{4.151}$$

One can easily check that for each i with $1 \le |i| \le \frac{i_*}{2}$ the set $\{j : (i,j) \in \mathcal{A}\}$ is an interval of the nonzero integers and that $(i,j) \in \mathcal{A}$ if and only if $(j,i) \in \mathcal{A}$. For each such fixed i we denote the smallest and the biggest j such that $(i,j) \in \mathcal{A}$ by $j_-(i)$ and $j_+(i)$, respectively. We will use the notation

$$\sum_j^{\mathcal{A},(i)} := \sum_{\substack{j:(i,j)\in\mathcal{A} \\ i\neq j}}, \qquad \sum_j^{\mathcal{A}^c,(i)} := \sum_{j:(i,j)\notin\mathcal{A}}.$$

Assuming for simplicity that i_* is divisible by four, we introduce the intervals

$$\mathcal{J}_z(t) := \left[\overline{\gamma}_{-\frac{3i_*}{4}}(t), \overline{\gamma}_{\frac{3i_*}{4}}(t)\right], \tag{4.152}$$

and for each $0 < |i| \le \frac{i_*}{2}$ we define

$$\mathcal{I}_{z,i}(t) := [\overline{\gamma}_{j_-(i)}(t), \overline{\gamma}_{j_+(i)}(t)]. \tag{4.153}$$

For a fixed $\alpha \in [0,1]$ and $N \geq |i| > \frac{i_*}{2}$ we let

$$\mathrm{d}\widehat{z}_i(t,\alpha) = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \Bigg[\frac{1}{N}\sum_j^{\mathcal{A},(i)} \frac{1}{\widehat{z}_i(t,\alpha)-\widehat{z}_j(t,\alpha)} + \frac{1}{N}\sum_j^{\mathcal{A}^c,(i)} \frac{1}{\widetilde{z}_i(t,\alpha)-\widetilde{z}_j(t,\alpha)} + \Phi_\alpha(t)\Bigg]\mathrm{d}t \tag{4.154}$$

for $0 < |i| \leq N^{\omega_A}$

$$\mathrm{d}\widehat{z}_i(t,\alpha) = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \Bigg[\frac{1}{N}\sum_j^{\mathcal{A},(i)} \frac{1}{\widehat{z}_i(t,\alpha)-\widehat{z}_j(t,\alpha)} + \int_{\mathcal{I}_{y,i}(t)^c} \frac{\rho_{y,t}(E+\mathfrak{e}_{y,t}^+)}{\widehat{z}_i(t,\alpha)-E}\,\mathrm{d}E + \Re[m_{y,t}(\mathfrak{e}_{y,t}^+)]\Bigg]\mathrm{d}t, \tag{4.155}$$

and for $N^{\omega_A} < |i| \leq \frac{i_*}{2}$

$$\mathrm{d}\widehat{z}_i(t,\alpha) = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \Bigg[\frac{1}{N}\sum_j^{\mathcal{A},(i)} \frac{1}{\widehat{z}_i(t,\alpha)-\widehat{z}_j(t)} + \int_{\mathcal{I}_{z,i}(t)^c\cap\mathcal{J}_z(t)} \frac{\overline{\rho}_t(E+\overline{\mathfrak{e}}_t^+)}{\widehat{z}_i(t,\alpha)-E}\,\mathrm{d}E + \frac{1}{N}\sum_{|j|\geq\frac{3}{4}i_*} \frac{1}{\widetilde{z}_i(t,\alpha)-\widetilde{z}_j(t,\alpha)} + \Phi_\alpha(t)\Bigg]\mathrm{d}t, \tag{4.156}$$

with initial data

$$\widehat{z}_i(0,\alpha) := \widetilde{z}_i(0,\alpha), \tag{4.157}$$

where we recall that $\widetilde{z}_i(0,\alpha) = \alpha\widetilde{x}_i(0) + (1-\alpha)\widetilde{y}_i(0)$ for any $\alpha \in [0,1]$. In particular, $\widehat{z}(t,1) = \widehat{x}(t)$ and $\widehat{z}(t,0) = \widehat{y}(t)$, that are the short range approximations of the $\widetilde{x}(t) := x(t) - \mathfrak{e}_{x,t}^+$ and $\widetilde{y}(t) := x(t) - \mathfrak{e}_{y,t}^+$ processes.

Using the rigidity estimates in (4.78) and (4.150) we will prove the following lemma in Appendix 4.C.

Lemma 4.6.9. *Assuming that the rigidity estimates* (4.78) *and* (4.150) *hold. Then, for any fixed* $\alpha \in [0,1]$ *we have*

$$\sup_{1\leq|i|\leq N}\sup_{0\leq t\leq t_*} |\widehat{z}_i(t,\alpha) - \widetilde{z}_i(t,\alpha)| \leq \frac{N^{C\omega_1}}{N^{\frac{3}{4}}}, \tag{4.158}$$

with very high probability.

In particular, since (4.78) and (4.150) have already been proven, we conclude from (4.150) and (4.158) that

$$\sup_{0\leq t\leq t_*} |\widehat{z}_i(t,\alpha) - \overline{\gamma}_i(t)| \leq \frac{N^{C\omega_1}}{N^{\frac{3}{4}}}, \qquad 1 \leq |i| \leq i_*, \tag{4.159}$$

for any fixed $\alpha \in [0,1]$.

Now we state the improved rigidity for $\widehat{z}$, the main result of Section 4.6.4:

Proposition 4.6.10. *Fix any $\alpha \in [0,1]$. There exists a constant $C > 0$ such that if $0 < \delta_1 < C\omega_\ell$ then*

$$\sup_{0\le t\le t_*} |\widehat{z}_i(t,\alpha) - \overline{\gamma}_i(t)| \lesssim \frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}}}, \qquad 1 \le |i| \le N^{4\omega_\ell+\delta_1} \tag{4.160}$$

for any $\xi > 0$ with very high probability.

Proof. Recall that initially $\widetilde{z}_i(0,\alpha)$ is a linear interpolation between $\widetilde{x}_i(0)$ and $\widetilde{y}_i(0)$ and thus for $\widetilde{z}_i(0,\alpha)$ optimal rigidity (4.80) holds. We define the derivative process

$$w_i(t,\alpha) := \partial_\alpha \widehat{z}_i(t,\alpha). \tag{4.161}$$

In particular, we find that $w = w(t,\alpha)$ is the solution of

$$\partial_t w = \mathcal{L}w + \zeta^{(0)}, \qquad \mathcal{L} := \mathcal{B} + \mathcal{V}, \tag{4.162}$$

with initial data

$$w_i(0,\alpha) = \widehat{x}_i(0) - \widehat{y}_i(0).$$

Here, for any $1 \le |i| \le N$, the (short range) operator $\mathcal{B}$ is defined on any vector $f \in \mathbf{C}^{2N}$ as

$$(\mathcal{B}f)_i := \sum_j^{\mathcal{A},(i)} \mathcal{B}_{ij}(f_i - f_j), \qquad \mathcal{B}_{ij} := -\frac{1}{N}\frac{1}{(\widehat{z}_i(t,\alpha) - \widehat{z}_j(t,\alpha))^2}. \tag{4.163}$$

Moreover, $\mathcal{V}$ is a multiplication operator, i.e. $(\mathcal{V}f)_i = \mathcal{V}_i f_i$, where $\mathcal{V}_i$ is defined in different regimes of i as follows:

$$\begin{aligned} \mathcal{V}_i &:= -\int_{\mathcal{I}_{y,i}(t)^c} \frac{\rho_{y,t}(E + \mathfrak{e}_{y,t}^+)}{(\widehat{z}_i(t,\alpha) - E)^2}\, \mathrm{d}E, \qquad 1 \le |i| \le N^{\omega_A} \\ \mathcal{V}_i &:= -\int_{\mathcal{I}_{z,i}(t)^c \cap \mathcal{J}_z(t)} \frac{\overline{\rho}_t(E + \overline{\mathfrak{e}}_t^+)}{(\widehat{z}_i(t,\alpha) - E)^2}\, \mathrm{d}E, \qquad N^{\omega_A} < |i| \le \frac{i_*}{2} \end{aligned} \tag{4.164}$$

and $\mathcal{V}_i = 0$ for $|i| > \frac{i_*}{2}$. The error term $\zeta_i^{(0)} = \zeta_i^{(0)}(t)$ in (4.162) is defined as follows: for $|i| > \frac{i_*}{2}$ we have

$$\zeta_i^{(0)} := \frac{1}{N}\sum_j^{\mathcal{A}^c,(i)} \frac{\partial_\alpha \widetilde{z}_j(t,\alpha) - \partial_\alpha \widetilde{z}_i(t,\alpha)}{(\widetilde{z}_i(t,\alpha) - \widetilde{z}_j(t,\alpha))^2} + \partial_\alpha \Phi_\alpha(t) =: Z_1 + \partial_\alpha \Phi_\alpha(t) \tag{4.165}$$

for $N^{\omega_A} < |i| \le \frac{i_*}{2}$ we have

$$\begin{aligned} \zeta_i^{(0)} := &\frac{1}{N}\sum_{|j|\ge \frac{3i_*}{4}} \frac{\partial_\alpha \widetilde{z}_j(t,\alpha) - \partial_\alpha \widetilde{z}_i(t,\alpha)}{(\widetilde{z}_i(t,\alpha) - \widetilde{z}_j(t,\alpha))^2} + \int_{\mathcal{I}_{z,i}(t)^c \cap \mathcal{J}_z(t)} \frac{\partial_\alpha[\overline{\rho}_t(E + \overline{\mathfrak{e}}_t^+)]}{\widehat{z}_i(t,\alpha) - E}\, \mathrm{d}E \\ &+ \Big(\partial_\alpha \int_{\mathcal{I}_{z,i}(t)^c \cap \mathcal{J}_z(t)}\Big) \frac{\overline{\rho}_t(E + \overline{\mathfrak{e}}_t^+)}{\widehat{z}_i(t,\alpha) - E}\, \mathrm{d}E + \partial_\alpha \Phi_\alpha(t) =: Z_2 + Z_3 + Z_4 + \partial_\alpha \Phi_\alpha(t), \end{aligned} \tag{4.166}$$

and finally for $1 \le |i| \le N^{\omega_A}$ we have $\zeta_i^{(0)} = 0$. We recall that $\mathcal{I}_{z,i}(t)$ and $\mathcal{J}_z(t)$ in (4.166) are defined by (4.153) and (4.152) respectively. Next, we prove that the error term $\zeta^{(0)}$ in (4.162) is bounded by some large power of N.

Lemma 4.6.11. *There exists a large constant $C > 0$ such that*

$$\sup_{0 \le t \le t_*} \max_{1 \le |i| \le N} |\zeta_i^{(0)}(t)| \le N^C. \tag{4.167}$$

Proof of Lemma 4.6.11. By (4.72), it follows that

$$\partial_\alpha \Phi_\alpha(t) = \partial_\alpha \Re[\overline{m}_t(\bar{\mathfrak{e}}_t^+ + \mathrm{i}N^{-100})] + h^{**}(t,1) - h^{**}(t,0),$$

with $h^{**}(t,\alpha)$ defined by (4.64). Since the two h^{**} terms are small by (4.63), for each fixed t, we have that

$$|\partial_\alpha \Phi_\alpha(t)| \lesssim \left| \partial_\alpha \int_{\mathbf{R}} \frac{\overline{\rho}_t(\bar{\mathfrak{e}}_t^+ + E)}{E - \mathrm{i}N^{-100}} \, \mathrm{d}E \right| + N^{-1} = U_1 + U_2 + N^{-1}, \tag{4.168}$$

where

$$\begin{aligned} U_1 &:= \left| \partial_\alpha \int_{\overline{\gamma}_{-i(\delta_*)}}^{\overline{\gamma}_{i(\delta_*)}} \frac{\overline{\rho}_t(\bar{\mathfrak{e}}_t^+ + E)}{E - \mathrm{i}N^{-100}} \, \mathrm{d}E \right| = \left| \partial_\alpha \int_{I_*} \frac{\overline{\rho}_t(\bar{\mathfrak{e}}_t^+ + \varphi_{\alpha,t}(s))}{\varphi_{\alpha,t}(s) - \mathrm{i}N^{-100}} \varphi_{\alpha,t}'(s) \, \mathrm{d}s \right| \\ U_2 &:= \left| \frac{1}{N} \sum_{i_*(\delta) < |i| \le N} \partial_\alpha \int_{\mathbf{R}} \frac{\psi(E - \overline{\gamma}_i^*(t))}{E - \mathrm{i}N^{-100}} \, \mathrm{d}E \right|, \end{aligned}$$

using the notation $\overline{\gamma}_{i(\delta_*)} = \overline{\gamma}_{i(\delta_*)}(t)$ and the definition of $\overline{\rho}_t$ from (4.56). In U_1 we changed variables, i.e. $E = \varphi_{\alpha,t}(s)$, using that $s \to \varphi_{\alpha,t}(s)$ is strictly increasing. In particular, in order to compute the limits of integration we used that $\varphi_{\alpha,t}(i/N) = \overline{\gamma}_i(t)$ by (4.18) and defined the α-independent interval $I_* := [-i(\delta_*)/N, i(\delta_*)/N]$. Furthermore, in U_1 we denoted by prime the s-derivative.

For U_1 we have that (omitting the t dependence, $\overline{\rho} = \overline{\rho}_t$, etc.)

$$\begin{aligned} U_1 \lesssim & \left| \int_{I_*} \frac{\partial_\alpha [\overline{\rho}(\bar{\mathfrak{e}}^+ + \varphi_\alpha(s))]}{\varphi_\alpha(s) - \mathrm{i}N^{-100}} \varphi_\alpha'(s) \, \mathrm{d}s \right| + \left| \int_{I_*} \frac{\overline{\rho}(\bar{\mathfrak{e}}^+ + \varphi_\alpha(s))}{(\varphi_\alpha(s) - \mathrm{i}N^{-100})^2} (\varphi_\alpha'(s))^2 \, \mathrm{d}s \right| \\ & + \left| \int_{I_*} \frac{\overline{\rho}(\bar{\mathfrak{e}}^+ + \varphi_\alpha(s))}{\varphi_\alpha(s) - \mathrm{i}N^{-100}} \partial_\alpha \varphi_\alpha'(s) \, \mathrm{d}s \right|. \end{aligned} \tag{4.169}$$

For $s \in I_*$, by the definition of $\varphi_\alpha(s)$ and (4.20) it follows that

$$1 = n_\alpha'(\varphi_\alpha(s))\varphi_\alpha'(s) = \rho_\alpha(\varphi_\alpha(s))\varphi_\alpha(s),$$

and so that

$$\varphi_\alpha'(s) = \frac{1}{\rho_\alpha(\varphi_\alpha(s))} \lesssim s^{-\frac{1}{4}}, \tag{4.170}$$

where in the last inequality we used that $\rho_\alpha(\omega) \sim \min\{\omega^{1/3}, \omega^{1/2}\Delta^{-1/6}\}$ and $\varphi_\alpha(s) \sim \max\{s^{\frac{3}{4}}, s^{\frac{2}{3}}\Delta^{1/9}\}$ by (4.25a).

In the first integral in (4.169) we use that

$$\overline{\rho}(\bar{\mathfrak{e}}^+ + \varphi_\alpha(s)) = \rho_\alpha(\bar{\mathfrak{e}}^+ + \varphi_\alpha(s)), \qquad s \in I_*$$

by (4.56) and that $\partial_\alpha[\rho_\alpha(\bar{\mathfrak{e}}^+ + \varphi_\alpha(s))]$ is bounded by the explicit relation in (4.26). For the other two integrals in (4.169) we use that $\overline{\rho}$ is bounded on the integration domain and that $(\varphi_\alpha'(s))^2 \lesssim s^{-1/2}$ from (4.170), hence it is integrable. In the third integral we also observe that

$$\partial_\alpha \varphi_\alpha(s) = \varphi_\lambda(s) - \varphi_\mu(s)$$

by (4.18), thus $|\partial_\alpha \varphi_\alpha'(s)| \lesssim s^{-1/4}$ similarly to (4.170). Using $|\varphi_\alpha(s) - \mathrm{i}N^{-100}| \gtrsim N^{-100}$, we thus conclude that

$$U_1 \lesssim N^{200}.$$

Next, we proceed with the estimate for U_2.

Notice that $|\partial_\alpha \psi(E - \overline{\gamma}_i^*(t))| \le \|\psi'\|_\infty |\widehat{\gamma}_{x,i}(t) - \widehat{\gamma}_{y,i}(t)|$ by (4.54). Furthermore, since $|E - \mathrm{i}N^{-100}| \gtrsim \delta_*$ on the domain of integration of U_2, we conclude that

$$U_2 \lesssim N^{200} \|\psi'\|_\infty,$$

and therefore from (4.168) we have

$$|\partial_\alpha \Phi_\alpha(t)| \lesssim N^{202}. \tag{4.171}$$

since $\|\psi'\|_\infty \lesssim N^2$ by the choice of ψ, see below (4.53).

Similarly, we conclude that

$$|Z_3| \lesssim N^{200} \|\psi'\|_\infty. \tag{4.172}$$

To estimate Z_2, by (4.71), it follows that

$$\mathrm{d}(\partial_\alpha \widetilde{z}_i) = \left[\frac{1}{N} \sum_j \frac{\partial_\alpha \widetilde{z}_j - \partial_\alpha \widetilde{z}_i}{(\widetilde{z}_i - \widetilde{z}_j)^2} + \partial_\alpha \Phi_\alpha(t) \right] \mathrm{d}t,$$

with initial data

$$\partial_\alpha \widetilde{z}_i(0,\alpha) = \widetilde{x}_i(0) - \widetilde{y}_i(0),$$

for all $1 \le |i| \le N$. Since $|\partial_\alpha \widetilde{z}_i(0,\alpha)| \lesssim N^{200}$ for all $1 \le |i| \le N$, by Duhamel principle and contraction, it follows that

$$|\partial_\alpha \widetilde{z}_i(t,\alpha)| \lesssim N^{200} + t_* \max_{0 \le \tau \le t_*} |\partial_\alpha \Phi_\alpha(\tau)| \lesssim N^{202} \tag{4.173}$$

for all $0 \le t \le t_*$. In particular, by (4.173) it follows that

$$|Z_2| \lesssim N^{202} \sqrt{N} \tag{4.174}$$

since for all j in the summation in Z_2 we have that $|i - j| \gtrsim i_* \sim N^{\frac{1}{2}}$ and thus $|\widetilde{z}_i - \widetilde{z}_j| \gtrsim |i - j|/N \gtrsim N^{-1/2}$.

Finally, we estimate Z_4 using the fact that the endpoints of $\mathcal{I}_{z,i}(t)^c \cap \mathcal{J}_z(t)$ are quantiles $\overline{\gamma}_i(t)$ whose α-derivatives are bounded by (4.54). Hence

$$|Z_4| \lesssim \left| \frac{\overline{\rho}_t(\overline{\gamma}_{j_+} + \bar{\mathfrak{e}}_t^+)}{\widehat{z}_i - \overline{\gamma}_{j_+}} \right| + \left| \frac{\overline{\rho}_t(\overline{\gamma}_{j_-} + \bar{\mathfrak{e}}_t^+)}{\widehat{z}_i - \overline{\gamma}_{j_-}} \right| + \left| \frac{\overline{\rho}_t(\overline{\gamma}_{\frac{3i_*}{4}} + \bar{\mathfrak{e}}_t^+)}{\widehat{z}_i - \overline{\gamma}_{\frac{3i_*}{4}}} \right| \lesssim N \tag{4.175}$$

by rigidity. Combining (4.171)-(4.175) we conclude (4.167), completing the proof of Lemma 4.6.11. □

Continuing the analysis of the equation (4.162), for any fixed α let us define $w^\# = w^\#(t,\alpha)$ as the solution of

$$\partial_t w^\# = \mathcal{L} w^\#, \tag{4.176}$$

with cutoff initial data

$$w_i^\#(0,\alpha) = \mathbf{1}_{\{|i|\le N^{4\omega_\ell+\delta}\}} w_i(0,\alpha),$$

with some $0<\delta<C\omega_\ell$ where $C>10$ a constant such that $(4+C)\omega_\ell < \omega_A$.

By the rigidity (4.159) the finite speed estimate (4.279), with $\delta' := \delta$, for the propagator $\mathcal{U}^{\mathcal{L}}$ of $\mathcal{L}$ holds. Let $0<\delta_1<\frac{\delta}{2}$, then, using Duhamel principle, that the error term $\zeta_i^{(0)}$ is bounded by (4.167) and that $\zeta_i^{(0)}=0$ for any $1\le|i|\le N^{\omega_A}$, it easily follows that

$$\sup_{0\le t\le t_*} \max_{|i|\le N^{4\omega_\ell+\delta_1}} |w_i^\#(t,\alpha) - w_i(t,\alpha)| \le N^{-100}, \tag{4.177}$$

for any $\alpha\in[0,1]$. In other words, the initial conditions far away do not influence the w-dynamics, hence they can be set zero.

Next, we use the heat kernel contraction for the equation in (4.176). By the optimal rigidity of $\widehat{x}_i(0)$ and $\widehat{y}_i(0)$, since $w_i^\#(0,\alpha)$ is non zero only for $1\le|i|\le N^{4\omega_\ell+\delta}$, it follows that

$$\max_{1\le|i|\le N} |w_i^\#(0,\alpha)| \le \frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}}}, \tag{4.178}$$

and so, by heat kernel contraction and Duhamel principle

$$\sup_{0\le t\le t_*} \max_{1\le|i|\le N} |w_i^\#(t,\alpha)| \le \frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}}}. \tag{4.179}$$

Next, we recall that $\widehat{z}(t,\alpha=0)=\widehat{y}(t)$.

Combining (4.177) and (4.179), integrating $w_i(t,\alpha')$ over $\alpha'\in[0,\alpha]$, by high moment Markov inequality as in (4.102)-(4.103), we conclude that

$$\sup_{0\le t\le t_*} |\widehat{z}_i(t,\alpha) - \widehat{y}_i(t)| \le \frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}}}, \qquad 1\le|i|\le N^{4\omega_\ell+\delta_1},$$

for any fixed $\alpha\in[0,1]$ with very high probability for any $\xi>0$. Since

$$|\widehat{z}_i(t,\alpha) - \overline{\gamma}_i(t)| \le |\widehat{y}_i(t) - \widehat{\gamma}_{y,i}(t)| + |\overline{\gamma}_i(t) - \widehat{\gamma}_{y,i}(t)| + \frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}}},$$

for all $1\le|i|\le N^{4\omega_\ell+\delta_1}$ and $\alpha\in[0,1]$, by (4.35) and the optimal rigidity of $\widehat{y}_i(t)$, see (4.60), we conclude that

$$\sup_{0\le t\le t_*} |\widehat{z}_i(t,\alpha) - \overline{\gamma}_i(t)| \le \frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}}}, \qquad 1\le|i|\le N^{4\omega_\ell+\delta_1} \tag{4.180}$$

for any fixed $\alpha\in[0,1]$, for any $\xi>0$ with very high probability. This concludes the proof of (4.160). □

4.6.5 Phase 3: Rigidity for $\widehat{z}$ with the correct i-dependence.

In this subsection we will prove almost optimal i-dependent rigidity for the short range approximation $\widehat{z}_i(t,\alpha)$ (see (4.154)–(4.157)) for $1 \le |i| \le N^{4\omega_\ell+\delta_1}$.

Proposition 4.6.12. *Let δ_1 be defined in Proposition 4.6.10, then, for any fixed $\alpha \in [0,1]$, we have that*

$$\sup_{0\le t\le t_*} |\widehat{z}_i(t,\alpha) - \overline{\gamma}_i(t)| \lesssim \frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}}|i|^{\frac{1}{4}}}, \qquad 1 \le |i| \le N^{4\omega_\ell+\delta_1}, \tag{4.181}$$

for any $\xi > 0$ with very high probability.

Proof. Define

$$K := \lceil N^\xi \rceil,$$

then (4.160) (with $\xi \to \xi/2$) implies (4.181) for all $1 \le |i| \le 2K$. Next, we prove (4.181) for all $2K \le |i| \le N^{4\omega_\ell+\delta_1}$ by coupling $\widetilde{x}_i(t)$ with $\widetilde{y}_{\langle i-K\rangle}(t)$, where we make the following notational convention:

$$\langle i-K\rangle := i-K \qquad \text{if } i \in [K+1,N]\cup[-N,-1], \qquad \langle i-K\rangle := i-K-1 \qquad \text{if } i \in [1,K]. \tag{4.182}$$

This slight complication is due to our indexing convention that excludes $i = 0$.

In order to couple the Brownian motion of $\widetilde{x}_i(t)$ with the one of $\widetilde{y}_{\langle i-K\rangle}(t)$ we construct a new process $\widetilde{z}^*(t,\alpha)$ satisfying

$$\mathrm{d}\widetilde{z}_i^*(t,\alpha) = \sqrt{\frac{2}{N}}\,\mathrm{d}B_{\langle i-K\rangle} + \left[\frac{1}{N}\sum_{j\ne i}\frac{1}{\widetilde{z}_i^*(t,\alpha) - \widetilde{z}_j^*(t,\alpha)} + \Phi_\alpha(t)\right]\mathrm{d}t, \quad 1 \le |i| \le N \tag{4.183}$$

with initial data

$$\widetilde{z}_i^*(0,\alpha) = \alpha\widetilde{x}_i(0) + (1-\alpha)\widetilde{y}_{\langle i-K\rangle}(0), \tag{4.184}$$

for any $\alpha \in [0,1]$. Notice that the only difference with respect to $\widetilde{z}_i(t,\alpha)$ from (4.71) is a shift in the index of the Brownian motion, i.e. $\widetilde{z}$ and $\widetilde{z}^*$ (almost) coincide in distribution, but their coupling to the y-process is different. The slight discrepancy comes from the effect of the few extreme indices. Indeed, to make the definition (4.183) unambiguous even for extreme indices, $i \in [-N, -N+K-1]$, additionally we need to define independent Brownian motions B_j and initial padding particles $\widetilde{y}_j(0) = -jN^{300}$ for $j = -N-1, \ldots -N-K$. Similarly to Lemma 4.5.1, the effect of these very distant additional particles is negligible on the dynamics of the particles for $1 \le |i| \le \epsilon N$ for some small ϵ.

Next, we define the process $\widehat{z}^*(t,\alpha)$ as the short range approximation of $\widetilde{z}^*(t,\alpha)$, given by (4.154)–(4.156) but B_i replaced with $B_{\langle i-K\rangle}$ and we use initial data $\widehat{z}^*(0,\alpha) = \widetilde{z}^*(0,\alpha)$. In particular,

$$\widehat{z}_i^*(t,1) = \widehat{x}_i(t) + O(N^{-100}), \qquad \widehat{z}_i^*(t,0) = \widehat{y}_{\langle i-K\rangle}(t) + O(N^{-100}), \qquad 1 \le |i| \le \epsilon N, \tag{4.185}$$

the discrepancy again coming from the negligible effect of the additional K distant particles on the particles near the cusp regime.

Let $w_i^*(t,\alpha) := \partial_\alpha \widehat{z}_i^*(t,\alpha)$, i.e. $w^* = w^*(t,\alpha)$ is a solution of

$$\partial_t w^* = \mathcal{B}w^* + \mathcal{V}w^* + \zeta^{(0)}$$

with initial data

$$w_i^*(0,\alpha) = \widehat{x}_i^*(0) - \widehat{y}_{\langle i-K\rangle}(0).$$

The operators $\mathcal{B}$, $\mathcal{L}$ and the error term $\zeta^{(0)}$ are defined as in (4.163)-(4.166) with all $\widetilde{z}$ and $\widehat{z}$ replaced by $\widetilde{z}^*$ and $\widehat{z}^*$, respectively.

We now define $(w^*)^{\#}$ as the solution of

$$\partial_t (w^*)^{\#} = \mathcal{L}(w^*)^{\#}, \tag{4.186}$$

with cutoff initial data

$$(w_i^*)^{\#}(0,\alpha) = \mathbf{1}_{\{|i|\le N^{4\omega_\ell+\delta}\}} w_i^*(0,\alpha),$$

with $0<\delta<C\omega_\ell$ with $C>10$ such that $(4+C)\omega_\ell < \omega_A$.

We claim that

$$(w_i^*)^{\#}(0,\alpha) \ge 0, \qquad 1\le |i| \le N. \tag{4.187}$$

We need to check it for $1\le|i|\le N^{4\omega_\ell+\delta}$, otherwise $(w_i^*)^{\#}(0,\alpha)=0$ by the cutoff. In the regime $1\le|i|\le N^{4\omega_\ell+\delta}$ we use the optimal rigidity (Lemma 4.6.1 with $\xi\to\xi/10$) for $\widehat{x}_i^*(0)$ and $\widehat{y}_{\langle i-K\rangle}(0)$ that yields

$$\begin{aligned}(w_i^*)^{\#}(0,\alpha) = \widehat{x}_i^*(0) - \widehat{y}_{\langle i-K\rangle}(0) \ge -N^{\frac{\xi}{10}}\eta_f(\gamma_{x,i}^*(0)) + \widehat{\gamma}_{x,i}(0) - \widehat{\gamma}_{y,\langle i-K\rangle}(0) \\ - N^{\frac{\xi}{10}}\eta_f(\gamma^*_{y,\langle i-K\rangle}(0)).\end{aligned} \tag{4.188}$$

We now check that $\widehat{\gamma}_{x,i}(0)-\widehat{\gamma}_{y,\langle i-K\rangle}(0)$ is sufficiently positive to compensate for the $N^{\frac{\xi}{10}}\eta_f$ error terms. Indeed, by (4.30a) and (4.35), for all $|i|\ge 2K$ we have

$$\widehat{\gamma}_{x,i}(t) - \widehat{\gamma}_{y,\langle i-K\rangle}(t) \gtrsim K\eta_f(\gamma^*_{x,i}(t)) \gg N^{\frac{\xi}{10}}\eta_f(\gamma^*_{x,i}(t))$$

and that

$$\eta_f(\gamma^*_{y,\langle i-K\rangle}(t)) \sim \eta_f(\gamma^*_{x,i}(t)).$$

This shows (4.187) in the $2K\le|i|\le N^{4\omega_\ell+\delta}$ regime. If $K\le|i|\le 2K$ or $-K\le i\le -1$ we have that $(w_i^*)^{\#}(0,\alpha)\ge 0$ since

$$\begin{aligned}\widehat{\gamma}_{x,i}(0) - \widehat{\gamma}_{y,\langle i-K\rangle}(0) &\gtrsim \max\Big\{\frac{K^{3/4}}{N^{3/4}}, (t_*-t)^{1/6}\frac{K^{2/3}}{N^{2/3}}\Big\} \\ &\gtrsim K\max\big\{\eta_f(\gamma^*_{x,i}(0)), \eta_f(\gamma^*_{y,\langle i-K\rangle}(0))\big\},\end{aligned}$$

so $\widehat{\gamma}_{x,i}(0)-\widehat{\gamma}_{y,\langle i-K\rangle}(0)$ beats the error terms $N^{\frac{\xi}{10}}\eta_f$ as well. Finally, if $1\le i\le K-1$ the bound in (4.188) is easy since $\widehat{\gamma}_{x,i}(0)$ and $\widehat{\gamma}_{y,\langle i-K\rangle}(0)$ have opposite sign, i.e. they are in two different sides of the small gap and one of them is at least of order $(K/N)^{3/4}$, beating $N^{\frac{\xi}{10}}\eta_f$. This proves (4.187). Hence, by the maximum principle we conclude that

$$(w_i^*)^{\#}(t,\alpha)\ge 0, \qquad 0\le t\le t_*, \qquad \alpha\in[0,1]. \tag{4.189}$$

Let $\delta_1<\frac{\delta}{2}$ be defined in Proposition 4.6.10. The rigidity estimate in (4.159) holds for $\widehat{z}^*$ as well, since $\widehat{z}$ and $\widehat{z}^*$ have the same distribution. Furthermore, by (4.159) the propagator

$\mathcal{U}$ of $\mathcal{L} := \mathcal{B} + \mathcal{V}$ satisfies the finite speed estimate in Lemma 4.B.3. Then, using Duhamel principle and (4.167), we obtain

$$\sup_{0 \le t \le t_*} \max_{1 \le |i| \le N^{4\omega_\ell + \delta_1}} |(w_i^*)^\#(t, \alpha) - w_i^*(t, \alpha)| \le N^{-100}, \tag{4.190}$$

for any $\alpha \in [0, 1]$ with very high probability.

By (4.190), integrating $w_i^*(t, \alpha')$ over $\alpha' \in [0, \alpha]$, we conclude that

$$\widehat{z}_i^*(t, \alpha) - \widehat{y}_{\langle i-K \rangle}(t) \ge -N^{-100}, \qquad 1 \le |i| \le N,^{4\omega_\ell + \delta_1} \tag{4.191}$$

for all $\alpha \in [0, 1]$ and $0 \le t \le t_*$ with very high probability. Note that in order to prove (4.191) with very high probability we used a Markov inequality as in (4.102)-(4.103). Hence,

$$\begin{aligned} \widehat{z}_i^*(t, \alpha) - \overline{\gamma}_i(t) \ge & [\widehat{y}_{\langle i-K \rangle}(t) - \widehat{\gamma}_{y, \langle i-K \rangle}(t)] + [\widehat{\gamma}_{y, \langle i-K \rangle}(t) - \widehat{\gamma}_{y,i}(t)] \\ & + [\widehat{\gamma}_{y,i}(t) - \overline{\gamma}_i(t)] - N^{-100} \\ \gtrsim & - K(\eta_f(\gamma_{y, \langle i-K \rangle}^*(t)) + \eta_f(\gamma_{y,i}^*(t))) - \gamma_i^*(t) t_*^{1/3} \\ \ge & -2K(\eta_f(\gamma_{y, \langle i-K \rangle}^*(t)) + \eta_f(\gamma_{y,i}^*(t))) \end{aligned} \tag{4.192}$$

for all $1 \le |i| \le N^{4\omega_\ell + \delta_1}$, where we used the optimal rigidity (4.60) and (4.35) in going to the second line. In particular, since for $|i| \ge 2K$ we have that $\eta_f(\gamma_{y,i}^*(t)) \sim \eta_f(\gamma_{y,i-K}^*(t))$, we conclude that

$$\widehat{z}_i^*(t, \alpha) - \overline{\gamma}_i(t) \ge -\frac{CKN^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}} |i|^{\frac{1}{4}}}, \qquad 2K \le |i| \le N^{4\omega_\ell + \delta_1}, \tag{4.193}$$

for all $0 \le t \le t_*$ and for any $\alpha \in [0, 1]$. This implies the lower bound in (4.181).

In order to prove the upper bound in (4.181) we consider a very similar process $\widetilde{z}_i^*(t, \alpha)$ (we continue to denote it by star) where the index shift in y is in the other direction. More precisely, it is defined as a solution of

$$\mathrm{d}\widetilde{z}_i^*(t, \alpha) = \sqrt{\frac{2}{N}} \mathrm{d}B_{\langle i+K \rangle} + \left[\frac{1}{N} \sum_{j \ne i} \frac{1}{\widetilde{z}_i^*(t, \alpha) - \widetilde{z}_j^*(t, \alpha)} + \Phi_\alpha(t) \right] \mathrm{d}t$$

with initial data

$$\widetilde{z}_i(0, \alpha) = \alpha \widetilde{y}_{\langle i+K \rangle}(0) + (1 - \alpha) \widetilde{x}_i(0),$$

for any $\alpha \in [0, 1]$. Here $\langle i + K \rangle$ is defined analogously to (4.182). Then, by similar computations, we conclude that

$$\widehat{z}_i^*(t, \alpha) - \overline{\gamma}_i(t) \le \frac{KN^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}} |i|^{\frac{1}{4}}}, \qquad 2K \le |i| \le N^{4\omega_\ell + \delta_1}, \tag{4.194}$$

for all $0 \le t \le t_*$ and for any $\alpha \in [0, 1]$. Combining (4.193) and (4.194) we conclude (4.181) and complete the proof of Proposition 4.6.12. □

4.7 Proof of Proposition 4.3.1: Dyson Brownian motion near the cusp

In this section $t_1 \leq t_*$, indicating that we are before the cusp formation, we recall that t_1 is defined as follows

$$t_1 := \frac{N^{\omega_1}}{N^{1/2}},$$

for a small fixed $\omega_1 > 0$ and t_* is the time of the formation of the exact cusp. The main result of this section is the following proposition from which we can quickly prove Proposition 4.3.1 for $t_1 \leq t_*$. If $t_1 > t_*$ we conclude Proposition 4.3.1 using the analogous Proposition 4.8.1 instead of Proposition 4.7.1 exactly in the same way.

Proposition 4.7.1. *For $t_1 \leq t_*$, with very high probability we have that*

$$|(\lambda_j(t_1) - \mathfrak{e}^+_{\lambda,t_1}) - (\mu_{j+i_\mu - i_\lambda}(t_1) - \mathfrak{e}^+_{\mu,t_1})| \leq N^{-\frac{3}{4}-c\omega_1} \tag{4.195}$$

for some small constant $c > 0$ and for any j such that $|j - i_\lambda| \leq N^{\omega_1}$.

Note that if $t_1 = t_*$ then $\mathfrak{e}^+_{r,t_*} = \mathfrak{e}^-_{r,t_*} = \mathfrak{c}_r$, for $r = \lambda, \mu$, with $\mathfrak{c}_r$ being the exact cusp point of the scDOSs ρ_{r,t_*}. The proof of Proposition 4.7.1 will be given at the end of the section after several auxiliary lemmas.

Proof of Proposition 4.3.1. Firstly, we recall the definition of the physical cusp

$$\mathfrak{b}_{r,t_1} := \begin{cases} \frac{1}{2}(\mathfrak{e}^+_{r,t_1} + \mathfrak{e}^-_{r,t_1}) & \text{if} \quad t_1 < t_*, \\ \mathfrak{c}_r & \text{if} \quad t_1 = t_*, \\ \mathfrak{m}_{r,t_1} & \text{if} \quad t_1 > t_*. \end{cases}$$

of ρ_{r,t_1} as in (4.5), for $r = \lambda, \mu$. Then, using the change of variables $\boldsymbol{x} = N^{\frac{3}{4}}(\boldsymbol{x}' - \mathfrak{b}_{r,t_1})$, for $r = \lambda, \mu$, and the definition of correlation function, for each Lipschitz continuous and compactly supported test function F, we have that

$$\begin{aligned} &\int_{\mathbf{R}^k} F(\boldsymbol{x}) \left[N^{k/4} p^{(N,\lambda)}_{k,t_1}\left(\mathfrak{b}_{\lambda,t_1} + \frac{\boldsymbol{x}}{N^{3/4}}\right) - N^{k/4} p^{(N,\mu)}_{k,t_1}\left(\mathfrak{b}_{\mu,t_1} + \frac{\boldsymbol{x}}{N^{3/4}}\right)\right] \mathrm{d}\boldsymbol{x} \\ &\quad = N^k \binom{N}{k}^{-1} \sum_{\{i_1,\dots,i_k\}\subset[N]} \Bigg[\mathbf{E}_{H^{(\lambda)}_{t_1}} F\left(N^{\frac{3}{4}}(\lambda_{i_1} - \mathfrak{b}_{\lambda,t_1}), \dots, N^{\frac{3}{4}}(\lambda_{i_k} - \mathfrak{b}_{\lambda,t_1})\right) \\ &\qquad\qquad - \mathbf{E}_{H^{(\mu)}_{t_1}} F(\lambda \to \mu)\Bigg], \end{aligned} \tag{4.196}$$

where $\lambda_1, \dots, \lambda_N$ and $\mu_1, \dots, \mu_N$ are the eigenvalues, labelled in increasing order, of $H^{(\lambda)}_{t_1}$ and $H^{(\mu)}_{t_1}$ respectively. In $\mathbf{E}_{H^{(\mu)}_{t_1}} F(\lambda \to \mu)$ we also replace $\mathfrak{b}_{\lambda,t_1}$ by $\mathfrak{b}_{\mu,t_1}$.

In order to apply Proposition 4.7.1 we split the sum in the rhs. of (4.196) into two sums:

$$\sum_{\substack{\{i_1,\dots,i_k\}\subset[N] \\ |i_1 - i_\lambda|,\dots,|i_k - i_\lambda| < N^\epsilon}} \quad \text{and its complement} \quad \sum{}', \tag{4.197}$$

where ϵ is a positive exponent with $\epsilon \ll \omega_1$.

We start with the estimate for the second sum of (4.197). In particular, we will estimate only the term $\mathbf{E}_{H_{t_1}^{(\lambda)}}(\cdot)$, the estimate for $\mathbf{E}_{H_{t_1}^{(\mu)}}(\cdot)$ will follow in an analogous way.

Since the test function F is compactly supported in some set $\Omega \subset \mathbf{R}^k$ and in $\sum'$ there is at least one index i_l such that $|i_l - i_\lambda| \geq N^\epsilon$, we have that

$$\begin{aligned}\sum{}' \mathbf{E}_{H_{t_1}^{(\lambda)}} F\left(N^{\frac{3}{4}}(\lambda_{i_1} - \mathfrak{b}_{\lambda,t_1}), \dots, N^{\frac{3}{4}}(\lambda_{i_k} - \mathfrak{b}_{\lambda,t_1})\right) \\ \lesssim N^{k-1}\|F\|_\infty \sum_{i_l\colon |i_l - i_\lambda| \geq N^\epsilon} \mathbf{P}_{H_{t_1}^{(\lambda)}}\left(|\lambda_{i_l} - \mathfrak{b}_{\lambda,t_1}| \lesssim C_\Omega N^{-\frac{3}{4}}\right).\end{aligned} \tag{4.198}$$

where C_Ω is the diameter of Ω. Let $\gamma_{\lambda,i} = \widehat{\gamma}_{\lambda,i} + \mathfrak{e}^+_{\lambda,t_1}$ be the classical eigenvalue locations of $\rho_\lambda(t_1)$ defined by (4.27) for all $1 - i_\lambda \leq i \leq N + 1 - i_\lambda$. Then, by the rigidity estimate from [83, Corollary 2.6], we have that

$$\mathbf{P}_{H_{t_1}^{(\lambda)}}\left(|\lambda_{i_l} - \mathfrak{b}_{\lambda,t_1}| \lesssim C_\Omega N^{-\frac{3}{4}}, |i_l - i_\lambda| \geq N^\epsilon\right) \leq N^{-D}, \tag{4.199}$$

for each $D > 0$ if N is large enough, depending on C_Ω. Indeed, by rigidity it follows that

$$|\lambda_{i_l} - \mathfrak{b}_{\lambda,t_1}| \geq |\gamma_{\lambda,i_l} - \gamma_{\lambda,i_\lambda}| - |\lambda_{i_l} - \gamma_{\lambda,i_l}| - |\mathfrak{b}_{\lambda,t_1} - \gamma_{\lambda,i_\lambda}| \gtrsim \frac{N^{c\epsilon}}{N^{\frac{3}{4}}} - \frac{N^{c\xi}}{N^{\frac{3}{4}}} \gtrsim \frac{N^{c\epsilon}}{N^{\frac{3}{4}}} \tag{4.200}$$

with very high probability, if $N^\epsilon \leq |i_l - i_\lambda| \leq \tilde{c}N$, for some $0 < \tilde{c} < 1$. In (4.200) we used the rigidity from [83, Corollary 2.6] in the form

$$|\lambda_i - \gamma_{\lambda,i}| \leq \frac{N^\xi}{N^{\frac{3}{4}}},$$

for any $\xi > 0$, with very high probability. Note that (4.199) and (4.200) hold for any $\epsilon \gtrsim \xi$. If $|i_l - i_\lambda| \geq \tilde{c}N$, then $|\gamma_{i_l} - \gamma_{i_\lambda}| \sim 1$ and the bound in (4.200) clearly holds. A similar estimate holds for $H_{t_1}^{(\mu)}$, hence, choosing $D > k + 1$ we conclude that the second sum in (4.197) is negligible.

Next, we consider the first sum in (4.197). For $t_1 \leq t_*$ we have, by (4.22a) that

$$|(\mathfrak{e}^+_{\lambda,t_1} - \mathfrak{b}_{\lambda,t_1}) - (\mathfrak{e}^+_{\mu,t_1} - \mathfrak{b}_{\mu,t_1})| = \frac{1}{2}|\Delta_{\lambda,t_1} - \Delta_{\mu,t_1}| \lesssim \Delta_{\mu,t_1}(t_* - t_1)^{1/3} \leq N^{-\frac{3}{4} - \frac{1}{6} + C\omega_1}.$$

Hence, by (4.195), using that $|F(\boldsymbol{x}) - F(\boldsymbol{x}')| \lesssim \|F\|_{C^1}\|\boldsymbol{x} - \boldsymbol{x}'\|$, we conclude that

$$\begin{aligned}\sum_{\substack{\{i_1,\dots,i_k\}\subset[N] \\ |i_1 - i_\lambda|,\dots,|i_k - i_\lambda| \leq N^\epsilon}} \left[\mathbf{E}_{H_{t_1}^{(\lambda)}} F\left(N^{\frac{3}{4}}(\lambda_{i_1} - \mathfrak{b}_{\lambda,t_1}), \dots, N^{\frac{3}{4}}(\lambda_{i_k} - \mathfrak{b}_{\lambda,t_1})\right) - \mathbf{E}_{H_{t_1}^{(\mu)}} F(\lambda \to \mu)\right] \\ \leq C_k \|F\|_{C^1} \frac{N^{k\epsilon}}{N^{c\omega_1}},\end{aligned} \tag{4.201}$$

for some $c > 0$. Then, using that

$$\frac{N^k(N-k)!}{N!} = 1 + \mathcal{O}_k(N^{-1}),$$

we easily conclude the proof of Proposition 4.3.1. □

4.7.1 Interpolation.

In order to prove Proposition 4.7.1 we recall a few concepts introduced previously. In Section 4.5 we introduced the padding particles $x_i(t)$, $y_i(t)$, for $1 \leq |i| \leq N$, that are good approximations of the eigenvalues $\lambda_j(t)$, $\mu_j(t)$ respectively, for $1 \leq j \leq N$, in the sense of Lemma 4.5.1. They satisfy a Dyson Brownian Motion equation (4.49), (4.51) mimicking the DBM of genuine eigenvalue processes (4.47), (4.48). It is more convenient to consider shifted processes where the edge motion is subtracted.

More precisely, for $r = x, y$ and $r(t) = x(t), y(t)$, we defined

$$\widetilde{r}_i(t) := r_i(t) - \mathfrak{e}^+_{r,t}, \qquad 1 \leq |i| \leq N,$$

for all $0 \leq t \leq t_*$. In particular, $\widetilde{r}(t)$ is a solution of

$$\mathrm{d}\widetilde{r}_i(t) = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \left(\frac{1}{N}\sum_{j\neq i}\frac{1}{\widetilde{r}_i(t) - \widetilde{r}_j(t)} + \Re[m_{r,t}(\mathfrak{e}^+_{r,t})]\right)\mathrm{d}t, \tag{4.202}$$

with initial data

$$\widetilde{r}_i(0) = r_i(0) - \mathfrak{e}^+_{r,0}, \tag{4.203}$$

for all $1 \leq |i| \leq N$.

Next, following a similar idea of [131], we also introduced in (4.71) an interpolation process between $\widetilde{x}(t)$ and $\widetilde{y}(t)$. For any $\alpha \in [0,1]$ we defined the process $\widetilde{z}(t,\alpha)$ as the solution of

$$\mathrm{d}\widetilde{z}_i(t,\alpha) = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \left(\frac{1}{N}\sum_{j\neq i}\frac{1}{\widetilde{z}_i(t,\alpha) - \widetilde{z}_j(t,\alpha)} + \Phi_\alpha(t)\right)\mathrm{d}t, \tag{4.204}$$

with initial data

$$\widetilde{z}_i(0,\alpha) = \alpha\widetilde{x}_i(0) + (1-\alpha)\widetilde{y}_i(0),$$

for each $1 \leq |i| \leq N$. Recall that $\Phi_\alpha(t)$ was defined in (4.72) and it is such that $\Phi_0(t) = \Re[m_{y,t}(\mathfrak{e}^+_{y,t})]$ and $\Phi_1(t) = \Re[m_{x,t}(\mathfrak{e}^+_{x,t})]$. Note that $\widetilde{z}_i(t,1) = \widetilde{x}_i(t)$ and $\widetilde{z}_i(t,0) = \widetilde{y}_i(t)$ for all $1 \leq |i| \leq N$ and $0 \leq t \leq t_*$.

We recall the definition of the interpolated quantiles from (4.54) of Section 4.5;

$$\overline{\gamma}_i(t) := \alpha\widehat{\gamma}_{x,i}(t) + (1-\alpha)\widehat{\gamma}_{y,i}(t), \qquad \alpha \in [0,1], \tag{4.205}$$

where $\widehat{\gamma}_{x,i}$ and $\widehat{\gamma}_{y,i}$ are the shifted quantiles of $\rho_{x,t}$ and $\rho_{y,t}$ respectively, defined in Section 4.5. In particular,

$$\overline{\mathfrak{e}}^\pm_t = \alpha\mathfrak{e}^\pm_{x,t} + (1-\alpha)\mathfrak{e}^\pm_{y,t}, \qquad \alpha \in [0,1].$$

We denoted the interpolated density, whose quantiles are the $\overline{\gamma}_i(t)$, by $\overline{\rho}_t$ (4.56), and its Stieltjes transform by $\overline{m}_t$.

Let $\widehat{z}(t,\alpha)$ be the short range approximation of $\widetilde{z}(t,\alpha)$ defined by (4.154)-(4.156), with exponents $\omega_1 \ll \omega_\ell \ll \omega_A \ll 1$ and with initial data $\widehat{z}(0,\alpha) = \widetilde{z}(0,\alpha)$ and $i_* = N^{\frac{1}{2}+C_*\omega_1}$, for some large constant $C_* > 0$. In particular, $\widehat{x}(t) = \widehat{z}(t,1)$ and $\widehat{y}(t) = \widehat{z}(t,0)$. Assuming optimal rigidity in (4.60) for $\widetilde{r}_i(t) = \widetilde{x}_i(t), \widetilde{y}_i(t)$, the following lemma shows that the process $\widetilde{r}$ and its short range approximation $\widehat{r} = \widehat{x}, \widehat{y}$ stay very close to each other, i.e. $|\widehat{r}_i - \widetilde{r}_i| \leq N^{-\frac{3}{4}-c}$, for some small $c > 0$. This is the analogue of Lemma 3.7 in [131] and its proof, given in Appendix 4.C, follows similar lines. It assumes the optimal rigidity, see (4.206) below, which is ensured by [83, Corollary 2.6], see Lemma 4.6.1.

Lemma 4.7.2. *Let* $i_* = N^{\frac{1}{2}+C_*\omega_1}$. *Assume that* $\widetilde{z}(t,0)$ *and* $\widetilde{z}(t,1)$ *satisfy the optimal rigidity*

$$\sup_{0\le t\le t_1} |\widetilde{z}_i(t,\alpha) - \widehat{\gamma}_{r,i}(t)| \le N^{\xi}\eta_{\mathrm{f}}^{\rho_{r,t}}(e^+_{r,t} + \widehat{\gamma}_{r,\pm i}(t)), \qquad 1\le |i| \le i_*, \quad \alpha = 0,1, \tag{4.206}$$

with $r = x, y$, *for any* $\xi > 0$, *with very high probability. Then, for* $\alpha = 0$ *or* $\alpha = 1$ *we have that*

$$\begin{aligned}\sup_{1\le |i|\le N} \sup_{0\le t\le t_1} &|\widetilde{z}_i(t,\alpha) - \widehat{z}_i(t,\alpha)| \\ &\lesssim \frac{N^{\frac{\omega_1}{6}}N^{\xi}}{N^{\frac{3}{4}}}\left(\frac{N^{\omega_1}}{N^{3\omega_\ell}} + \frac{N^{\omega_1}}{N^{\frac{1}{8}}} + \frac{N^{C\omega_1}N^{\frac{\omega_A}{2}}}{N^{\frac{1}{6}}} + \frac{N^{\frac{\omega_A}{2}}N^{C\omega_1}}{N^{\frac{1}{4}}} + \frac{N^{C\omega_1}}{N^{\frac{1}{18}}}\right),\end{aligned} \tag{4.207}$$

for any $\xi > 0$, *with very high probability.*

In particular, (4.207) implies that there exists a small fixed universal constant $c > 0$ such that

$$\sup_{1\le |i|\le N} \sup_{0\le t\le t_1} |\widetilde{z}_i(t,\alpha) - \widehat{z}_i(t,\alpha)| \lesssim N^{-\frac{3}{4}-c}, \qquad \alpha = 0,1 \tag{4.208}$$

with very high probability.

Remark 4.7.3. *Note the denominator in the first error term in* (4.207)*: the factor* $N^{3\omega_\ell}$ *is better than* $N^{2\omega_\ell}$ *in Lemma 3.7 in [131], this is because of the natural cusp scaling. The fact that this power is at least* $N^{(1+\epsilon)\omega_\ell}$ *was essential in [131] since this allowed to transfer the optimal rigidity from* $\widetilde{z}$ *to the* $\widehat{z}$ *process for all* $\alpha \in [0,1]$. *Optimal rigidity for* $\widehat{z}$ *is essential (i) for the heat kernel bound for the propagator of* $\mathcal{L}$, *see* (4.162)–(4.163), *and (ii) for a good* ℓ^p*-norm for the initial condition in* (4.219). *With our approach, however, this power in* (4.207) *is not critical since we have already obtained an even better,* i*-dependent rigidity for the* $\widehat{z}$ *process for any* α *by using maximum principle, see Proposition 4.6.12. We still need* (4.207) *for the* x *and* y *processes (i.e. only for* $\alpha = 0,1$*), but only with a precision below the rigidity scale, therefore the denominator in the first term has only to beat* $N^{\frac{7}{6}\omega_1+\xi}$.

4.7.2 Differentiation.

Next, we consider the α-derivative of the process $\widehat{z}(t,\alpha)$. Let

$$u_i(t) = u_i(t,\alpha) := \partial_\alpha \widehat{z}_i(t,\alpha), \qquad 1\le |i| \le N,$$

then u is a solution of the equation

$$\partial_t u = \mathcal{L}u + \zeta^{(0)}, \tag{4.209}$$

where $\zeta^{(0)}$, defined by (4.165)-(4.166), is an error term that is non zero only for $|i| > N^{\omega_A}$ and such that $|\zeta_i^{(0)}| \lesssim N^C$, for some large constant $C > 0$ with very high probability, by (4.167), and the operator $\mathcal{L} = \mathcal{B} + \mathcal{V}$ acting on $\mathbf{R}^{2N}$ is defined by (4.163)-(4.164).

In the following with $\mathcal{U}^{\mathcal{L}}$ we denote the semigroup associated to (4.209), i.e. by Duhamel principle

$$u(t) = \mathcal{U}^{\mathcal{L}}(0,t)u(0) + \int_0^t \mathcal{U}^{\mathcal{L}}(s,t)\zeta^{(0)}(s)\,\mathrm{d}s$$

and $\mathcal{U}^{\mathcal{L}}(s,s) = \mathrm{Id}$ for all $0 \leq s \leq t$. Furthermore, for each a, b such that $|a|, |b| \leq N$, with $\mathcal{U}^{\mathcal{L}}_{ab}$ we denote the entries of $\mathcal{U}^{\mathcal{L}}$, which can be either seen as the solution of the equation (4.209) with initial condition $u_a(0) = \delta_{ab}$.

By Proposition 4.6.3 and Lemma 4.C.1, for any fixed $\alpha \in [0,1]$, it follows that

$$\sup_{0 \leq t \leq t_*} |\widehat{z}_i(t,\alpha) - \overline{\gamma}_i(t)| \lesssim \frac{N^{C\omega_1}}{N^{\frac{1}{2}}}, \qquad 1 \leq |i| \leq N, \tag{4.210}$$

and

$$\sup_{0 \leq t \leq t_*} |\widehat{z}_i(t,\alpha) - \overline{\gamma}_i(t)| \lesssim \frac{N^{C\omega_1}}{N^{\frac{3}{4}}}, \qquad 1 \leq |i| \leq i_*, \tag{4.211}$$

with very high probability. Then, using (4.211), as a consequence of Lemma 4.B.3 we have the following:

Lemma 4.7.4. *There exists a constant $C > 0$ such that for any $0 < \delta < C\omega_\ell$, if $1 \leq |a| \leq \frac{1}{2}N^{4\omega_\ell+\delta}$ and $|b| \geq N^{4\omega_\ell+\delta}$, then*

$$\sup_{0 \leq s \leq t \leq t_*} \mathcal{U}^{\mathcal{L}}_{ab}(s,t) + \mathcal{U}^{\mathcal{L}}_{ba}(s,t) \leq N^{-D} \tag{4.212}$$

for any $D > 0$ with very high probability.

Furthermore, by Proposition 4.6.12, for any fixed $\alpha \in [0,1]$, we have that

$$\sup_{0 \leq t \leq t_*} |\widehat{z}_i(t,\alpha) - \overline{\gamma}_i(t)| \lesssim \frac{N^{\xi} N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}} |i|^{\frac{1}{4}}}, \qquad 1 \leq |i| \leq N^{4\omega_\ell+\delta_1}, \tag{4.213}$$

for some small fixed $\delta_1 > 0$ and for any $\xi > 0$ with very high probability.

Next, we introduce the ℓ^p norms

$$\|u\|_p := \left(\sum_i |u_i|^p \right)^{\frac{1}{p}}, \quad \|u\|_\infty := \max_i |u_i|.$$

Following a similar scheme to [41], [91] with some minor modifications we will prove the following Sobolev-type inequalities in Appendix 4.D.

Lemma 4.7.5. *For any small $\eta > 0$ there exists $c_\eta > 0$ such that*

$$\sum_{i \neq j \in \mathbf{Z}_+} \frac{(u_i - u_j)^2}{|i^{\frac{3}{4}} - j^{\frac{3}{4}}|^{2-\eta}} \geq c_\eta \left(\sum_{i \geq 1} |u_i|^p \right)^{\frac{2}{p}}, \quad \sum_{i \neq j \in \mathbf{Z}_-} \frac{(u_i - u_j)^2}{||i|^{\frac{3}{4}} - |j|^{\frac{3}{4}}|^{2-\eta}} \geq c_\eta \left(\sum_{i \leq -1} |u_i|^p \right)^{\frac{2}{p}} \tag{4.214}$$

hold, with $p = \frac{8}{2+3\eta}$, for any function $\|u\|_p < \infty$.

Using the Sobolev inequality in (4.214) and the finite speed estimate of Lemma 4.7.4, in Appendix 4.E we prove the energy estimates for the heat kernel in Lemma 4.7.6 via a Nash-type argument.

Lemma 4.7.6. *Assume* (4.210), (4.211) *and* (4.213). *Let* $0 < \delta_4 < \frac{\delta_1}{10}$ *and* $w_0 \in \mathbf{R}^{2N}$ *such that* $|(w_0)_i| \le N^{-100}\|w_0\|_1$, *for* $|i| \ge \ell^4 N^{\delta_4}$. *Then, for any small* $\eta > 0$ *there exists a constant* $C > 0$ *independent of* η *and a constant* c_η *such that for all* $0 \le s \le t \le t_*$

$$\|\mathcal{U}^{\mathcal{L}}(s,t)w_0\|_2 \le \left(\frac{N^{C\eta + \frac{\omega_1}{3}}}{c_\eta N^{\frac{1}{2}}(t-s)}\right)^{1-3\eta} \|w_0\|_1, \tag{4.215}$$

and

$$\|\mathcal{U}^{\mathcal{L}}(0,t)w_0\|_\infty \le \left(\frac{N^{C\eta + \frac{\omega_1}{3}}}{c_\eta N^{\frac{1}{2}} t}\right)^{\frac{2(1-3\eta)}{p}} \|w_0\|_p, \tag{4.216}$$

for each $p \ge 1$.

Let $0 < \delta_v < \frac{\delta_4}{2}$. Define $v_i = v_i(t,\alpha)$ to be the solution of

$$\partial_t v = \mathcal{L} v, \qquad v_i(0,\alpha) = u_i(0,\alpha)\mathbf{1}_{\{|i| \le N^{4\omega_\ell + \delta_v}\}}. \tag{4.217}$$

Then, by Lemma 4.7.4 the next lemma follows.

Lemma 4.7.7. *Let* u *be the solution of the equation in* (4.209) *and* v *defined by* (4.217), *then we have that*

$$\sup_{0 \le t \le t_1} \sup_{|i| \le \ell^4} |u_i(t) - v_i(t)| \le N^{-100}, \tag{4.218}$$

with very high probability.

Proof. By (4.209) and (4.217) have that

$$u_i(t) - v_i(t) = \sum_{j=-N}^{N} \mathcal{U}^{\mathcal{L}}_{ij}(0,t)u_j(0) - \sum_{j=-N^{4\omega_\ell+\delta_v}}^{N^{4\omega_\ell+\delta_v}} \mathcal{U}^{\mathcal{L}}_{ij}(0,t)u_j(0) + \int_0^t \sum_{|j| \ge N^{\omega_A}} \mathcal{U}^{\mathcal{L}}_{ij}(s,t)\zeta_j^{(0)}(s)\, \mathrm{d}s.$$

Then, using that $\zeta_i^{(0)} = 0$ for $1 \le |i| \le N^{\omega_A}$ and (4.167), the bound in (4.218) follows by Lemma 4.7.4. □

Proof of Proposition 4.7.1. We consider only the $j = i_\lambda$ case. By Lemma 4.5.1 and (4.208) we have that

$$\begin{aligned} |(\lambda_{i_\lambda}(t_1) - \mathfrak{e}^+_{\lambda,t_1}) - (\mu_{i_\mu}(t_1) - \mathfrak{e}^+_{\mu,t_1})| &\le |\widetilde{x}_1(t_1) - \widehat{x}_1(t_1)| + |\widehat{x}_1(t_1) - \widehat{y}_1(t_1)| + |\widehat{y}_1(t_1) - \widetilde{y}_1(t_1)| \\ &\le |\widehat{x}_1(t_1) - \widehat{y}_1(t_1)| + N^{-\frac{3}{4}-c} \end{aligned}$$

with very high probability.

Since $\widehat{z}_i(t_1,1) = \widehat{x}_i(t_1)$ and $\widehat{z}_i(t_1,0) = \widehat{y}_i(t_1)$ for all $1 \le |i| \le N$, by the definition of $u_i(t,\alpha)$, it follows that

$$\widehat{x}_1(t_1) - \widehat{y}_1(t_1) = \int_0^1 u_1(t_1,\alpha)\, \mathrm{d}\alpha.$$

Furthermore, by a high moment Markov inequality as in (4.102)-(4.103) and Lemma 4.7.7, we get

$$\int_0^1 |u_1(t_1,\alpha)|\, \mathrm{d}\alpha \lesssim N^{-100} + \int_0^1 |v_1(t_1,\alpha)|\, \mathrm{d}\alpha.$$

Since $v_i(0) = u_i(0)\mathbf{1}_{\{|i|\le N^{4\omega_\ell+\delta_v}\}}$ and, by (4.35) and (4.60), for $1 \le |i| \le N^{4\omega_\ell+\delta_v}$ we have that

$$\begin{aligned} |u_i(0)| &\lesssim |\widehat{x}_i(0) - \widehat{\gamma}_{x,i}(0)| + |\widehat{y}_i(0) - \widehat{\gamma}_{y,i}(0)| + |\widehat{\gamma}_{x,i}(0) - \widehat{\gamma}_{y,i}(0)| \\ &\lesssim \frac{N^{\frac{\omega_1}{6}}}{|i|^{\frac{1}{4}}N^{\frac{3}{4}}} + \frac{|i|^{\frac{3}{4}}N^{\frac{\omega_1}{2}}}{N^{\frac{11}{12}}} \lesssim \frac{N^{\frac{\omega_1}{6}}}{|i|^{\frac{1}{4}}N^{\frac{3}{4}}}, \end{aligned}$$

we conclude that

$$\|v(0)\|_5 \lesssim \frac{N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}}} \tag{4.219}$$

with very high probability. Hence, reaclling that $t_1 = N^{-1/2+\omega_1}$, by (4.216) and Markov's inequality again, we get

$$\begin{aligned} \int_0^1 |v_1(t_1,\alpha)| \,\mathrm{d}\alpha \le \sup_{\alpha\in[0,1]} \|v(t_1,\alpha)\|_\infty &\le \left(\frac{N^{C\eta+\omega_1/3}}{N^{1/2}t_1}\right)^{\frac{2(1-3\eta)}{5}} \|v(0)\|_5 \\ &\lesssim \frac{N^{\frac{\omega_1}{6}+\frac{\eta}{5}(2C+3\omega_1-6\eta C)}}{N^{\frac{3}{4}}N^{\frac{4\omega_1}{15}}} = \frac{1}{N^{\frac{3}{4}}N^{\frac{\omega_1}{20}}}, \end{aligned} \tag{4.220}$$

with very high probability, for η small enough, say $\eta \le \omega_1(8C + 12\omega_1)^{-1}$. Notice that the constant in front of the ω_1 in the exponents play a crucial role: eventually the constant $(1 - \frac{1}{3})\frac{2}{5} = \frac{4}{15}$ from the Nash estimate beats the constant $\frac{1}{6}$ from (4.219). This completes the proof of Proposition 4.7.1. □

4.8 Case of $t \ge t_*$: small minimum

In this section we consider the case when the densities $\rho_{x,t}$, $\rho_{y,t}$, hence their interpolation $\overline{\rho}_t$ as well, have a small minimum, i.e. $t_* \le t \le 2t_*$. We deal with the small minimum case in this separate section mainly for notational reasons: for $t_* \le t \le 2t_*$ the processes $x(t)$ and $y(t)$, and consequently the associated quantiles and densities, are shifted by $\widetilde{\mathfrak{m}}_{r,t}$, for $r = x, y$, instead of $\mathfrak{e}^+_{r,t}$. We recall that $\widetilde{\mathfrak{m}}_{r,t}$, defined in (4.31a), denotes a close approximation of the actual local minimum $\mathfrak{m}_{r,t}$ near the physical cusp. We chose to shift $x(t)$ and $y(t)$ by the tilde approximation of the minimum instead of the minimum itself for technical reasons, namely because the t-derivative of $\widetilde{\mathfrak{m}}_{r,t}$, $r = x, y$, satisfies the convenient relation in (4.31d).

As we explained at the beginning of Section 4.7, in order to prove universality, i.e. Proposition 4.3.1 at time $t_1 \ge t_*$, it is enough to prove the following:

Proposition 4.8.1. *For $t_1 \ge t_*$, we have, with very high probability, that*

$$|(\lambda_j(t_1) - \mathfrak{m}_{\lambda,t_1}) - (\mu_{j+i_\mu-i_\lambda}(t_1) - \mathfrak{m}_{\mu,t_1})| \le N^{-\frac{3}{4}-c} \tag{4.221}$$

for some small constant $c > 0$ and for any j such that $|j - i_\lambda| \le N^{\omega_1}$. Here $\mathfrak{m}_{\lambda,t_1}$ and $\mathfrak{m}_{\mu,t_1}$ are the local minimum of ρ_{λ,t_1} and ρ_{μ,t_1}, respectively.

We introduce the shifted process $\widetilde{r}_i(t) = \widetilde{x}_i(t), \widetilde{y}_i(t)$ for $t \ge t_*$. Let us define

$$\widetilde{r}_i(t) := r_i(t) - \widetilde{\mathfrak{m}}_{r,t}, \qquad 1 \le |i| \le N, \tag{4.222}$$

for $r = x, y$, hence, by (4.31d), the shifted points satisfy the following DBM

$$\mathrm{d}\widetilde{r}_i(t) = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \frac{1}{N}\sum_{j\neq i}\frac{1}{\widetilde{r}_i(t)-\widetilde{r}_j(t)}\,\mathrm{d}t - \Big(\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\mathfrak{m}}_{r,t}\Big)\,\mathrm{d}t. \tag{4.223}$$

Furthermore we recall that by $\widehat{\gamma}_{r,i}(t)$ we the denote the quantiles of $\rho_{r,t}$, with $r = x, y$, for all $t_* \leq t \leq 2t_*$, i.e.

$$\widehat{\gamma}_{r,i} = \gamma_{r,i} - \widetilde{\mathfrak{m}}_{r,t}, \qquad 1 \leq |i| \leq N.$$

By the rigidity estimate of [83, Corollary 2.6], using Lemma 4.5.1 and the fluctuation scale estimate in (4.34a) the proof of the following lemma is immediate.

Lemma 4.8.2. *Let $\widetilde{r}(t) = \widetilde{x}(t), \widetilde{y}(t)$. There exists a fixed small $\epsilon > 0$, such that for each $1 \leq |i| \leq \epsilon N$, we have*

$$\sup_{t_* \leq t \leq t_1} |\widetilde{r}_i(t) - \widehat{\gamma}_{r,i}(t)| \leq N^{\xi}\eta_{\mathrm{f}}^{\rho_{r,t}}(\gamma_{r,i}(t)), \tag{4.224}$$

for any $\xi > 0$ with very high probability, where we recall that the behavior of $\eta_{\mathrm{f}}^{\rho_{r,t}}(\mathfrak{e}_{r,t}^{+} + \widehat{\gamma}_{r,\pm i}(t))$, with $r = x, y$, is given by (4.34b).

In order to prove Proposition 4.8.1, by Lemma 4.5.1 and (4.31b), it is enough to prove the following proposition.

Proposition 4.8.3. *For $t_1 \geq t_*$ we have, with very high probability, that*

$$|(x_i(t_1) - \widetilde{\mathfrak{m}}_{x,t_1}) - (y_i(t_1) - \widetilde{\mathfrak{m}}_{y,t_1})| \leq N^{-\frac{3}{4}-c} \tag{4.225}$$

for some small constant $c > 0$ and for any $1 \leq |i| \leq N^{\omega_1}$.

The remaining part of this section is devoted to the proof of Proposition 4.8.3. We start with some preparatory lemmas. We recall the definition of the interpolated quantiles given in Section 4.5,

$$\overline{\gamma}_i(t) := \alpha\widehat{\gamma}_{x,i}(t) + (1-\alpha)\widehat{\gamma}_{y,i}(t), \tag{4.226}$$

for all $\alpha \in [0,1]$ and $t_* \leq t \leq 2t_*$, as well as

$$\overline{\mathfrak{m}}_t := \alpha\widetilde{\mathfrak{m}}_{x,t} + (1-\alpha)\widetilde{\mathfrak{m}}_{y,t},$$

for all $\alpha \in [0,1]$ and $t_* \leq t \leq 2t_*$. Furthermore by $\overline{\rho}_t$ from (4.56) we denote the interpolated density between $\rho_{x,t}$ and $\rho_{y,t}$ and by $\overline{m}_t$ its Stieltjes transform.

We now define the process $\widetilde{z}_i(t,\alpha)$ whose initial data are given by the linear interpolation of $\widetilde{x}(0)$ and $\widetilde{y}(0)$. Analogously to the small gap case, we define the function $\Psi_\alpha(t)$, for $t_* \leq t \leq 2t_*$, that represents the correct shift of the process $\widetilde{z}(t,\alpha)$, in order to compensate the discrepancy of our choice of the interpolation for $\overline{\rho}_t$ with respect to the semicircular flow evolution of the density $\overline{\rho}_0$.

Analogously to the edge case, see (4.62)-(4.68), we define $h(t,\alpha)$ with the following properties

$$h(t,\alpha) = \alpha\Re[m_{x,t}(\widetilde{\mathfrak{m}}_{x,t})] + (1-\alpha)\Re[m_{y,t}(\widetilde{\mathfrak{m}}_{y,t})] - \Re[\overline{m}_t(\overline{\mathfrak{m}}_t + \mathrm{i}N^{-100})] + \mathcal{O}\left(N^{-1}\right) \tag{4.227}$$

and $h(t,0)=h(t,1)=0$. Then, similarly to the edge case, we define

$$\Psi_\alpha(t) := -\alpha \frac{\mathrm{d}}{\mathrm{d}t}[m_{x,t}(\widetilde{\mathfrak{m}}_{x,t})] - (1-\alpha)\frac{\mathrm{d}}{\mathrm{d}t}[m_{y,t}(\widetilde{\mathfrak{m}}_{y,t})] - h(t,\alpha). \tag{4.228}$$

In particular, by our definition of $h(t,\alpha)$ in (4.227) it follows that $\Psi_0(t) = \frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\mathfrak{m}}_{y,t}$, $\Psi_1(t) = \frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\mathfrak{m}}_{x,t}$ and that

$$\Psi_\alpha(t) = \Re[\overline{m}_t(\overline{\mathfrak{m}}_t)] + \mathcal{O}(N^{-\frac{1}{2}+\omega_1}). \tag{4.229}$$

Note that the error in (4.229) is somewhat weaker than in the analogous equation (4.73) due to the additional error in (4.31d) compared with (4.31e).

More precisely, the process $\widetilde{z}(t,\alpha)$ is defined by

$$\mathrm{d}\widetilde{z}_i(t,\alpha) = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \left[\frac{1}{N}\sum_{j\neq i}\frac{1}{\widetilde{z}_i(t,\alpha)-\widetilde{z}_j(t,\alpha)} + \Psi_\alpha(t)\right]\mathrm{d}t, \tag{4.230}$$

with initial data

$$\widetilde{z}_i(t_*,\alpha) := \alpha\widetilde{x}_i(t_*) + (1-\alpha)\widetilde{y}_i(t_*), \tag{4.231}$$

for all $1\le |i|\le N$ and for all $\alpha\in[0,1]$.

We recall that $\omega_1 \ll \omega_\ell \ll \omega_A \ll 1$ and that $i_* = N^{\frac{1}{2}+C_*\omega_1}$ with some large constant C_*.

Next, we define the analogue of $\mathcal{J}_z(t)$ and $\mathcal{I}_{z,i}(t)$ for the small minimum by (4.152) and (4.153) using the definition in (4.226) for the quantiles. Then, for each $t_*\le t\le t_1$, we define the short range approximation $\widehat{z}_i(t,\alpha)$ of $\widetilde{z}(t,\alpha)$ by the following SDE.

For $|i| > \frac{i_*}{2}$ we let

$$\mathrm{d}\widehat{z}_i(t,\alpha) = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \left[\frac{1}{N}\sum_j^{\mathcal{A},(i)}\frac{1}{\widehat{z}_i(t,\alpha)-\widehat{z}_j(t,\alpha)} + \frac{1}{N}\sum_j^{\mathcal{A}^c,(i)}\frac{1}{\widetilde{z}_i(t,\alpha)-\widetilde{z}_j(t,\alpha)} + \Psi_\alpha(t)\right]\mathrm{d}t, \tag{4.232}$$

for $|i|\le N^{\omega_A}$

$$\begin{aligned}\mathrm{d}\widehat{z}_i(t,\alpha) = {}& \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \left[\frac{1}{N}\sum_j^{\mathcal{A},(i)}\frac{1}{\widehat{z}_i(t,\alpha)-\widehat{z}_j(t,\alpha)} + \int_{\mathcal{I}_{y,i}(t)^c}\frac{\rho_{y,t}(E+\widetilde{\mathfrak{m}}_{y,t}^+)}{\widehat{z}_i(t,\alpha)-E}\,\mathrm{d}E\right]\mathrm{d}t \\ & - \Big(\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\mathfrak{m}}_{r,t}\Big)\,\mathrm{d}t,\end{aligned} \tag{4.233}$$

and for $N^{\omega_A} < |i| \le \frac{i_*}{2}$

$$\begin{aligned}\mathrm{d}\widehat{z}_i(t,\alpha) = {}& \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \left[\frac{1}{N}\sum_j^{\mathcal{A},(i)}\frac{1}{\widehat{z}_i(t,\alpha)-\widehat{z}_j(t,\alpha)} + \int_{\mathcal{I}_{z,i}(t)^c\cap\mathcal{J}_z(t)}\frac{\overline{\rho}_t(E+\overline{\mathfrak{m}}_t^+)}{\widehat{z}_i(t,\alpha)-E}\,\mathrm{d}E\right. \\ & \left. + \sum_{|j|\ge\frac{3}{4}i_*}\frac{1}{\widetilde{z}_i(t,\alpha)-\widetilde{z}_j(t,\alpha)} + \Psi_\alpha(t)\right]\mathrm{d}t,\end{aligned} \tag{4.234}$$

with initial data

$$\widehat{z}_i(t_*, \alpha) := \widetilde{z}_i(t_*, \alpha). \tag{4.235}$$

Next, by Lemma 4.C.2, by the optimal rigidity in (4.224) for $\widetilde{x}(t)$ and $\widetilde{y}(t)$, the next lemma follows immediately.

Lemma 4.8.4. *For $\alpha = 0$ and $\alpha = 1$, with very high probability, we have*

$$\sup_{1 \leq |i| \leq N} \sup_{t_* \leq t \leq t_1} |\widetilde{z}_i(t, \alpha) - \widehat{z}_i(t, \alpha)| \lesssim \frac{N^\xi}{N^{\frac{3}{4}}} \left(\frac{N^{\omega_1}}{N^{3\omega_\ell}} + \frac{N^{C\omega_1}}{N^{\frac{1}{24}}} \right), \tag{4.236}$$

for any $\xi > 0$ and $C > 0$ a large universal constant.

In order to proceed with the heat-kernel estimates we need an optimal i-dependent rigidity for $\widehat{z}_i(t, \alpha)$ for $1 \leq |i| \leq N^{4\omega_\ell + \delta}$, for some $0 < \delta < C\omega_\ell$. In particular, analogously to Proposition 4.6.12 we have:

Proposition 4.8.5. *Fix any $\alpha \in [0, 1]$. There exists a small fixed $0 < \delta_1 < C\omega_\ell$, for some constant $C > 0$, such that*

$$\sup_{t_* \leq t \leq 2t_*} |\widehat{z}_i(t, \alpha) - \overline{\gamma}_i(t)| \lesssim \frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}} |i|^{\frac{1}{4}}}, \qquad 1 \leq |i| \leq N^{4\omega_\ell + \delta_1} \tag{4.237}$$

for any $\xi > 0$ with very high probability.

Proof. We can adapt the arguments in Section 4.6 to the case of the small minimum, $t \geq t_*$, in a straightforward way. In Section 4.6, as the main input, we used the precise estimates on the density $\rho_{r,t}$ (4.22b), (4.37), on the quantiles $\widehat{\gamma}_{r,i}(t)$ (4.30a), on the quantile gaps (4.35), on the fluctuation scale (4.34a) and on the Stieltjes transform (4.39a); all formulated for the small gap case, $0 \leq t \leq t_*$. In the small minimum case, $t \geq t_*$, the corresponding estimates are all available in Section 4.4, see (4.22d),(4.38),(4.30b), (4.36),(4.34b) and (4.39b), respectively. In fact, the semicircular flow is more regular after the cusp formation, see e.g. the better (larger) exponent in the $(t - t_*)$ error terms when comparing (4.22b) with (4.22d). This makes handling the small minimum case easier. The most critical part in Section 4.6 is the estimate of the forcing term (Proposition 4.6.7), where the derivative of the density (4.23a) was heavily used. The main mechanism of this proof is the delicate cancellation between the contributions to S_2 from the intervals $[\gamma_{i-n-1}, \gamma_{i-n}]$ and $[\gamma_{i+n-1}, \gamma_{i+n}]$, see (4.140). This cancellation takes place away from the edge. The proof is divided into two cases; the so-called "edge regime", where the gap length Δ is relatively large and the "cusp regime", where Δ is small or zero. The adaptation of this argument to the small minimum case, $t \geq t_*$, will be identical to the proof for the small gap case in the "cusp regime". In this regime the derivative bound (4.23a) is used only in the form $|\rho'| \lesssim \rho^{-2}$ which is available in the small minimum case, $t \geq t_*$, as well, see (4.24a). This proves Proposition 4.6.7 for $t \geq t_*$. The rest of the argument is identical to the proof in the small minimum case up to obvious notational changes; the details are left to the reader. □

Let us define $u_i(t, \alpha) := \partial_\alpha \widehat{z}_i(t, \alpha)$, for $t_* \leq t \leq 2t_*$. In particular, u is a solution of the equation

$$\partial_t u = \mathcal{L} u + \zeta^{(0)} \tag{4.238}$$

with initial condition $u(t_*, \alpha) = \widetilde{x}(t_*) - \widetilde{y}(t_*)$ from (4.231). The error term $\zeta^{(0)}$ is defined analogously to (4.165)-(4.166) but replacing Φ_α and $\overline{\mathfrak{e}}_t^+$ with Ψ_α and $\widetilde{\mathfrak{m}}_t$, respectively. Note that this error term is non zero only for $|i| \geq N^{\omega_A}$ and for any i we have $|\zeta_i^{(0)}| \leq N^C$ with very high probability, for some large $C > 0$. Furthermore, $\mathcal{L} = \mathcal{B} + \mathcal{V}$ is defined as in (4.163)-(4.164) replacing $\mathfrak{e}_{y,t}^+$ and $\overline{\mathfrak{e}}_t^+$ by $\widetilde{\mathfrak{m}}_{y,t}$ and $\overline{\mathfrak{m}}_t$, respectively. In the following by $\mathcal{U}^{\mathcal{L}}$ we denote the propagator of the operator $\mathcal{L}$.

Let $0 < \delta_v < \frac{\delta_4}{2}$, with δ_4 defined in Lemma 4.7.6. Define $v_i = v_i(t, \alpha)$ to be the solution of

$$\partial_t v = \mathcal{L} v, \qquad v_i(t_*, \alpha) = u_i(t_*, \alpha) \mathbf{1}_{\{|i| \leq N^{4\omega_\ell + \delta_v}\}}. \tag{4.239}$$

By the finite speed of propagation estimate in Lemma 4.B.3, similarly to the proof of Lemma 4.7.7, we immediately obtain the following:

Lemma 4.8.6. *Let u be the solution of the equation in (4.238) and v defined by (4.239), then we have that*

$$\sup_{t_* \leq t \leq 2t_*} \sup_{1 \leq |i| \leq \ell^4} |u_i(t) - v_i(t)| \leq N^{-100} \tag{4.240}$$

with very high probability.

Collecting all the previous lemmas we conclude this section with the proof of Proposition 4.8.3.

Proof of Proposition 4.8.3. We consider only the $i = 1$ case. By Lemma 4.5.1 and Lemma 4.8.4 we have that

$$\begin{aligned} |(x_1(t_1) - \widetilde{\mathfrak{m}}_{x,t_1}) - (y_1(t_1) - \widetilde{\mathfrak{m}}_{y,t_1})| &\leq |\widetilde{x}_1(t_1) - \widehat{x}_1(t_1)| + |\widehat{x}_1(t_1) - \widehat{y}_1(t_1)| \\ &\quad + |\widehat{y}_1(t_1) - \widetilde{y}_1(t_1)| \\ &\leq |\widehat{x}_1(t_1) - \widehat{y}_1(t_1)| + \frac{1}{N^{\frac{3}{4}+c}} \end{aligned}$$

with very high probability. Since $u(t, \alpha) = \partial_\alpha \widehat{z}(t, \alpha)$, using $\widehat{x}_1(t_1) - \widehat{y}_1(t_1) = \int_0^1 u(t_1, \alpha)\, \mathrm{d}\alpha$ and Lemma 4.8.6 it will be sufficient to estimate $\int_0^1 |v_1(t_1, \alpha)|\, \mathrm{d}\alpha$. By rigidity from (4.224), we have

$$|v_i(t_*, \alpha)| = |u_i(t_*, \alpha)| = |\widetilde{y}_i(t_*) - \widetilde{x}_i(t_*)| \lesssim \frac{N^\xi}{N^{\frac{3}{4}} |i|^{\frac{1}{4}}},$$

for any $1 \leq |i| \leq N^{4\omega_\ell + \delta_v}$ hence

$$\|v(t_*, \alpha)\|_5 \lesssim \frac{N^\xi}{N^{\frac{3}{4}}},$$

for any $\xi > 0$ with very high probability.

Finally, using the heat kernel estimate in (4.216) for $\mathcal{U}^{\mathcal{L}}(0, t)$ for $t_* \leq t \leq 2t_*$, we conclude, after a Markov inequality as in (4.102)-(4.103),

$$\int_0^1 |v_1(t_1, \alpha)|\, \mathrm{d}\alpha \lesssim \frac{N^\xi}{N^{\frac{3}{4}} N^{\frac{4\omega_1}{15}}}, \tag{4.241}$$

with very high probability. □

4.A Proof of Theorem 4.2.4

We now briefly outline the changes required for the proof of Theorem 4.2.4 compared to the proof of Theorem 4.2.2. We first note that for $0 \le \tau_1 \le \cdots \le \tau_k \lesssim N^{-1/2}$ in distribution $(H^{(\tau_1)}, \ldots, H^{(\tau_k)})$ agrees with

$$\Big(H+\sqrt{\tau_1}U_1, H+\sqrt{\tau_1}U_1+\sqrt{\tau_2-\tau_1}U_2, \ldots, H+\sqrt{\tau_1}U_1+\cdots+\sqrt{\tau_k-\tau_{k-1}}U_k\Big), \quad (4.242)$$

where $U_1, \ldots, U_k$ are independent GOE matrices. Next, we claim and prove later by Green function comparison that the time-dependent k-point correlation function of (4.242) asymptotically agrees with the one of

$$\Big(\widetilde{H}_t+\sqrt{\tau_1}U_1, \widetilde{H}_t+\sqrt{\tau_1}U_1+\sqrt{\tau_2-\tau_1}U_2, \ldots, \widetilde{H}_t+\sqrt{\tau_1}U_1+\cdots+\sqrt{\tau_k-\tau_{k-1}}U_k\Big), \quad (4.243)$$

and thereby also with the one of

$$\Big(H_t+\sqrt{ct}U+\sqrt{\tau_1}U_1, \ldots, H_t+\sqrt{ct}U+\sqrt{\tau_1}U_1+\cdots+\sqrt{\tau_k-\tau_{k-1}}U_k\Big), \quad (4.244)$$

for any fixed $t \le N^{-1/4-\epsilon}$, where we recall that $\widetilde{H}_t$ and H_t constructed as in Section 4.3 (see (4.10)). Finally, we notice that the joint eigenvalue distribution of the matrices in (4.244) is precisely given by the joint distribution of

$$\Big(\lambda_i(ct+\tau_1), \ldots, \lambda_i(ct+\tau_k),\ i \in [N]\Big)$$

of eigenvalues λ_i^s evolved according to the DBM

$$\mathrm{d}\lambda_i(s) = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \sum_{j\ne i}\frac{1}{\lambda_i(s)-\lambda_j(s)}\,\mathrm{d}s, \quad \lambda_i(0)=\lambda_i(H_t). \quad (4.245)$$

The high probability control on the eigenvalues evolved according to (4.245) in Propositions 4.7.1 and 4.8.1 allows to simultaneously compare eigenvalues at different times with those of the Gaussian reference ensemble automatically.

In order to establish Theorem 4.2.4 it thus only remains to argue that the k-point functions of (4.242) and (4.243) are asymptotically equal. For the sake of this argument we consider only the randomness in H and condition on the randomness in $U_1, \ldots, U_k$. Then the OU-flow

$$\mathrm{d}\widetilde{H}_s' = -\frac{1}{2}(\widetilde{H}_s' - A - \sqrt{\tau_1}U_1 - \cdots - \sqrt{\tau_l-\tau_{l-1}}U_l)\,\mathrm{d}s + \Sigma^{1/2}[\mathrm{d}\mathfrak{B}_s]$$

with initial conditions

$$\widetilde{H}_0' = H + \sqrt{\tau_1}U_1 + \cdots + \sqrt{\tau_l-\tau_{l-1}}U_l$$

for fixed $U_1, \ldots, U_l$ is given by

$$\widetilde{H}_s' = \widetilde{H}_s + \sqrt{\tau_1}U_1 + \cdots + \sqrt{\tau_l-\tau_{l-1}}U_l,$$

i.e. we view $\sqrt{\tau_1}U_1 + \cdots + \sqrt{\tau_l-\tau_{l-1}}U_l$ as an additional expectation matrix. Thus we can appeal to the standard Green function comparison technique already used in Section 4.3

to compare the k-point functions of (4.242) and (4.243). Here we can follow the standard resolvent expansion argument from [83, Eq. (116)] and note that the proof therein verbatim also allows us to compare products of traces of resolvents with differing expectations. Finally we then take the $\mathbf{E}_{U_1}\dots\mathbf{E}_{U_k}$ expectation to conclude that not only the conditioned k-point functions of (4.242) and (4.243) asymptotically agree, but also the k-point functions themselves.

4.B Finite speed of propagation estimate

In this section we prove a finite speed of propagation estimate for the time evolution of the α-derivative of the short range dynamics defined in (4.154)–(4.156). It is an adjustment to the analogous proof of Lemma 4.1 in [131]. For concreteness, we present the proof for the propagator $\mathcal{U}^{\mathcal{L}}$ where $\mathcal{L} = \mathcal{B} + \mathcal{V}$ is defined in (4.162)–(4.164). The point is that once the dynamics is localized, i.e. the range of the interaction term $\mathcal{B}$ is restricted to a local scale $|i-j| \le |j_+(i) - j_-(i)|$, with $|j_+(i) - j_-(i)| \gtrsim N^{4\omega_\ell} =: L$, and the time is also restricted, $0 \le t \le 2t_* \lesssim N^{-\frac{1}{2}+\omega_1}$, then the propagation cannot go beyond a scale that is much bigger than the interaction scale. This mechanism is very general and will also be used in a slightly different (simpler) setup of Lemma 4.6.5 and Proposition 4.6.8 where the interaction scale is much bigger $L \sim \sqrt{N}$. We will give the necessary changes for the proof of Lemma 4.6.5 and Proposition 4.6.8 at the end of this section.

Lemma 4.B.1. *Let $\widehat{z}(t) = \widehat{z}(t,\alpha)$ be the solution to the short range dynamics* (4.154)–(4.156) *with $i_* = N^{\frac{1}{2}+C_*\omega_1}$, exponents $\omega_1 \ll \omega_\ell \ll \omega_A \ll 1$ and propagator $\mathcal{L} = \mathcal{B}+\mathcal{V}$ from* (4.162)–(4.164). *Let us assume that*

$$\sup_{0\le t\le t_*} |\widehat{z}_i(t) - \overline{\gamma}_i(t)| \le \frac{N^{C\omega_1}}{N^{\frac{3}{4}}}, \qquad 1 \le |i| \le i_*, \tag{4.246}$$

where $\overline{\gamma}_i(t)$ are the quantiles from (4.54). *Then, there exists a constant $C' > 0$ such that for any $0 < \delta < C'\omega_\ell$, $|a| \ge LN^\delta$ and $|b| \le \frac{3}{4}LN^\delta$, for any fixed $0 \le s \le t_*$, we have that*

$$\sup_{s\le t\le t_*} \mathcal{U}^{\mathcal{L}}_{ab}(s,t) + \mathcal{U}^{\mathcal{L}}_{ba}(s,t) \le N^{-D} \tag{4.247}$$

for any $D > 0$, with very high probability. The same result holds for the short range dynamics after the cusp defined in (4.238) *for $t_* \le s \le 2t_*$.*

Proof of Lemma 4.B.1. For concreteness we assume that $0 \le s \le t \le t_*$, i.e. we are in the small gap regime. For $t_* \le s \le t \le 2t_*$ the proof is analogous using the definition (4.226) for the $\overline{\gamma}_i(t)$, the definition of the short range approximation in (4.232)-(4.235) for the $\widehat{z}_i(t,\alpha)$ and replacing $\overline{\mathfrak{e}}_t^+$ by $\overline{\mathfrak{m}}_t$. With these adjustments the proof follows in the same way except for (4.270) below, where we have to use the estimates in (4.39b) instead of (4.39a).

First we consider the $s = 0$ case, then in Lemma 4.B.3 below we extend the proof for all $0 \le s \le t$. Let $\psi(x)$ be an even 1-Lipschitz real function, i.e. $\psi(x) = \psi(-x)$, $\|\psi'\|_\infty \le 1$ such that

$$\psi(x) = |x| \quad \text{for} \quad |x| \le \frac{L^{\frac{3}{4}}N^{\frac{3}{4}\delta}}{N^{\frac{3}{4}}}, \qquad \psi'(x) = 0 \quad \text{for} \quad |x| \ge 2\frac{L^{\frac{3}{4}}N^{\frac{3}{4}\delta}}{N^{\frac{3}{4}}}. \tag{4.248}$$

and

$$\|\psi''\|_\infty \lesssim \frac{N^{\frac{3}{4}}}{L^{\frac{3}{4}}N^{\frac{3\delta}{4}}}. \tag{4.249}$$

We consider a solution of the equation

$$\partial_t f = \mathcal{L}f, \qquad 0 \le t \le t_*$$

with some discrete Dirac delta initial condition $f_i(0) = \delta_{ip_*}$ at p_* for any $|p_*| \ge N^{4\omega_\ell}N^\delta$. For concreteness, assume $p_* > 0$ and set $p := N^{4\omega_\ell}N^\delta$. Define

$$\phi_i = \phi_i(t,\alpha) := e^{\frac{1}{2}\nu\psi(\widehat{z}_i(t,\alpha)-\overline{\gamma}_p(t))}, \quad m_i = m_i(t,\alpha) := f_i(t,\alpha)\phi_i(t,\alpha), \quad \nu = \frac{N^{\frac{3}{4}}}{L^{\frac{3}{4}}N^{\delta'}} \tag{4.250}$$

with some $\delta' \ge \frac{\delta}{2}$ to be chosen later. Let $\widehat{z}_i = \widehat{z}_i(t,\alpha)$ and set

$$F(t) := \sum_i f_i^2 e^{\nu\psi(\widehat{z}_i - \overline{\gamma}_p(t))} = \sum_i m_i^2. \tag{4.251}$$

Since

$$2\sum_i f_i(\mathcal{B}f)_i\phi_i^2 = \sum_{(i,j)\in\mathcal{A}} \mathcal{B}_{ij}(m_i - m_j)^2 - \sum_{(i,j)\in\mathcal{A}} \mathcal{B}_{ij}m_im_j\left(\frac{\phi_i}{\phi_j} + \frac{\phi_j}{\phi_i} - 2\right),$$

using Ito's formula, we conclude that

$$\mathrm{d}F = \sum_{(i,j)\in\mathcal{A}} \mathcal{B}_{ij}(m_i - m_j)^2\,\mathrm{d}t + 2\sum_i \mathcal{V}_i m_i^2\,\mathrm{d}t \tag{4.252}$$

$$- \sum_{(i,j)\in\mathcal{A}} \mathcal{B}_{ij}m_im_j\left(\frac{\phi_i}{\phi_j} + \frac{\phi_j}{\phi_i} - 2\right)\mathrm{d}t \tag{4.253}$$

$$+ \sum_i \nu m_i^2\psi'(\widehat{z}_i - \overline{\gamma}_p)\,\mathrm{d}(\widehat{z}_i - \overline{\gamma}_p) \tag{4.254}$$

$$+ \sum_i m_i^2\left(\frac{\nu^2}{N}\psi'(\widehat{z}_i - \overline{\gamma}_p)^2 + \frac{\nu}{N}\psi''(\widehat{z}_i - \overline{\gamma}_p)\right)\mathrm{d}t. \tag{4.255}$$

Let $\tau_1 \le t_*$ be the first time such that $F \ge 5$ and let τ_2 be stopping time so that the estimate (4.246) holds with $t \le \tau_2$ instead of $t \le t_*$; the condition (4.246) then says that $\tau_2 = t_*$ with very high probability. Define $\tau := \tau_1 \wedge \tau_2 \wedge t_*$, our goal is to show that $\tau = t_*$. In the following we assume $t \le \tau$.

Now we estimate the terms in (4.252)–(4.255) one by one. We start with (4.253). Note that the rigidity scale $N^{-\frac{3}{4}+C\omega_1}$ in (4.246) is much smaller than $N^{-\frac{3}{4}(1-\delta)+3\omega_\ell}$, the range of the support of ψ', which, in turn, is comparable with $|\overline{\gamma}_i - \overline{\gamma}_p| \gtrsim (p/N)^{3/4}$ for any $i \ge 2p = 2LN^\delta$. Therefore $\psi'(\widehat{z}_i - \overline{\gamma}_p) = 0$ unless $|i| \lesssim LN^\delta$. Moreover, if $|i| \lesssim LN^\delta$ and $(i,j) \in \mathcal{A}$, then $|j| \lesssim LN^\delta$. Hence, the nonzero terms in the sum in (4.253) have both indices $|i|, |j| \lesssim N^{4\omega_\ell+\delta}$. By (4.246) and $C\omega_1 \ll \omega_\ell$, for such terms we have

$$|\widehat{z}_i - \widehat{z}_j| \lesssim \frac{|i-j|}{N^{\frac{3}{4}}\min\{|i|,|j|\}^{\frac{1}{4}}} + \frac{N^{C\omega_1}}{N^{\frac{3}{4}}} \lesssim \frac{L^{\frac{3}{4}}N^{\frac{\delta}{2}}}{N^{\frac{3}{4}}}. \tag{4.256}$$

Note that $\nu|\widehat{z}_i - \widehat{z}_j| \lesssim 1$. Therefore, by Taylor expanding in the exponent, we have

$$\begin{aligned}\left|\frac{\phi_i}{\phi_j} + \frac{\phi_j}{\phi_i} - 2\right| &= \left(e^{\frac{\nu}{2}(\psi(\widehat{z}_j - \overline{\gamma}_p) - \psi(\widehat{z}_i - \overline{\gamma}_p))} - e^{\frac{\nu}{2}(\psi(\widehat{z}_i - \overline{\gamma}_p) - \psi(\widehat{z}_j - \overline{\gamma}_p))}\right)^2 \\ &\lesssim \nu^2 |\psi(\widehat{z}_i - \overline{\gamma}_p) - \psi(\widehat{z}_j - \overline{\gamma}_p)|^2,\end{aligned}$$

and thus

$$\left|\mathcal{B}_{ij}\left(\frac{\phi_i}{\phi_j} + \frac{\phi_j}{\phi_i} - 2\right)\right| \lesssim \nu^2 \frac{|\psi(\widehat{z}_i - \overline{\gamma}_p) - \psi(\widehat{z}_j - \overline{\gamma}_p)|^2}{N(\widehat{z}_i - \widehat{z}_j)^2} \lesssim \frac{\nu^2}{N}, \tag{4.257}$$

where in the last inequality we used that ψ is Lipschitz continuous. Hence we conclude the estimate of (4.253) as

$$\left|\sum_{(i,j)\in\mathcal{A}} \mathcal{B}_{ij} m_i m_j \left(\frac{\phi_i}{\phi_j} + \frac{\phi_j}{\phi_i} - 2\right)\right| \lesssim \frac{\nu^2}{N} \sum_i m_i^2 \sum_j^{\mathcal{A},(i)} \mathbf{1}_{\{\phi_j \neq \phi_i\}} \lesssim \frac{\nu^2 L N^{\frac{3}{4}\delta}}{N} F(t), \tag{4.258}$$

since the number of j's in the summation is at most

$$|j_+(i) - j_-(i)| \leq \ell^4 + \ell |i|^{3/4} \leq L N^{3\delta/4}. \tag{4.259}$$

By (4.249) and since $|\psi'(x)| \leq 1$, (4.255) is bounded as follows

$$\left|\sum_i m_i^2 \left(\frac{\nu^2}{N}\psi'(\widehat{z}_i - \overline{\gamma}_p)^2 + \frac{\nu}{N}\psi''(\widehat{z}_i - \overline{\gamma}_p)\right)\right| \lesssim \left(\frac{\nu^2}{N} + \frac{\nu}{N^{\frac{1}{4}} L^{\frac{3}{4}} N^{\frac{3}{4}\delta}}\right) F(t). \tag{4.260}$$

The next step is to get a bound for (4.254). Since $\psi'(\widehat{z}_i - \overline{\gamma}_p) = 0$ unless $|i| \lesssim N^{4\omega_\ell + \delta} \ll N^{\omega_A}$, choosing $C > 0$ such that $(4 + C)\omega_\ell < \omega_A$ and using (4.155) we get

$$\mathrm{d}(\widehat{z}_i(t) - \overline{\gamma}_p(t)) = \sqrt{\frac{2}{N}}\,\mathrm{d}B_i + \frac{1}{N}\sum_j^{\mathcal{A},(i)} \frac{1}{\widehat{z}_i(t) - \widehat{z}_j(t)} + Q_i(t) \tag{4.261}$$

with

$$\begin{aligned}Q_i(t) := &\int_{\mathcal{I}_{y,i}(t)^c} \frac{\rho_{y,t}(E + \mathfrak{e}_{y,t}^+)}{\widehat{z}_i(t) - E}\,\mathrm{d}E + \Re[m_{y,t}(\mathfrak{e}_{y,t}^+)] + \alpha\Big(\Re[m_{x,t}(\widehat{\gamma}_{x,p}(t) + \mathfrak{e}_{x,t}^+) - m_{x,t}(\mathfrak{e}_{x,t}^+)]\Big) \\ &+ (1-\alpha)\Big(\Re[m_{y,t}(\widehat{\gamma}_{y,p}(t) + \mathfrak{e}_{y,t}^+) - m_{y,t}(\mathfrak{e}_{y,t}^+)]\Big).\end{aligned} \tag{4.262}$$

We insert (4.261) into (4.254) and estimate all three terms separately in the regime $|i| \lesssim LN^\delta$. For the stochastic differential, by the definition of $\tau \leq t_*$ and the Burkholder-Davis-Gundy inequality we have that

$$\sup_{0\leq t\leq\tau} \int_0^t \sqrt{\frac{2}{N}}\nu \sum_i m_i^2 \psi'(\widehat{z}_i - \overline{\gamma}_p)\,\mathrm{d}B_i \leq N^{\epsilon'} \frac{\nu}{\sqrt{N}}\sqrt{t_*} \sup_{0\leq t\leq\tau} F(t) \lesssim \nu N^{\epsilon'} N^{-\frac{3}{4}+\frac{1}{2}\omega_1}, \tag{4.263}$$

for any $\epsilon' > 0$, with very high probability. In (4.263) we used that $\tau \leq t_* \sim N^{-\frac{1}{2}+\omega_1}$, and that, by the definition of τ, $F(t)$ is bounded for all $0 \leq t \leq \tau$.

The contribution of the second term in (4.261) to (4.254) is written, after symmetrisation, as follows

$$\begin{aligned}\frac{\nu}{N}\sum_{(i,j)\in\mathcal{A}}\frac{\psi'(\widehat{z}_i-\overline{\gamma}_p)m_i^2}{\widehat{z}_j-\widehat{z}_i}&=\frac{\nu}{2N}\sum_{(i,j)\in\mathcal{A}}\frac{\psi'(\widehat{z}_i-\overline{\gamma}_p)(m_i^2-m_j^2)}{\widehat{z}_j-\widehat{z}_i}\\&\quad+\frac{\nu}{2N}\sum_{(i,j)\in\mathcal{A}}m_i^2\frac{\psi'(\widehat{z}_i-\overline{\gamma}_p)-\psi'(\widehat{z}_j-\overline{\gamma}_p)}{\widehat{z}_j-\widehat{z}_i}.\end{aligned}\tag{4.264}$$

Using (4.249) and (4.259), the second sum in (4.264) is bounded by

$$\begin{aligned}\left|\frac{\nu}{2N}\sum_{(i,j)\in\mathcal{A}}m_i^2\frac{\psi'(\widehat{z}_i-\overline{\gamma}_p)-\psi'(\widehat{z}_j-\overline{\gamma}_p)}{\widehat{z}_j-\widehat{z}_i}\right|&\lesssim\frac{\nu L^{-\frac{3}{4}}}{N^{\frac{1}{4}+\frac{3}{4}\delta}}\sum_i m_i^2\sum_j^{\mathcal{A},(i)}\mathbf{1}_{\{\psi'(\widehat{z}_i-\overline{\gamma}_p)\neq\psi'(\widehat{z}_j-\overline{\gamma}_p)\}}\\&\lesssim\frac{\nu L^{\frac{1}{4}}}{N^{\frac{1}{4}}}F.\end{aligned}\tag{4.265}$$

Using $m_i^2-m_j^2=(m_i-m_j)(m_i+m_j)$ and Schwarz inequality, the first sum in (4.264) is bounded as follows

$$\begin{aligned}\frac{\nu}{2N}\sum_{(i,j)\in\mathcal{A}}\frac{\psi'(\widehat{z}_i-\overline{\gamma}_p)(m_i^2-m_j^2)}{\widehat{z}_j-\widehat{z}_i}&\le-\frac{1}{100}\sum_{(i,j)\in\mathcal{A}}\mathcal{B}_{ij}(m_i-m_j)^2\\&\quad+\frac{C\nu^2}{2N}\sum_{(i,j)\in\mathcal{A}}\psi'(\widehat{z}_i-\overline{\gamma}_p)^2(m_i^2+m_j^2).\end{aligned}\tag{4.266}$$

The second sum in (4.266), using (4.259), is bounded by

$$\frac{C\nu^2}{2N}\sum_{(i,j)\in\mathcal{A}}\psi'(\widehat{z}_i-\overline{\gamma}_p)(m_i^2+m_j^2)\le\frac{C\nu^2LN^{\frac{3\delta}{4}}}{2N}F,\tag{4.267}$$

hence we conclude that

$$\frac{\nu}{N}\sum_{(i,j)\in\mathcal{A}}\frac{\psi'(\widehat{z}_i-\overline{\gamma}_p)m_i^2}{\widehat{z}_j-\widehat{z}_i}\le-\frac{1}{100}\sum_{(i,j)\in\mathcal{A}}\mathcal{B}_{ij}(m_i-m_j)^2+C\left(\frac{\nu L^{\frac{1}{4}}}{N^{\frac{1}{4}}}+\frac{\nu^2LN^{\frac{3\delta}{4}}}{N}\right)F.\tag{4.268}$$

Note that the first term on the right-hand side of (4.268) can be incorporated in the first, dissipative term in (4.252).

To conclude the estimate of (4.254) we write the third term in (4.261)

$$\begin{aligned}Q_i(t)&=\left(\int_{\mathcal{I}_{y,i}(t)^c}\frac{\rho_{y,t}(E+\mathfrak{e}_{y,t}^+)}{\widehat{z}_i(t)-E}\,\mathrm{d}E+\Re[m_{y,t}(\overline{\gamma}_p(t)+\mathfrak{e}_{y,t}^+)]\right)\\&\quad+\alpha\Big(\Re[m_{x,t}(\widehat{\gamma}_{x,p}(t)+\mathfrak{e}_{x,t}^+)-m_{x,t}(\mathfrak{e}_{x,t}^+)]-\Re[m_{y,t}(\widehat{\gamma}_{x,p}(t)+\mathfrak{e}_{y,t}^+)-m_{y,t}(\mathfrak{e}_{y,t}^+)]\Big)\\&\quad+\alpha\Big(\Re[m_{y,t}(\widehat{\gamma}_{x,p}(t)+\mathfrak{e}_{y,t}^+)]-\Re[m_{y,t}(\overline{\gamma}_p(t)+\mathfrak{e}_{y,t}^+)]\Big)\\&\quad+(1-\alpha)\Big(\Re[m_{y,t}(\widehat{\gamma}_{y,p}(t)+\mathfrak{e}_{y,t}^+)]-\Re[m_{y,t}(\overline{\gamma}_p(t)+\mathfrak{e}_{y,t}^+)\Big)\\&\quad=:A_1+A_2+A_3+A_4.\end{aligned}\tag{4.269}$$

Similarly to the estimates in (4.115), for A_2 we use (4.39a) while for A_3, A_4 we use (4.23b), then we use the asymptotic behavior of $\widehat{\gamma}_p, \overline{\gamma}_p$ by (4.30a) and $p = LN^\delta$ to conclude that

$$|A_2| + |A_3| + |A_4| \lesssim \frac{L^{\frac{1}{4}} N^{\frac{\delta}{4}} N^{C\omega_1} \log N}{N^{\frac{1}{4}} N^{\frac{1}{6}}}. \tag{4.270}$$

For the A_1 term we write it as

$$A_1 = \int_{\mathcal{I}_{y,i}(t)^c} \frac{\overline{\gamma}_p(t) - \widehat{z}_i(t)}{(\widehat{z}_i(t) - E)(\overline{\gamma}_p(t) - E)} \rho_{y,t}(E + \mathfrak{e}_{y,t}^+) \, \mathrm{d}E + \int_{\mathcal{I}_{y,i}(t)} \frac{\rho_{y,t}(E + \mathfrak{e}_{y,t}^+)}{\overline{\gamma}_p(t) - E} \, \mathrm{d}E. \tag{4.271}$$

Since $i \leq Cp$, we have $\rho_{y,t}(E + \mathfrak{e}_{y,t}^+) \leq \rho_{y,t}(\overline{\gamma}_{Cp}(t) + \mathfrak{e}_{y,t}^+) \lesssim L^{\frac{1}{4}} N^{-\frac{1}{4} + \frac{\delta}{4}}$ for any $E \in \mathcal{I}_{y,i}(t)$, the second term in (4.271) is bounded by $L^{\frac{1}{4}} N^{-\frac{1}{4} + \frac{\delta}{4}} \log N$. In the first term in (4.271) we use that

$$|\widehat{z}_i(t) - E| \geq |\overline{\gamma}_i(t) - E| - |\widehat{z}_i(t) - \overline{\gamma}_i(t)| \gtrsim \overline{\gamma}_p(t)$$

for $E \notin \mathcal{I}_{y,i}(t)$, by rigidity (4.246) and by the fact that in the $i \leq Cp$ regime $|\overline{\gamma}_i(t) - \overline{\gamma}_{i \pm j_\pm(i)}(t)| \gtrsim \overline{\gamma}_p(t) \gg N^{-\frac{3}{4} + C\omega_1}$ since $\omega_1 \ll \omega_\ell$ and $= LN^\delta = N^{4\omega_\ell + \omega_1}$.

We thus conclude that the first term in (4.271) is bounded by

$$|\widehat{z}_i(t) - \overline{\gamma}_p(t)| \frac{\Im[m_{y,t}(\mathfrak{e}_{y,t}^+ + \mathrm{i}\overline{\gamma}_p(t))]}{\overline{\gamma}_p(t)} \lesssim \overline{\gamma}_p^{\frac{1}{3}} \lesssim L^{\frac{1}{4}} N^{-\frac{1}{4} + \frac{\delta}{4}},$$

where we used again the rigidity (4.246). In summary, we have

$$|A_1| \lesssim L^{\frac{1}{4}} N^{-\frac{1}{4} + \frac{\delta}{4}} \log N. \tag{4.272}$$

In particular (4.269)-(4.272) imply that

$$Q := \sup_{0 \leq t \leq t_*} \sup_{|i| \lesssim LN^\delta} |Q_i(t)| \lesssim L^{\frac{1}{4}} N^{-\frac{1}{4} + \frac{\delta}{4}} \log N. \tag{4.273}$$

Collecting all the previous estimates using the choice of ν from (4.250) with $\delta' \geq \frac{\delta}{2}$ and that F is bounded up to $t \leq \tau$, we integrate (4.252)–(4.255) from 0 up to time $0 \leq t \leq t_*$ and conclude that

$$\begin{aligned} \sup_{0 \leq t \leq \tau} F(t) - F(0) &\lesssim \left(\frac{\nu^2 L N^{\frac{3\delta}{4} + \omega_1}}{N^{\frac{3}{2}}} + \frac{\nu L^{\frac{1}{4}} N^{\omega_1}}{N^{\frac{3}{4}}} + \frac{\nu Q N^{\omega_1}}{N^{\frac{1}{2}}} \right) \\ &\lesssim \frac{N^{\frac{3\delta}{4} + \omega_1}}{L^{\frac{1}{2}} N^{2\delta'}} + \frac{N^{\omega_1}}{L^{\frac{1}{2}} N^{\delta'}} + \frac{N^{\omega_1 + \frac{\delta}{4}}}{L^{\frac{1}{2}} N^{\delta'}} \log N \leq 1 \end{aligned} \tag{4.274}$$

for large N and with very high probability, where we used the choice of ν (4.250) and that $\omega_1 \ll \omega_\ell$ in the last line. Since $F(0) = 1$, we get that $\tau = t_*$ with very high probability, and so

$$\sup_{0 \leq t \leq t_*} F(t) \leq 5, \tag{4.275}$$

with very high probability.

Furthermore, since $p = LN^\delta$, if $i \leq \frac{3}{4}LN^\delta$, choosing $\delta' = \frac{3\delta}{4} - \epsilon_1$, with $\epsilon_1 \leq \frac{\delta}{4}$, then by Proposition 4.6.3 we have that

$$\nu\psi(\widehat{z}_i(t) - \overline{\gamma}_p) = \nu|\widehat{z}_i(t) - \overline{\gamma}_p| \gtrsim \nu \frac{|i-p|}{N^{\frac{3}{4}}|p|^{\frac{1}{4}}} \gtrsim \frac{N^{\frac{3\delta}{4}}}{N^{\delta'}} = N^{\epsilon_1}$$

with very high probability.

Note that (4.275) implies

$$f_i(t) \leq 5e^{-\frac{1}{2}\nu\psi(\widehat{z}_i(t)-\overline{\gamma}_p)}.$$

Therefore, if $i \leq \frac{3LN^\delta}{4}$ and $p_* \geq p$, then for each fixed $0 \leq t \leq t_*$ we have that

$$\mathcal{U}^{\mathcal{L}}_{ip_*}(0,t) \leq N^{-D}, \tag{4.276}$$

for any $D > 0$ with very high probability. Similar estimate holds if i and p_* are negative or have opposite sign. This proves the estimate on the first term in (4.247) for any fixed s. The estimate for $\mathcal{U}^{\mathcal{L}}_{p_* i}(s,t)$ is analogous with initial condition $f = \delta_i$. This proves Lemma 4.B.1.

□

Next, we enhance this result to a bound uniform in $0 \leq s \leq t_*$. We first have:

Lemma 4.B.2. *Let u be a solution of*

$$\partial_t u = \mathcal{L}u, \tag{4.277}$$

with non-negative initial condition $u_i(0) \geq 0$. Then, for each $0 \leq t \leq t_*$ we have

$$\frac{1}{2}\sum_i u_i(0) \leq \sum_i u_i(t) \leq \sum_i u_i(0) \tag{4.278}$$

with very high probability.

Proof. Since $\mathcal{U}^{\mathcal{L}}$ is a contraction semigroup the upper bound in (4.278) is trivial. Notice that $\partial_t \sum_i u_i = \sum_i \mathcal{V}_i u_i$. Thus the lower bound will follow once we prove $-\mathcal{V}_i \lesssim N^{\frac{1}{2}}L^{-\frac{1}{2}}$ with very high probability since $t_* N^{\frac{1}{2}}L^{-\frac{1}{2}}$ is much smaller than 1 by $\omega_1 \ll \omega_\ell$.

The estimate $-\mathcal{V}_i \lesssim N^{\frac{1}{2}}L^{-\frac{1}{2}}$ proceeds similarly to (4.271). Indeed, for $1 \leq |i| < N^{\omega_A}$, we use $\rho_{y,t}(E + \mathfrak{e}^+_{y,t}) \lesssim |E|^{\frac{1}{3}}$ and that $|\widehat{z}_i(t) - E| \sim |\overline{\gamma}_i(t) - E|$ by rigidity (4.246) and by the fact that

$$|j_+(i) - i|,\ |j_-(i) - i| \gtrsim N^{4\omega_\ell} + N^{\omega_\ell}|i|^{\frac{3}{4}}$$

is much bigger than the rigidity scale. Therefore, we have

$$\begin{aligned} -\mathcal{V}_i = \int_{\mathcal{I}_{y,i}(t)^c} \frac{\rho_{r,t}(E + \mathfrak{e}^+_{r,t})}{(\widehat{z}_i(t) - E)^2}\,\mathrm{d}E &\lesssim \int_{\mathcal{I}_{y,i}(t)^c} \frac{1}{|E - \overline{\gamma}_i(t)|^{\frac{5}{3}}}\,\mathrm{d}E \\ &\quad + \int_{\mathcal{I}_{y,i}(t)^c} \frac{|\overline{\gamma}_i|^{\frac{1}{3}}}{(E - \overline{\gamma}_i(t))^2}\,\mathrm{d}E \\ &\lesssim \frac{N^{\frac{1}{2}}}{N^{2\omega_\ell}} = \frac{N^{\frac{1}{2}}}{L^{\frac{1}{2}}}. \end{aligned}$$

The estimate of $-\mathcal{V}_i$ for $N^{\omega_A} < |i| \leq \frac{i_*}{2}$ is similar. This concludes the proof of Lemma 4.B.2.

□

Finally we prove the following version of Lemma 4.B.1 that is uniform in s:

Lemma 4.B.3. *Under the same hypotheses of Lemma 4.B.1, for any $\delta' > 0$, such that $\delta' < C'\omega_\ell$, with $C' > 0$ the constant defined in Lemma 4.B.1, $|a| \le \frac{LN^{\delta'}}{2}$ and $|b| \ge LN^{\delta'}$ we have that*

$$\sup_{0 \le s \le t \le t_*} \mathcal{U}^{\mathcal{L}}_{ab}(s,t) + \mathcal{U}^{\mathcal{L}}_{ba}(s,t) \le N^{-D} \tag{4.279}$$

with very high probability. The same result holds for $t_* \le s \le t \le 2t_*$ as well.

Proof. By the semigroup property for any $0 \le s \le t \le t_*$ and any j we have that

$$\mathcal{U}^{\mathcal{L}}_{aj}(0,t) \ge \mathcal{U}^{\mathcal{L}}_{ab}(s,t)\mathcal{U}^{\mathcal{L}}_{bj}(0,s). \tag{4.280}$$

Furthermore, by Lemma 4.B.2 for the dual dynamics we have that

$$\frac{1}{2}\sum_j u_j(0) \le \sum_j u_j(s) = \sum_i \sum_j \left(\mathcal{U}^{\mathcal{L}}_{ji}(0,s)\right)^T u_i(0),$$

and so, by choosing $u(0) = \delta_b$ we conclude that

$$\sum_j \mathcal{U}^{\mathcal{L}}_{bj}(0,s) \ge \frac{1}{2}, \qquad \forall\, 0 \le s \le t_*.$$

From the last inequality and since $\sup_{s \le t_*} \mathcal{U}^{\mathcal{L}}_{bj}(0,s) \le N^{-100}$ with very high probability for any $|j| \le \frac{3}{4}LN^{\delta'}$ by Lemma 4.B.1, it follows that there exists an $j_* = j_*(s)$, maybe depending on s, with $|j_*(s)| \ge \frac{3}{4}LN^{\delta'}$, such that $\mathcal{U}^{\mathcal{L}}_{bj_*(s)}(0,s) \ge \frac{1}{4N}$. Furthermore, by the finite speed propagation estimate in Lemma 4.B.1 (this time with $|a| \ge \frac{3}{4}LN^{\delta}$ and $|b| \le \frac{1}{2}LN^{\delta}$; note that its proof only used that $|a-b| \gtrsim LN^{\delta}$), we have that

$$\sup_{t \le t_*} \mathcal{U}^{\mathcal{L}}_{aj_*}(0,t) \le N^{-D}, \qquad \forall |j_*| \ge \frac{3}{4}LN^{\delta'}$$

with very high probability. Hence we get from (4.280) with $j = j_*(s)$ that $\sup_{s \le t} \mathcal{U}^{\mathcal{L}}_{ab}(s,t) \lesssim N^{-D+1}$ with very high probability. The estimate for $\mathcal{U}^{\mathcal{L}}_{ba}(s,t)$ follows in a similar way. This concludes the proof of Lemma 4.B.3. □

Finally, we prove Lemma 4.6.5 and Proposition 4.6.8 which are versions of Lemma 4.B.3 but for the short range approximation on scale $L = N^{1/2 + C_1\omega_1}$ needed in Section 4.6.3.2.

Proof of Lemma 4.6.5. Choosing $L = N^{\frac{1}{2} + C_1\omega_1}$, the proof of Lemma 4.B.1 is exactly the same except for the estimate of Q in (4.273), since, for any $\alpha \in [0,1]$, $Q_i(t)$ from (4.98) is now defined as

$$Q_i(t) := \frac{\beta}{N} \sum_{j:|j-i|>L} \frac{1}{\overline{\gamma}^*_i - \overline{\gamma}^*_j} + \frac{1-\beta}{N} \sum_{j:|j-i|>L} \frac{1}{\widetilde{z}_i - \widetilde{z}_j} \mathrm{d}t + \Phi_\alpha(t), \tag{4.281}$$

with $\Phi_\alpha(t)$ given in (4.72) instead of (4.262). Then Lemma 4.B.2 and Lemma 4.B.3 follow exactly in the same way.

By (4.281) it easily follows that

$$Q := \sup_{0\le t\le t_*} \sup_{|i|\le LN^{\delta'}} |Q_i(t)| \lesssim \log N. \tag{4.282}$$

Hence, by an estimate similar to (4.274), we conclude that

$$\begin{aligned}\sup_{0\le t\le\tau} F(t) - F(0) &\lesssim \left(\frac{\nu^2 L N^{\frac{3\delta}{4}+\omega_1}}{N^{\frac{3}{2}}} + \frac{\nu L^{\frac{1}{4}} N^{\omega_1}}{N^{\frac{3}{4}}} + \frac{\nu Q N^{\omega_1}}{N^{\frac{1}{2}}} \right) \\ &\lesssim \frac{N^{\frac{3\delta}{4}+\omega_1}}{L^{\frac{1}{2}} N^{\delta'}} + \frac{N^{\omega_1}}{L^{\frac{1}{2}} N^{\delta'}} + \frac{N^{\frac{3}{4}+\omega_1}}{L^{\frac{3}{4}} N^{\frac{1}{2}} N^{\delta'}} \log N \le 1,\end{aligned} \tag{4.283}$$

with very high probability. Note that in the last inequality we used that $L = N^{\frac{1}{2}+C_1\omega_1}$. □

Proof of Proposition 4.6.8. This proof is almost identical to the previous one, except that $Q_i(t)$ is now defined from (4.109) as

$$Q_i(t) := \beta \left[\frac{1}{N} \sum_{j:|j-i|>L} \frac{1}{\overline{\gamma}_i^* - \overline{\gamma}_j^*} + \Phi(t) \right] + (1-\beta) \left[\frac{\mathrm{d}}{\mathrm{d}t} \overline{\gamma}_i^*(t) - \frac{1}{N} \sum_{j:|j-i|\le L} \frac{1}{\overline{\gamma}_i^* - \overline{\gamma}_j^*} \right],$$

which satisfies the same bound (4.282). The rest of the proof is unchanged. □

4.C Short-long approximation

In this section we estimate the difference of the solution of the DBM $\widetilde{z}(t,\alpha)$ and its short range approximation $\widehat{z}(t,\alpha)$, closely following the proof of Lemma 3.7 in [131] and adapting it to the more complicated cusp situation. In particular, in Section 4.C.1 we estimate $|\widetilde{z}(t,\alpha) - \widehat{z}(t,\alpha)|$ for $0 \le t \le t_*$, i.e. until the formation of an exact cusp; in Section 4.C.2, instead, we estimate $|\widetilde{z}(t,\alpha) - \widehat{z}(t,\alpha)|$ for $t_* < t \le 2t_*$, i.e. after the formation of a small minimum. The precision of this approximation depends on the rigidity bounds we put as a condition. We consider a two-scale rigidity assumption, a weaker rigidity valid for all indices and a stronger rigidity valid for $1 \le |i| \lesssim i_* = N^{\frac{1}{2}+C_*\omega_1}$; both described by an exponent.

4.C.1 Short-long approximation: Small gap and exact cusp.

In this subsection we estimate the difference of the solution of the DBM $\widetilde{z}(t,\alpha)$ defined in (4.71) and its short range approximation $\widehat{z}(t,\alpha)$ defined by (4.154)-(4.157) for $0 \le t \le t_*$. We formulate Lemma 4.C.1 (for $0 \le t \le t_*$) below a bit more generally than we need in order to indicate the dependence of the approximation precision on these two exponents. For our actual application in Lemma 4.6.9 and Lemma 4.7.2 we use specific exponents.

Lemma 4.C.1. *Let* $\omega_1 \ll \omega_\ell \ll \omega_A \ll 1$. *Let* $0 < a_0 \le \frac{1}{4} + C\omega_1$, $C > 0$ *a universal constant and* $0 < a \le C\omega_1$. *Let* $i_* := N^{\frac{1}{2}+C_*\omega_1}$ *with* C_* *defined in Proposition 4.6.3. We assume that*

$$|\widetilde{z}_i(t,\alpha) - \overline{\gamma}_i(t)| \le \frac{N^{a_0}}{N^{\frac{3}{4}}}, \qquad 1 \le |i| \le N, \quad 0 \le t \le t_* \tag{4.284}$$

and that

$$|\widetilde{z}_i(t,\alpha) - \overline{\gamma}_i(t)| \leq \frac{N^a}{N^{\frac{3}{4}}}, \qquad 1 \leq |i| \leq i_*, \quad 0 \leq t \leq t_*. \tag{4.285}$$

Then, for any $\alpha \in [0,1]$, *we have that*

$$\sup_{1\leq|i|\leq N} \sup_{0\leq t\leq t_*} |\widehat{z}_i(t,\alpha) - \widetilde{z}_i(t,\alpha)|$$
$$\leq \frac{N^a N^{C\omega_1}}{N^{\frac{3}{4}}} \left(\frac{1}{N^{2\omega_\ell}} + \frac{N^{\frac{\omega_A}{2}} \log N}{N^{\frac{1}{6}} N^a} + \frac{N^{\frac{\omega_A}{2}} \log N}{N^{\frac{1}{4}} N^a} + \frac{1}{N^{\frac{2a}{5}} i_*^{\frac{1}{5}}} + \frac{N^{a_0}}{N^a i_*^{\frac{1}{2}}} + \frac{1}{N^{\frac{1}{18}} N^a} \right), \tag{4.286}$$

with very high probability.

Proof of Lemma 4.6.9. Use Lemma 4.C.1 with the choice $a_0 = \frac{1}{4} + C\omega_1$ and $a = C\omega_1$, for some universal constant $C > 0$. The conditions (4.284) and (4.285) are guaranteed by (4.78) and (4.79). □

Proof of Lemma 4.C.1. Let $w_i := \widehat{z}_i - \widetilde{z}_i$, hence w is a solution of

$$\partial_t w = \mathcal{B}_1 w + \mathcal{V}_1 w + \zeta, \tag{4.287}$$

where the operator $\mathcal{B}_1$ is defined for any $f \in \mathbf{C}^{2N}$ by

$$(B_1 f)_i = \frac{1}{N} \sum_j^{\mathcal{A},(i)} \frac{f_j - f_i}{(\widetilde{z}_i(t,\alpha) - \widetilde{z}_j(t,\alpha))(\widehat{z}_i(t,\alpha) - \widehat{z}_j(t,\alpha))}. \tag{4.288}$$

The diagonal operator $\mathcal{V}_1$ is defined by $(\mathcal{V}_1 f)_i = \mathcal{V}_1(i) f_i$, where

$$\mathcal{V}_1(i) := -\int_{\mathcal{I}_{y,i}(t)^c} \frac{\rho_{y,t}(E + \mathfrak{e}_{y,t}^+)}{(\widetilde{z}_i(t,\alpha) - E)(\widehat{z}_i(t,\alpha) - E)} \, \mathrm{d}E, \qquad \text{for} \quad 0 < |i| \leq N^{\omega_A}, \tag{4.289}$$

and

$$\mathcal{V}_1(i) := -\int_{\mathcal{I}_{z,i}(t)^c \cap \mathcal{J}_z(t)} \frac{\overline{\rho}_t(E + \overline{\mathfrak{e}}_t^+)}{(\widetilde{z}_i(t,\alpha) - E)(\widehat{z}_i(t,\alpha) - E)} \, \mathrm{d}E, \qquad \text{for} \quad N^{\omega_A} < |i| \leq \frac{i_*}{2}. \tag{4.290}$$

Finally, $\mathcal{V}_1(i) = 0$ for $|i| \geq \frac{i_*}{2}$. The vector ζ in (4.287) collects various error terms.

We define the stopping time

$$T := \max\{t \in [0,t_*] | \sup_{0\leq s\leq t} |\widetilde{z}_i(s,\alpha) - \widehat{z}_i(s,\alpha)| \leq \frac{1}{2} \min\{|\mathcal{I}_{z,i}(t)|, |\mathcal{I}_{y,i}(t)|, \} \, \forall \alpha \in [0,1]\}, \tag{4.291}$$

where we recall that $|\mathcal{I}_{z,i}(t)| \sim |\mathcal{I}_{y,i}(t)| \sim N^{-\frac{3}{4}+3\omega_\ell}$.

For $0 \leq t \leq T$ we have that $\mathcal{V}_1 \leq 0$. Therefore, since $\sum_i (\mathcal{B}f)_i = 0$, by the symmetry of $\mathcal{A}$, the semigroup of $\mathcal{B}_1 + \mathcal{V}_1$, denoted by $\mathcal{U}^{\mathcal{B}_1+\mathcal{V}_1}$, is a contraction on every ℓ^p space. Hence, since $w(0) = 0$ by (4.157), we have that

$$w(t) = \int_0^t \mathcal{U}^{\mathcal{B}_1+\mathcal{V}_1}(s,t)\zeta(s) \, \mathrm{d}s$$

and so

$$\|w(t)\|_\infty \le t \sup_{0\le s\le t} \|\zeta(s)\|_\infty \le N^{-\frac{1}{2}+\omega_1} \sup_{0\le s\le t} \|\zeta(s)\|_\infty. \tag{4.292}$$

Thus, to prove (4.286) it is enough to estimate $\|\zeta(s)\|_\infty$, for all $0 \le s \le t_*$.

The error term ζ is given by $\zeta_i = 0$ for $|i| > \frac{i_*}{2}$, then for $1 \le |i| \le N^{\omega_A}$, ζ_i is defined as

$$\zeta_i = \int_{\mathcal{I}_{y,i}(t)^c} \frac{\rho_{y,t}(E+\mathfrak{e}_{y,t}^+)}{\widetilde{z}_i(t,\alpha)-E}\,\mathrm{d}E - \frac{1}{N}\sum_j^{\mathcal{A}^c,(i)} \frac{1}{\widetilde{z}_i(t,\alpha)-\widetilde{z}_j(t,\alpha)} + \Phi_\alpha(t) - \Re[m_{y,t}(\mathfrak{e}_{y,t}^+)], \tag{4.293}$$

with $\Phi_\alpha(t)$ defined in (4.72), and for $N^{\omega_A} < |i| \le \frac{i_*}{2}$ as

$$\zeta_i = \int_{\mathcal{I}_{z,i}(t)^c \cap \mathcal{J}_z(t)} \frac{\overline{\rho}_t(E+\overline{\mathfrak{e}}_t^+)}{\widetilde{z}_i(t,\alpha)-E}\,\mathrm{d}E - \frac{1}{N}\sum_{1\le|j|<\frac{3i_*}{4}}^{\mathcal{A}^c,(i)} \frac{1}{\widetilde{z}_i(t,\alpha)-\widetilde{z}_j(t,\alpha)}. \tag{4.294}$$

Note that in the sum in (4.294) we do not have the summation over $|j| \ge \frac{3i_*}{4}$ since if $1 \le |i| \le \frac{i_*}{2}$ and $|j| \ge \frac{3i_*}{4}$ then $(i,j) \in \mathcal{A}^c$.

In the following we will often omit the t and the α arguments from $\widetilde{z}_i$ and $\overline{\gamma}_i$ for notational simplicity.

First, we consider the error term (4.294) for $N^{\omega_A} < |i| \le \frac{i_*}{2}$. We start with the estimate

$$\begin{aligned}|\zeta_i| &= \left| \int_{\mathcal{I}_{z,i}^c(t)\cap\mathcal{J}_z(t)} \frac{\overline{\rho}_t(E+\overline{\mathfrak{e}}_t^+)}{\widetilde{z}_i-E}\,\mathrm{d}E - \frac{1}{N}\sum_{1\le|j|<\frac{3i_*}{4}}^{\mathcal{A}^c,(i)} \frac{1}{\widetilde{z}_i-\widetilde{z}_j}\right| \\ &\lesssim \left|\sum_{1\le|j|<\frac{3i_*}{4}}^{\mathcal{A}^c,(i)} \int_{\overline{\gamma}_j}^{\overline{\gamma}_{j+1}} \frac{\overline{\rho}_t(E+\overline{\mathfrak{e}}_t^+)(E-\overline{\gamma}_j)}{(\widetilde{z}_i-E)(\widetilde{z}_i-\overline{\gamma}_j)}\,\mathrm{d}E\right| + \left|\frac{1}{N}\sum_{1\le|j|<\frac{3i_*}{4}}^{\mathcal{A}^c,(i)} \frac{\widetilde{z}_j-\overline{\gamma}_j}{(\widetilde{z}_i-\widetilde{z}_j)(\widetilde{z}_i-\overline{\gamma}_j)}\right| \\ &\quad + \left|\int_{\overline{\gamma}_{j_+}}^{\overline{\gamma}_{j_+ +1}} \frac{\overline{\rho}_t(E+\overline{\mathfrak{e}}_t^+)}{\widetilde{z}_i-E}\,\mathrm{d}E\right| + \left|\int_{\overline{\gamma}_{-\frac{3i_*}{4}}}^{\overline{\gamma}_{-\frac{3i_*}{4}+1}} \frac{\overline{\rho}_t(E+\overline{\mathfrak{e}}_t^+)}{\widetilde{z}_i-E}\,\mathrm{d}E\right| + \left|\int_0^{\overline{\gamma}_1} \frac{\overline{\rho}_t(E+\overline{\mathfrak{e}}_t^+)}{\widetilde{z}_i-E}\,\mathrm{d}E\right|.\end{aligned} \tag{4.295}$$

Since $|j_+ - i| \ge N^{4\omega_\ell} + N^{\omega_\ell}|i|^{\frac{3}{4}}$ and N^{ω_A}, i.e.

$$|\overline{\gamma}_{j_+} - \overline{\gamma}_i| \ge \frac{N^{\omega_\ell}|i|^{\frac{1}{2}}}{N^{\frac{3}{4}}}$$

is bigger than the rigidity scale (4.285), all terms in the last line of (4.295) are bounded by $N^{-\frac{1}{4}-3\omega_\ell}$.

Then, using the rigidity estimate in (4.285) for the first and the second term of the rhs. of (4.295), we conclude that

$$|\zeta_i| \lesssim \frac{N^a}{N^{\frac{7}{4}}} \sum_{1\le|j|<\frac{3i_*}{4}}^{\mathcal{A}^c,(i)} \frac{1}{(\overline{\gamma}_i-\overline{\gamma}_j)^2} + N^{-\frac{1}{4}-3\omega_\ell}. \tag{4.296}$$

The sum on the rhs. of (4.296) is over all the j, negative and positive, but the main contribution comes from i and j with the same sign, because if i and j have opposite sign then

$$\frac{1}{(\overline{\gamma}_i-\overline{\gamma}_j)^2}\le\frac{1}{(\overline{\gamma}_{-i}-\overline{\gamma}_j)^2}.$$

Hence, assuming that i is positive (for negative i's we proceed exactly in the same way), we conclude that

$$|\zeta_i|\lesssim\frac{N^a}{N^{\frac{7}{4}}}\sum_{1\le j<\frac{3i_*}{4}}^{\mathcal{A}^c,(i)}\frac{1}{(\overline{\gamma}_i-\overline{\gamma}_j)^2}+N^{-\frac{1}{4}-3\omega_\ell}. \tag{4.297}$$

From now we assume that both i and j are positive. In order to estimate (4.297) we use the explicit expression of the quantiles from (4.30a), i.e.

$$\overline{\gamma}_j\sim\max\left\{\Big(\frac{j}{N}\Big)^{2/3}\overline{\Delta}_t^{\frac{1}{9}},\Big(\frac{j}{N}\Big)^{3/4}\right\},$$

where $\overline{\Delta}_t\lesssim t_*^{3/2}$ denotes the length of the small gap of $\overline{\rho}_t$, for all $|j|\le i_*\sim N^{\frac{1}{2}}$. A simple calculation from (4.30a) shows that in the regime $i\ge N^{\omega_A}$ and $j\in\mathcal{A}^c$ we may replace $|\overline{\gamma}_i-\overline{\gamma}_j|\sim|\gamma_{y,i}(t)-\gamma_{y,j}(t)|\sim|i^{3/4}-j^{3/4}|/N^{3/4}$, hence

$$|\zeta_i|\lesssim\frac{N^a}{N^{\frac{1}{4}}}\sum_{1\le j<\frac{3i_*}{4}}^{\mathcal{A}^c,(i)}\frac{i^{\frac{1}{2}}+j^{\frac{1}{2}}}{(i-j)^2}+N^{-\frac{1}{4}-3\omega_\ell}. \tag{4.298}$$

In fact, the same replacement works if either $i\ge N^{4\omega_\ell}$ or $j\ge N^{4\omega_\ell}$ and at least one of these two inequalities always hold as $(i,j)\in\mathcal{A}^c$. Using $i\le\frac{i_*}{2}$ and that by the restriction $(i,j)\in\mathcal{A}^c$ we have $|j-i|\ge\ell(\ell^3+i^{\frac{3}{4}})$, elementary calculation gives

$$|\zeta_i|\lesssim\frac{N^a}{N^{\frac{1}{4}}N^{2\omega_\ell}}. \tag{4.299}$$

Since analogous computations hold for i and j both negative, we have

$$|\zeta_i|\lesssim\frac{N^a}{N^{\frac{1}{4}}N^{2\omega_\ell}},\qquad\text{for any}\quad N^{\omega_A}<|i|\le\frac{i_*}{2}. \tag{4.300}$$

with very high probability.

Next, we proceed with the bound for ζ_i for $|i|\le N^{\omega_A}$. From (4.293) we have

$$\begin{aligned}\zeta_i=&\left(\int_{\mathcal{I}_{z,i}(t)^c\cap\mathcal{J}_z(t)}\frac{\overline{\rho}_t(E+\overline{\mathfrak{e}}_t^+)}{\widetilde{z}_i-E}\,\mathrm{d}E-\frac{1}{N}\sum_{|j|<\frac{3i_*}{4}}^{\mathcal{A}^c,(i)}\frac{1}{\widetilde{z}_i-\widetilde{z}_j}\right)\\&+\left(\int_{\mathcal{J}_z(t)^c}\frac{\overline{\rho}_t(E+\overline{\mathfrak{e}}_t^+)}{\widetilde{z}_i-E}\,\mathrm{d}E-\frac{1}{N}\sum_{|j|\ge\frac{3i_*}{4}}^{\mathcal{A}^c,(i)}\frac{1}{\widetilde{z}_i-\widetilde{z}_j}\right)\\&+\Phi_\alpha(t)-\Re[\overline{m}_t(\widetilde{z}_i+\overline{\mathfrak{e}}_t^+)]+\Re[m_{y,t}(\widetilde{z}_i+\mathfrak{e}_{y,t}^+)]-\Re[m_{y,t}(\mathfrak{e}_{y,t}^+)]\\&+\left(\int_{\mathcal{I}_{z,i}(t)}\frac{\overline{\rho}_t(E+\overline{\mathfrak{e}}_t^+)}{\widetilde{z}_i-E}\,\mathrm{d}E-\int_{\mathcal{I}_{y,i}(t)}\frac{\rho_{y,t}(E+\mathfrak{e}_{y,t}^+)}{\widetilde{z}_i-E}\,\mathrm{d}E\right)=:A_1+A_2+A_3+A_4.\end{aligned} \tag{4.301}$$

By the remark after (4.298), the estimate of A_1 proceeds as in (4.298) and so we conclude that

$$|A_1| \lesssim \frac{N^a}{N^{\frac{1}{4}} N^{2\omega_\ell}}. \tag{4.302}$$

To estimate A_2, we first notice that the restriction $(i,j) \in \mathcal{A}^c$ in the summation is superfluous for $|i| \le N^{\omega_A}$ and $|j| \ge \frac{3}{4} i_*$. Let $\eta_1 \in [N^{-\frac{3}{4}+\frac{3}{4}\omega_A}, N^{-\delta}]$, for some small fixed $\delta > 0$, be an auxiliary scale we will determine later in the proof, then we write A_2 as follows:

$$\begin{aligned} A_2 = &\left(\int_{\mathcal{J}_z(t)^c} \frac{\overline{\rho}_t(E+\mathfrak{e}_t^+)}{\widetilde{z}_i - E} \, \mathrm{d}E - \int_{\mathcal{J}_z(t)^c} \frac{\overline{\rho}_t(E+\mathfrak{e}_t^+)}{\widetilde{z}_i - E + \mathrm{i}\eta_1} \, \mathrm{d}E \right) \\ &+ \left(\frac{1}{N} \sum_{|j| \ge \frac{3i_*}{4}} \frac{1}{\widetilde{z}_i - \widetilde{z}_j + \mathrm{i}\eta_1} - \frac{1}{N} \sum_{|j| \ge \frac{3i_*}{4}} \frac{1}{\widetilde{z}_i - \widetilde{z}_j} \right) \\ &+ \left(\frac{1}{N} \sum_{|j| < \frac{3i_*}{4}} \frac{1}{\widetilde{z}_i - \widetilde{z}_j + \mathrm{i}\eta_1} - \int_{\mathcal{J}_z(t)} \frac{\overline{\rho}_t(E+\mathfrak{e}_t^+)}{\widetilde{z}_i - E + \mathrm{i}\eta_1} \, \mathrm{d}E \right) \\ &+ (\overline{m}_t(\widetilde{z}_i + \mathrm{i}\eta_1) - m_{2N}(\widetilde{z}_i + \mathrm{i}\eta_1, t, \alpha)) =: A_{2,1} + A_{2,2} + A_{2,3} + A_{2,4}, \end{aligned} \tag{4.303}$$

where we introduced

$$m_{2N}(z,t,\alpha) := \frac{1}{N} \sum_{|j| \le N} \frac{1}{z_j(t,\alpha) - z}, \qquad z \in \mathbf{H}.$$

For $1 \le |i| \le N^{\omega_A}$ and $|j| > \frac{3i_*}{4}$, the term $A_{2,2}$ is bounded by the crude rigidity (4.284) as

$$|A_{2,2}| \le \frac{1}{N} \sum_{|j| > \frac{3i_*}{4}} \frac{\eta_1}{(\widetilde{z}_i - \widetilde{z}_j)^2} \lesssim \frac{N^{\frac{1}{2}} \eta_1}{i_*^{\frac{1}{2}}}. \tag{4.304}$$

Exactly the same estimate holds for $A_{2,1}$.

Next, using the rigidity estimates in (4.284) and (4.285) we conclude that

$$\begin{aligned} |A_{2,4}| &\lesssim \frac{1}{N} \sum_{1 \le |j| \le i_*} \frac{|\widetilde{z}_j - \overline{\gamma}_j|}{|\widetilde{z}_i - \widetilde{z}_j + \mathrm{i}\eta_1|^2} + \frac{1}{N} \sum_{i_* \le |j| \le N} \frac{|\widetilde{z}_j - \overline{\gamma}_j|}{|\widetilde{z}_i - \widetilde{z}_j + \mathrm{i}\eta_1|^2} \\ &\lesssim \frac{N^a}{N^{\frac{3}{4}} \eta_1} \Im m_N(\overline{\gamma}_i + \mathrm{i}\eta_1) + \frac{N^{a_0}}{N^{\frac{7}{4}}} \sum_{i_* \le |j| \le N} \frac{1}{(\overline{\gamma}_i - \overline{\gamma}_j)^2} \\ &\lesssim \frac{N^a}{N^{\frac{3}{4}} \eta_1} \left(\frac{N^{\frac{3\omega_A}{4}}}{N^{\frac{3}{4}}} + \eta_1 \right)^{\frac{1}{3}} + \frac{N^{a_0}}{N^{\frac{1}{4}} i_*^{\frac{1}{2}}} \lesssim \frac{N^a}{N^{\frac{3}{4}} \eta_1^{\frac{2}{3}}} + \frac{N^{a_0}}{i_*^{\frac{1}{2}} N^{\frac{1}{4}}}. \end{aligned} \tag{4.305}$$

Here we used that the rigidity scale near i for $1 \le |i| \le N^{\omega_A}$ is much smaller than $\eta_1 \ge N^{-\frac{3}{4}+\frac{3}{4}\omega_A}$. In particular, we know that $\Im m_N(\overline{\gamma}_i + \mathrm{i}\eta_1)$ can be bounded by the density $\overline{\rho}_t(\overline{\gamma}_i + \eta_1)$ which in turn is bounded by $(\overline{\gamma}_i + \eta_1)^{1/3}$. Similarly we conclude that

$$|A_{2,3}| \le \frac{N^a}{N^{\frac{3}{4}} \eta_1^{\frac{2}{3}}}.$$

Optimizing (4.304) and (4.305) for η_1, we choose $\eta_1 = (i_*^{1/2} N^{a-5/4})^{3/5}$ which falls into the required interval for η_1. Collecting all estimates for the parts of A_2 in (4.303), we therefore conclude that

$$|A_2| \le \frac{N^{\frac{3a}{5}}}{i_*^{\frac{1}{5}} N^{\frac{1}{4}}} + \frac{N^{a_0}}{i_*^{\frac{1}{2}} N^{\frac{1}{4}}}. \tag{4.306}$$

Next, we treat A_3 from (4.301). $\Phi_\alpha(t) = \Re[\overline{m}_t(\overline{\mathfrak{e}}_t^+)] + \mathcal{O}(N^{-1})$ by (4.73), then by (4.39a) we conclude that

$$\begin{aligned} |A_3| &= |\Re[\overline{m}_t(\overline{\mathfrak{e}}_t^+)] - \Re[\overline{m}_t(\widetilde{z}_i + \overline{\mathfrak{e}}_t^+)] + \Re[m_{y,t}(\widetilde{z}_i + \mathfrak{e}_{y,t}^+)] - \Re[m_{y,t}(\mathfrak{e}_{y,t}^+)]| \\ &\lesssim \left(\frac{|i|^{\frac{1}{4}} N^{\frac{7\omega_1}{18}}}{N^{\frac{1}{4}} N^{\frac{1}{6}}} + \frac{|i|^{\frac{1}{2}}}{N^{\frac{1}{2}}} \right) |\log|\overline{\gamma}_i|| \lesssim \frac{N^{\frac{\omega_A}{4}} N^{\frac{7\omega_1}{18}} \log N}{N^{\frac{1}{4}} N^{\frac{1}{6}}} + \frac{N^{\frac{\omega_A}{2}} \log N}{N^{\frac{1}{2}}}. \end{aligned} \tag{4.307}$$

We proceed writing A_4 as

$$\begin{aligned} A_4 = &\left(\int_{\mathcal{I}_{z,i}(t)} \frac{\overline{\rho}_t(E + \overline{\mathfrak{e}}_t^+)}{\widetilde{z}_i - E} \,\mathrm{d}E - \int_{\mathcal{I}_{z,i}(t)} \frac{\rho_{y,t}(E + \mathfrak{e}_{y,t}^+)}{\widetilde{z}_i - E} \,\mathrm{d}E \right) \\ &+ \left(\int_{\mathcal{I}_{z,i}(t)} \frac{\rho_{y,t}(E + \mathfrak{e}_{y,t}^+)}{\widetilde{z}_i - E} \,\mathrm{d}E - \int_{\mathcal{I}_{y,i}(t)} \frac{\rho_{y,t}(E + \mathfrak{e}_{y,t}^+)}{\widetilde{z}_i - E} \,\mathrm{d}E \right) =: A_{4,1} + A_{4,2}. \end{aligned} \tag{4.308}$$

We start with the estimate for $A_{4,2}$. By (4.153) and the comparison estimates between $\overline{\gamma}_{z,i}$ and $\widehat{\gamma}_{y,i}$ by (4.35) we have that

$$|\mathcal{I}_{z,i}(t) \Delta \mathcal{I}_{y,i}(t)| \lesssim |\overline{\gamma}_{z,i-j_-(i)} - \widehat{\gamma}_{y,i-j_-(i)}| + |\overline{\gamma}_{z,i+j_+(i)} - \widehat{\gamma}_{y,i+j_+(i)}| \lesssim \frac{N^{\frac{\omega_1}{2}} (\ell^3 + |i|^{\frac{3}{4}})}{N^{\frac{11}{12}}}, \tag{4.309}$$

where Δ is the symmetric difference. In the second inequality of (4.309) we used that $|i \pm j_\pm(i)| \lesssim N^{\omega_A}$ and $\omega_A \ll 1$. For $E \in \mathcal{I}_{z,i} \Delta \mathcal{I}_{y,i}$ we have that

$$\left|\frac{\rho_{y,t}(E + \mathfrak{e}_{y,t}^+)}{\widetilde{z}_i - E}\right| \lesssim \frac{N^{\frac{1}{2}} (\ell^2 + |i|^{\frac{1}{2}})}{\ell^3 + |i|^{\frac{3}{4}}}, \tag{4.310}$$

and so, using $|i| \le N^{\omega_A}$,

$$|A_{4,2}| \lesssim \frac{N^{\frac{\omega_1}{2}} N^{\frac{\omega_A}{2}}}{N^{\frac{5}{12}}} = \frac{N^{\frac{\omega_1}{2}} N^{\frac{\omega_A}{2}}}{N^{\frac{1}{4}} N^{\frac{1}{6}}} \tag{4.311}$$

with very high probability.

To estimate the integral in $A_{4,1}$ we have to deal with the logarithmic singularity due to the values of E close to $\widetilde{z}_i(t)$. For $\max\{\overline{\mathfrak{e}}_t^-, \mathfrak{e}_{y,t}^-\} < E \le 0$ we have that

$$\rho_{y,t}(E + \mathfrak{e}_{y,t}^+) = \overline{\rho}_t(E + \overline{\mathfrak{e}}_t^+) = 0. \tag{4.312}$$

For $\min\{\overline{\mathfrak{e}}_t^-, \mathfrak{e}_{y,t}^-\} \le E \le \max\{\overline{\mathfrak{e}}_t^-, \mathfrak{e}_{y,t}^-\}$, using the $\frac{1}{3}$-Hölder continuity of $\overline{\rho}_t$ and $\rho_{y,t}$ and (4.22a) we have that

$$|\rho_{y,t}(E + \mathfrak{e}_{y,t}^+) - \overline{\rho}_t(E + \overline{\mathfrak{e}}_t^+)| \lesssim \Delta_{y,t}^{\frac{1}{3}} (t_* - t)^{\frac{1}{9}} \lesssim \frac{N^{\frac{11\omega_1}{18}}}{N^{\frac{11}{36}}}, \tag{4.313}$$

for all $0 \leq t \leq t_*$. In the last inequality we used that $\Delta_{y,t} \leq \Delta_{y,0} \lesssim N^{-\frac{3}{4}+\frac{3\omega_1}{2}}$ for all $t \leq t_*$. Similarly, for $E \leq \min\{\bar{\mathfrak{e}}_t^-, \mathfrak{e}_{y,t}^-\}$ we have that

$$|\rho_{y,t}(E+\mathfrak{e}_{y,t}^+) - \overline{\rho}_t(E+\bar{\mathfrak{e}}_t^+)| \lesssim |\rho_{y,t}(E'+\mathfrak{e}_{y,t}^-) - \overline{\rho}_t(E'+\bar{\mathfrak{e}}_t^-)| + \Delta_{y,t}^{\frac{1}{3}}(t_*-t)^{\frac{1}{9}}, \quad (4.314)$$

with $E' \leq 0$.

Using (4.22b) for $E \geq 0$ and combining (4.22b) with (4.312)-(4.314) for $E < 0$, we have that

$$\begin{aligned} |A_{4,1}| \lesssim & \left(\frac{(\ell+|i|^{\frac{1}{4}})N^{\frac{\omega_1}{3}}}{N^{\frac{1}{4}}N^{\frac{1}{6}}} + \frac{(\ell^2+|i|^{\frac{1}{2}})}{N^{\frac{1}{2}}} + \frac{N^{\frac{11\omega_1}{18}}}{N^{\frac{11}{36}}} \right) \int_{\mathcal{I}_{,i}(t)\cap\{|E-\widetilde{z}_i|>N^{-60}\}} \frac{1}{|\widetilde{z}_i - E|} \mathrm{d}E \\ & + \left| \int_{|E-\widetilde{z}_i|\leq N^{-60}} \frac{\overline{\rho}_t(E+\bar{\mathfrak{e}}_t^+) - \rho_{y,t}(E+\mathfrak{e}_{y,t}^+)}{\widetilde{z}_i - E} \mathrm{d}E \right|. \end{aligned} \quad (4.315)$$

The two singular integrals in the second line are estimated separately. By the $\frac{1}{3}$-Hölder continuity $\rho_{y,t}$ we conclude that

$$\begin{aligned} \left| \int_{|E-\widetilde{z}_i|\leq N^{-60}} \frac{\rho_{y,t}(E+\mathfrak{e}_{y,t}^+)}{\widetilde{z}_i - E} \mathrm{d}E \right| &= \left| \int_{|E-\widetilde{z}_i|\leq N^{-60}} \frac{\rho_{y,t}(E+\mathfrak{e}_{y,t}^+) - \rho_{y,t}(\widetilde{z}_i+\mathfrak{e}_{y,t}^+)}{\widetilde{z}_i - E} \mathrm{d}E \right| \\ &\lesssim \int_{|E-\widetilde{z}_i|\leq N^{-60}} \frac{1}{|\widetilde{z}_i - E|^{\frac{2}{3}}} \mathrm{d}E \lesssim N^{-20}. \end{aligned}$$

The same bound holds for the other singular integral in (4.315) by using the $\frac{1}{3}$-Hölder continuity of $\overline{\rho}_t$. Hence, for $1 \leq |i| \leq N^{\omega_A}$, by (4.315) we have that

$$|A_{4,1}| \leq \frac{N^{\frac{\omega_A}{4}}N^{\frac{\omega_1}{3}}\log N}{N^{\frac{1}{4}}N^{\frac{1}{6}}} + \frac{N^{\frac{\omega_A}{2}}\log N}{N^{\frac{1}{2}}} + \frac{N^{\frac{11\omega_1}{18}}\log N}{N^{\frac{11}{36}}}, \quad (4.316)$$

with very high probability.

Collecting all the estimates (4.300), (4.302), (4.306), (4.307), (4.311) and (4.316), and recalling $\omega_1 \ll \omega_\ell \ll \omega_A \ll 1$, we see that (4.302) is the largest term and thus $|\zeta| \lesssim N^{-\frac{1}{4}-2\omega_\ell}N^{C\omega_1}$ as $a \leq C\omega_1$. Thus, using (4.292), we conclude that the estimate in (4.286) is satisfied for all $0 \leq t \leq T$. In particular, this means that

$$|\widehat{z}_i(t,\alpha) - \widetilde{z}_i(t,\alpha)| \leq N^{-\frac{3}{4}+C\omega_1}, \qquad 0 \leq t \leq T,$$

for some small constant $C > 0$. We conclude the proof of this lemma showing that $T \geq t_*$.

Suppose by contradiction that $T < t_*$, then, since the solution of the DBM have continuous paths (see Theorem 12.2 of [90]), we have that

$$|\widehat{z}_i(T+\tilde{t},\alpha) - \widetilde{z}_i(T+\tilde{t},\alpha)| \leq \frac{N^a N^{c\omega_1}}{N^{\frac{3}{4}}N^{2\omega_\ell}},$$

for some tiny $\tilde{t} > 0$ and for any $\alpha \in [0,1]$. This bound is much smaller than the threshold $|\mathcal{I}_{y,i}(t)|, |\mathcal{I}_{z,i}(t)| \sim N^{-\frac{3}{4}+3\omega\ell}$ in the definition of T. But this is a contradiction by the maximality in the definition of T, hence $T = t_*$, proving (4.286) for all $0 \leq t \leq t_*$. This completes the proof of Lemma 4.C.1. □

Proof of Lemma 4.7.2. The proof of this lemma is very similar to that of Lemma 4.C.1, hence we will only sketch the proof by indicating the differences. The main difference is that in this lemma we have optimal i-dependent rigidity for all $1 \leq |i| \leq i_*$. Hence, we can give a better estimate on the first two terms in (4.295) as follows (recall that $N^{\omega_A} \leq i \leq \frac{i_*}{2}$)

$$|\zeta_i| \lesssim \frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}}} \sum_{|j|<\frac{3i_*}{4}} \frac{1}{(\overline{\gamma}_i - \overline{\gamma}_j)^2 |j|^{\frac{1}{4}}} \lesssim \frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{3}{4}}} \sum_{|j|<\frac{3i_*}{4}} \frac{|i|^{\frac{1}{2}} + |j|^{\frac{1}{2}}}{(|i| - |j|)^2 |j|^{\frac{1}{4}}} \lesssim \frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{1}{4}} N^{3\omega_\ell}}.$$

Compared with (4.299), the additional N^{ω_ℓ} factor in the denominator comes from the $|j|^{1/4}$ factor beforehand that is due to the optimal dependence of the rigidity on the index. Consequently, using the optimal rigidity in (4.60), we improve the denominator in the first term on the rhs. of (4.286) from $N^{2\omega_\ell}$ to $N^{3\omega_\ell}$ with respect Lemma 4.C.1.

Furthermore, by (4.60),

$$|A_{2,3}|, |A_{2,4}| \leq \frac{N^\xi}{N\eta_1}, \qquad \text{and} \qquad |A_{2,1}|, |A_{2,2}| \lesssim \frac{N^{\frac{1}{2}}\eta_1}{i_*^{\frac{1}{2}}} \lesssim N^{\frac{1}{4} - \frac{C_*\omega_1}{2}} \eta_1,$$

since $i_* = N^{\frac{1}{2} + C_*\omega_1}$, hence, choosing $\eta_1 = N^{-\frac{5}{8}}$, we conclude that

$$|A_1| + |A_2| \lesssim \frac{N^\xi N^{\frac{\omega_1}{6}}}{N^{\frac{1}{4}} N^{3\omega_\ell}} + \frac{N^\xi}{N^{\frac{3}{8}}}.$$

All other estimates follow exactly in the same way of the proof of Lemma 4.C.1. This concludes the proof of Lemma 4.7.2. □

4.C.2 Short-long approximation: Small minimum.

In this subsection we estimate the difference of the solution of the DBM $\widetilde{z}(t, \alpha)$ defined by (4.230) and its short range approximation $\widehat{z}(t, \alpha)$ defined by (4.232)-(4.235) for $t_* \leq t \leq 2t_*$.

Lemma 4.C.2. *Under the same assumption of Lemma 4.C.1 and assuming that the rigidity bounds* (4.284) *and* (4.285) *hold for the* $\widetilde{z}(t, \alpha)$ *dynamics* (4.230) *for all* $t_* \leq t \leq 2t_*$, *we conclude that*

$$\sup_{1 \leq |i| \leq N} \sup_{t_* \leq t \leq 2t_*} |\widetilde{z}_i(t, \alpha) - \widehat{z}_i(t, \alpha)| \lesssim \frac{N^a N^{C\omega_1}}{N^{\frac{3}{4}}} \left(\frac{1}{N^{2\omega_\ell}} + \frac{1}{N^{\frac{2a}{5}} i_*^{\frac{1}{5}}} + \frac{N^{a_0}}{N^a i_*^{\frac{1}{2}}} + \frac{1}{N^a N^{\frac{1}{24}}} \right), \tag{4.317}$$

with very high probability, for any $\alpha \in [0, 1]$.

Proof. The proof of this lemma is similar to the proof of Lemma 4.C.1, but some estimates for the semicircular flow are slightly different mainly because in this lemma the $\widetilde{z}_i(t, \alpha)$ are shifted by $\overline{\mathfrak{m}}_t$ instead of $\overline{\mathfrak{e}}_t^+$. Hence, we will skip some details in this proof, describing carefully only the estimates that are different respect to Lemma 4.C.1.

Let $w_i := \widehat{z}_i - \widetilde{z}_i$, hence w is a solution of

$$\partial_t = \mathcal{B}_1 w + \mathcal{V}_1 w + \zeta,$$

where $\mathcal{B}_1$ and $\mathcal{V}_1$ are defined as in (4.288)-(4.290) substituting $\bar{\mathfrak{e}}_t^+$ with $\overline{\mathfrak{m}}_t$.

Without loss of generality we assume that $\mathcal{V}_1 \leq 0$ for all $t_* \leq t \leq T$ (see (4.291) in the proof of Lemma 4.C.1 but now we have $t_* \leq t \leq 2t_*$ in the definition of the stopping time). This implies that $\mathcal{U}^{\mathcal{B}_1+\mathcal{V}_1}$ is a contraction semigroup and so in order to prove (4.317) it is enough to estimate

$$\sup_{t_* \leq t \leq T} \|\zeta(s)\|_\infty.$$

At the end, exactly as at the end of the proof of Lemma 4.C.1, by continuity of the paths, we can easily establish $T = 2t_*$ for the stopping time.

The error term ζ is given by $\zeta_i = 0$ for $|i| > \frac{i_*}{2}$, then ζ_i for $1 \leq |i| \leq N^{\omega_A}$ is defined as

$$\zeta_i = \int_{\mathcal{I}_{y,i}(t)^c} \frac{\rho_{y,t}(E + \widetilde{\mathfrak{m}}_{y,t})}{\widetilde{z}_i - E} \,\mathrm{d}E - \frac{1}{N} \sum_{j}^{\mathcal{A}^c,(i)} \frac{1}{\widetilde{z}_i - \widetilde{z}_j} + \Psi_\alpha(t) + \frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\mathfrak{m}}_{y,t}, \tag{4.318}$$

with $\Psi_\alpha(t)$ defined in (4.228), and for $N^{\omega_A} < |i| \leq \frac{i_*}{2}$ as

$$\zeta_i = \int_{\mathcal{I}_{z,i}(t)^c \cap \mathcal{J}_z(t)} \frac{\overline{\rho}_t(E + \overline{\mathfrak{m}}_t)}{\widetilde{z}_i - E} \,\mathrm{d}E - \frac{1}{N} \sum_{|j|<\frac{3i_*}{4}}^{\mathcal{A}^c,(i)} \frac{1}{\widetilde{z}_i - \widetilde{z}_j}. \tag{4.319}$$

We start to estimate the error term for $N^{\omega_A} < |i| \leq \frac{i_*}{2}$. A similar computation as the one leading to (4.300) in Lemma 4.C.1, using (4.285), we conclude that

$$|\zeta_i| = \left| \int_{\mathcal{I}_{i,z}^c(t) \cap \mathcal{J}_z(t)} \frac{\overline{\rho}_t(E + \overline{\mathfrak{m}}_t)}{\widetilde{z}_i - E} \,\mathrm{d}E - \frac{1}{N} \sum_{|j|<\frac{3i_*}{4}}^{\mathcal{A}^c,(i)} \frac{1}{\widetilde{z}_i - \widetilde{z}_j} \right| \lesssim \frac{N^a}{N^{\frac{1}{4}} N^{2\omega_\ell}}, \quad N^{\omega_A} < |i| \leq \frac{i_*}{2}. \tag{4.320}$$

Next, we proceed with the bound for ζ_i for $1 \leq |i| \leq N^{\omega_A}$. We rewrite ζ_i as

$$\begin{aligned} \zeta_i = &\left(\int_{\mathcal{I}_{i,z}^c(t)} \frac{\overline{\rho}_t(E + \overline{\mathfrak{m}}_t)}{\widetilde{z}_i - E} \,\mathrm{d}E - \frac{1}{N} \sum_{j}^{\mathcal{A}^c,(i)} \frac{1}{\widetilde{z}_i - \widetilde{z}_j} \right) \\ &+ \Re[m_{y,t}(\widetilde{z}_i + \widetilde{\mathfrak{m}}_{y,t})] + \frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\mathfrak{m}}_{y,t} + \Psi_\alpha(t) - \Re[\overline{m}_t(\widetilde{z}_i + \overline{\mathfrak{m}}_t)] \\ &+ \left(\int_{\mathcal{I}_{z,i}(t)} \frac{\overline{\rho}_t(E + \overline{\mathfrak{m}}_t)}{\widetilde{z}_i - E} \,\mathrm{d}E - \int_{\mathcal{I}_{y,i}(t)} \frac{\rho_{y,t}(E + \widetilde{\mathfrak{m}}_{y,t})}{\widetilde{z}_i - E} \,\mathrm{d}E \right) =: (A_1 + A_2) + A_3 + A_4. \end{aligned} \tag{4.321}$$

where $(A_1 + A_2)$ indicates that for the actual estimates we split the first line in (4.321) into two terms as in (4.301). By similar computations as in Lemma 4.C.1, see (4.302) and (4.306), we conclude that

$$|A_1| + |A_2| \lesssim \frac{N^a}{N^{\frac{1}{4}} N^{2\omega_\ell}} + \frac{N^{\frac{3a}{5}}}{N^{\frac{1}{4}} i_*^{\frac{1}{5}}} + \frac{N^{a_0}}{i_*^{\frac{1}{2}} N^{\frac{1}{4}}}. \tag{4.322}$$

By (4.31b), (4.31d), (4.39b) and the definition of $\Psi_\alpha(t)$ in (4.228) it follows that

$$\begin{aligned} |A_3| &\lesssim |\Re[m_{y,t}(\widetilde{z}_i + \widetilde{\mathfrak{m}}_{y,t}) - m_{y,t}(\widetilde{\mathfrak{m}}_{y,t})] - \Re[\overline{m}_t(\overline{\mathfrak{m}}_t) - \overline{m}_t(\widetilde{z}_i + \overline{\mathfrak{m}}_t)]| + \frac{N^{\omega_1}}{N} \\ &\lesssim \left(\frac{N^{\frac{\omega_A}{4}} N^{\frac{\omega_1}{4}}}{N^{\frac{1}{4}} N^{\frac{1}{8}}} + \frac{N^{\frac{3\omega_1}{4}}}{N^{\frac{3}{8}}} + \frac{N^{\frac{\omega_A}{2}}}{N^{\frac{1}{2}}} \right) |\log|\widehat{\gamma}_i(t)|| + \frac{N^{\frac{7\omega_1}{12}}}{N^{\frac{7}{24}}} \lesssim \frac{N^{\frac{7\omega_1}{12}}}{N^{\frac{7}{24}}}. \end{aligned} \tag{4.323}$$

We proceed writing A_4 as

$$A_4 = \left(\int_{\mathcal{I}_{z,i}(t)} \frac{\overline{\rho}_t(E+\overline{\mathfrak{m}}_t)}{\widetilde{z}_i - E} \,\mathrm{d}E - \int_{\mathcal{I}_{z,i}(t)} \frac{\rho_{y,t}(E+\widetilde{\mathfrak{m}}_{y,t})}{\widetilde{z}_i - E} \,\mathrm{d}E \right) + \left(\int_{\mathcal{I}_{z,i}(t)} \frac{\rho_{y,t}(E+\widetilde{\mathfrak{m}}_{y,t})}{\widetilde{z}_i - E} \,\mathrm{d}E - \int_{\mathcal{I}_{y,i}(t)} \frac{\rho_{y,t}(E+\widetilde{\mathfrak{m}}_{y,t})}{\widetilde{z}_i - E} \,\mathrm{d}E \right) =: A_{4,1} + A_{4,2}. \tag{4.324}$$

We start with the estimate for $A_{4,2}$.

By (4.36) we have that

$$|\mathcal{I}_{z,i}(t) \Delta \mathcal{I}_{y,i}(t)| \lesssim \frac{N^\xi(\ell + |i|)}{N}, \tag{4.325}$$

where Δ is the symmetric difference. Note that this bound is somewhat better than the analogous (4.309) due to the better bound in (4.36) compared with (4.35). For $E \in \mathcal{I}_{z,i}(t)\Delta\mathcal{I}_{y,i}(t)$ we have that

$$\left|\frac{\rho_{y,t}(E+\overline{\mathfrak{m}}_t)}{\widetilde{z}_i - E}\right| \lesssim \frac{N^{\frac{1}{2}}(\ell^2 + |i|^{\frac{1}{2}})}{\ell^3 + |i|^{\frac{3}{4}}}, \tag{4.326}$$

and so

$$|A_{4,2}| \lesssim \frac{N^{\frac{3\omega_A}{4}}}{N^{\frac{1}{2}}} \tag{4.327}$$

with very high probability.

To estimate the integral in $A_{4,1}$, we combine (4.22d) and (4.31b) to obtain that

$$|\overline{\rho}_t(\overline{\mathfrak{m}}_t + E) - \rho_{y,t}(\widetilde{\mathfrak{m}}_{y,t} + E)| \leq |\rho_{x,t}(\alpha\mathfrak{m}_{x,t} + (1-\alpha)\mathfrak{m}_{y,t} + E) - \rho_{y,t}(\mathfrak{m}_{y,t} + E)| + (t - t_*)^{\frac{7}{12}}. \tag{4.328}$$

Proceeding similarly to the estimate of $|A_{4,1}|$ at the end of the proof of Lemma 4.C.1, we conclude that

$$|A_{4,1}| \lesssim \left(\frac{N^\xi(\ell^2 + |i|^{\frac{1}{2}})}{N^{\frac{1}{2}}} + \frac{N^{\frac{7\omega_1}{12}}}{N^{\frac{7}{24}}} \right) \int_{\mathcal{I}_{z,i}(t) \cap \{|E - \widetilde{z}_i| > N^{-60}\}} \frac{1}{|\widetilde{z}_i - E|} \,\mathrm{d}E + \left| \int_{|E-\widetilde{z}_i| \leq N^{-60}} \frac{\overline{\rho}_t(E+\overline{\mathfrak{m}}_t) - \rho_{y,t}(E+\widetilde{\mathfrak{m}}_{y,t})}{\widetilde{z}_i - E} \,\mathrm{d}E \right|. \tag{4.329}$$

Furthermore, similarly to the estimate in the singular integral in (4.315), but substituting $\overline{\mathfrak{e}}_t^+$ and $\mathfrak{e}_{y,t}^+$ by $\overline{\mathfrak{m}}_t$ and $\widetilde{\mathfrak{m}}_{y,t}$ respectively, we conclude that that the last term in (4.329) is bounded by N^{-20}. Therefore,

$$|A_{41}| \lesssim \frac{N^\xi(\ell^2 + |i|^{\frac{1}{2}})}{N^{\frac{1}{2}}} + \frac{N^{\frac{7\omega_1}{12}}}{N^{\frac{7}{24}}} \lesssim \frac{N^{\frac{7\omega_1}{12}}}{N^{\frac{7}{24}}}, \tag{4.330}$$

for any $|i| \leq N^{\omega_A}$. Collecting (4.322), (4.323), (4.327) and (4.330) completes the proof of Lemma 4.C.2. □

4.D Sobolev-type inequality

The proof of the Sobolev-type inequality in the cusp case is essentially identical to that in the edge case presented in Appendix B of [41]; only the exponents need adjustment to the cusp scaling. We give some details for completeness.

Proof of Lemma 4.7.5. We will prove only the first inequality in (4.214). The proof for the second one is exactly the same. We start by proving a continuous version of (4.214) and then we will conclude the proof by linear interpolation. We claim that for any small η there exists a constant $c_\eta > 0$ such that for any real function $f \in L^p(\mathbf{R}_+)$ we have that

$$\int_0^{+\infty}\int_0^{+\infty} \frac{(f(x)-f(y))^2}{|x^{\frac{3}{4}}-y^{\frac{3}{4}}|^{2-\eta}}\,\mathrm{d}x\,\mathrm{d}y \geq c_\eta \left(\int_0^{+\infty} |f(x)|^p\,\mathrm{d}x\right)^{\frac{2}{p}}. \tag{4.331}$$

We recall the representation formula for fractional powers of the Laplacian: for any $0<\alpha<2$ and for any function $f \in L^p(\mathbf{R})$ for some $p \in [1,\infty)$ we have

$$\langle f, |p|^\alpha f\rangle = C(\alpha)\int_{\mathbf{R}}\int_{\mathbf{R}} \frac{(f(x)-f(y))^2}{|x-y|^{1+\alpha}}\,\mathrm{d}x\,\mathrm{d}y, \tag{4.332}$$

with some explicit constant $C(\alpha)$, where $|p| := \sqrt{-\Delta}$.

Since for $0<x<y$ we have that

$$y^{\frac{3}{4}} - x^{\frac{3}{4}} = \frac{4}{3}\int_x^y s^{-\frac{1}{4}}\,\mathrm{d}s \leq C(y-x)(xy)^{-\frac{1}{8}},$$

in order to prove (4.331) it is enough to show that

$$\int_0^{+\infty}\int_0^{+\infty} \frac{(f(x)-f(y))^2}{|x-y|^{2-\eta}}(xy)^q\,\mathrm{d}x\,\mathrm{d}y \geq c_\eta \left(\int_0^{+\infty} |f(x)|^p\,\mathrm{d}x\right)^{\frac{2}{p}}, \tag{4.333}$$

where $q := \frac{1}{4} - \frac{\eta}{8}$ and $p := \frac{8}{2+3\eta}$. Let $\tilde{f}(x)$ be the symmetric extension of f to the whole real line, i.e. $\tilde{f}(x) := f(x)$ for $x>0$ and $\tilde{f}(x) := f(-x)$ for $x<0$. Then, by a simple calculation we have

$$4\int_0^{+\infty}\int_0^{+\infty} \frac{(f(x)-f(y))^2}{|x-y|^{2-\eta}}(xy)^q\,\mathrm{d}x\,\mathrm{d}y \geq \int_{\mathbf{R}}\int_{\mathbf{R}} \frac{(\tilde{f}(x)-\tilde{f}(y))^2}{|x-y|^{2-\eta}}|xy|^q\,\mathrm{d}x\,\mathrm{d}y.$$

Introducing $\phi(x) := |x|^q$ and dropping the tilde for f the estimate in (4.333) would follow from

$$\int_{\mathbf{R}}\int_{\mathbf{R}} \frac{(f(x)-f(y))^2}{|x-y|^{2-\eta}}\phi(x)\phi(y)\,\mathrm{d}x\,\mathrm{d}y \geq c_\eta' \left(\int_{\mathbf{R}} |f(x)|^p\,\mathrm{d}x\right)^{\frac{2}{p}}. \tag{4.334}$$

By the same computation as in the proof of Proposition 10.5 in [41] we conclude that

$$\int_{\mathbf{R}}\int_{\mathbf{R}} \frac{(f(x)-f(y))^2}{|x-y|^{2-\eta}}\phi(x)\phi(y)\,\mathrm{d}x\,\mathrm{d}y = \langle \phi f, |p|^{1-\eta}\phi f\rangle + C_0(\eta)\int_{\mathbf{R}} \frac{|\phi(x)f(x)|^2}{|x|^{1-\eta}}\,\mathrm{d}x$$

with some $C_0(\eta)>0$, hence for the proof of (4.334) it is enough to show that

$$\langle \phi f, |p|^{1-\eta}\phi f\rangle \geq c_\eta \left(\int_{\mathbf{R}} |f|^p\right)^{\frac{2}{p}}.$$

Let $g := |p|^{\frac{1}{2}(1-\eta)}|x|^q f$, we need to prove that

$$\|g\|_2 \geq c_\eta \| |x|^{-q} |p|^{-\frac{1}{2}(1-\eta)} g\|_p.$$

By the n-dimensional Hardy-Littlewood-Sobolev inequality in [189] we have that

$$\left\| |x|^{-q} \int |x-y|^{-a} g(y) \, \mathrm{d}y \right\|_p \leq C \|g\|_r,$$

where $\frac{1}{r} + \frac{a+q}{n} = 1 + \frac{1}{p}$, $0 \leq q < \frac{n}{p}$ and $0 < a < n$. In our case $a = \frac{1+\eta}{2}$, $r = 2$, $n = 1$ and all the conditions are satisfied if we take $0 < \eta < 1$. This completes the proof of (4.331).

Next, in order to prove (4.214), we proceed by linear interpolation as in Proposition B.2 in [91]. Given $u : \mathbf{Z} \to \mathbf{R}$, let $\psi : \mathbf{R} \to \mathbf{R}$ be its linear interpolation, i.e. $\psi(i) := u_i$ for $i \in \mathbf{Z}$ and

$$\psi(x) := u_i + (u_{i+1} - u_i)(x - i) = u_{i+1} - (u_{i+1} - u_i)(i + 1 - x), \tag{4.335}$$

for $x \in [i, i+1]$. It is easy to see that for each $p \in [2, +\infty]$ (i.e. $\eta \leq 2/3$), there exists a constant C_p such that

$$C_p^{-1} \|\psi\|_{L^p(\mathbf{R})} \leq \|u\|_{L^p(\mathbf{Z})} \leq C_p \|\psi\|_{L^p(\mathbf{R})}. \tag{4.336}$$

In order to prove (4.214) we claim that

$$\int_0^{+\infty} \int_0^{+\infty} \frac{|\psi(x) - \psi(y)|^2}{|x^{\frac{3}{4}} - y^{\frac{3}{4}}|^{2-\eta}} \, \mathrm{d}x \, \mathrm{d}y \leq c_\eta \sum_{i \neq j \in \mathbf{Z}_+} \frac{(u_i - u_j)^2}{|i^{\frac{3}{4}} - j^{\frac{3}{4}}|^{2-\eta}}, \tag{4.337}$$

for some constant $c_\eta > 0$. Indeed, combining (4.336) and (4.337) with (4.331) we conclude (4.214). Finally, the proof of (4.337) is a simple exercise along the lines of the proof of Proposition B.2 in [91]. □

4.E Heat-kernel estimates

The proof of the heat kernel estimates relies on the Nash method. In the edge scaling regime a similar bound was proven in [41] for a compact interval, extended to non-compact interval but with compactly supported initial data w_0 in [131]. Here we closely follow the latter proof, adjusted to the cusp regime, where interactions on both sides of the cusp play a role unlike in the edge regime.

Proof of Lemma 4.7.6. We start proving (4.215), then (4.216) follows by (4.215) by duality. Without loss of generality we assume $\|w_0\|_1 = 1$ and that

$$\|w(\tilde{s})\|_p \geq N^{-100} \tag{4.338}$$

for each $s \leq \tilde{s} \leq t$, where $w(\tilde{s}) = \mathcal{U}^{\mathcal{L}}(s, \tilde{s}) w_0$. Otherwise, by ℓ^p-contraction we had $\|w(\tilde{s})\|_p \leq N^{-100}$ implying (4.215) directly.

In the following we use the convention $w := w(\tilde{s})$ if there is no confusion. By (4.214), we have that

$$\|w\|_p^2 \lesssim \sum_{\substack{i,j \geq 1 \\ i \neq j}} \frac{(w_i - w_j)^2}{|i^{\frac{3}{4}} - j^{\frac{3}{4}}|^{2-\eta}} + \sum_{\substack{i,j \leq -1 \\ i \neq j}} \frac{(w_i - w_j)^2}{||i|^{\frac{3}{4}} - |j|^{\frac{3}{4}}|^{2-\eta}}.$$

First we assume that both i and j are positive. Let $\delta_4 < \delta_2 < \delta_3 < \frac{\delta_1}{2}$. We start with the following estimate

$$\sum_{\substack{i,j\geq 1\\ i\neq j}} \frac{(w_i - w_j)^2}{|i^{\frac{3}{4}} - j^{\frac{3}{4}}|^{2-\eta}} \lesssim \sum_{\substack{(i,j)\in\mathcal{A}\\ i,j\geq 1}} \frac{(w_i - w_j)^2}{|i^{\frac{3}{4}} - j^{\frac{3}{4}}|^{2-\eta}} + \sum_{i\geq 1} \sum_{j\geq 1}^{\mathcal{A}^c,(i)} \frac{w_i^2}{|i^{\frac{3}{4}} - j^{\frac{3}{4}}|^{2-\eta}}. \tag{4.339}$$

We proceed by writing

$$\sum_{\substack{(i,j)\in\mathcal{A}\\ i,j\geq 1}} \frac{(w_i - w_j)^2}{|i^{\frac{3}{4}} - j^{\frac{3}{4}}|^{2-\eta}} \lesssim \sum_{\substack{(i,j)\in\mathcal{A}:\, i,j\geq 1\\ i \text{ or } j\leq \ell^4 N^{\delta_2}}} \frac{(w_i - w_j)^2}{|i^{\frac{3}{4}} - j^{\frac{3}{4}}|^{2-\eta}} + \sum_{\substack{(i,j)\in\mathcal{A}\\ i,j\geq \ell^4 N^{\delta_2}}} \frac{(w_i - w_j)^2}{|i^{\frac{3}{4}} - j^{\frac{3}{4}}|^{2-\eta}}. \tag{4.340}$$

By Lemma 4.B.3 we have that

$$\sum_{\substack{(i,j)\in\mathcal{A}\\ i,j\geq \ell^4 N^{\delta_2}}} \frac{(w_i - w_j)^2}{|i^{\frac{3}{4}} - j^{\frac{3}{4}}|^{2-\eta}} \lesssim N^{-200}, \tag{4.341}$$

since $i \geq \ell^4 N^{\delta_2}$ and $|(w_0)_j| \leq N^{-100}$ for $j \geq \ell^4 N^{\delta_4}$ by our hypotheses. Indeed, for $i \geq \ell^4 N^{\delta_2}$, we have that

$$w_i = \left(\mathcal{U}^{\mathcal{L}}(s,\tilde{s})w_0\right)_i = \sum_{j=-N}^{N} \mathcal{U}_{ij}^{\mathcal{L}}(w_0)_j = \sum_{j=-\ell^4 N^{\delta_4}}^{\ell^4 N^{\delta_4}} \mathcal{U}_{ij}^{\mathcal{L}}(w_0)_j + N^{-100} \lesssim N^{-100}, \tag{4.342}$$

with very high probability. If $(i,j) \in \mathcal{A}$, $i,j \geq 1$ and i or j are smaller than $\ell^4 N^{\delta_2}$ then both i and j are smaller than $\ell^4 N^{\delta_3}$. Hence, for such i and j, by (4.213), we have that

$$|\widehat{z}_i(t,\alpha) - \widehat{z}_j(t,\alpha)| \lesssim \frac{N^{\frac{\omega_1}{6}} |i^{\frac{3}{4}} - j^{\frac{3}{4}}|}{N^{\frac{3}{4}}}, \tag{4.343}$$

for any fixed $\alpha \in [0,1]$ and for all $0 \leq t \leq t_*$, where $\widehat{z}_i(t,\alpha)$ is defined by (4.163)-(4.164).

If i and j are both negative the estimates in (4.339)-(4.343) follow in the same way.

In the following of the proof $\mathcal{B}$, $\mathcal{B}_{ij}$ and $\mathcal{V}_i$ are defined in (4.163)-(4.164). By (4.343) it follows that

$$\begin{aligned} &\sum_{\substack{(i,j)\in\mathcal{A}:\, i,j\geq 1\\ i \text{ or } j\leq \ell^4 N^{\delta_2}}} \frac{(w_i - w_j)^2}{|i^{\frac{3}{4}} - j^{\frac{3}{4}}|^{2-\eta}} + \sum_{\substack{(i,j)\in\mathcal{A}:\, i,j\leq -1\\ i \text{ or } j\geq -\ell^4 N^{\delta_2}}} \frac{(w_i - w_j)^2}{|i^{\frac{3}{4}} - j^{\frac{3}{4}}|^{2-\eta}} \\ &\quad \lesssim -N^{-\frac{1}{2}} N^{\frac{\omega_1}{3}+C\eta} \sum_{(i,j)\in\mathcal{A}} \mathcal{B}_{ij}(w_i - w_j)^2 = -2N^{-\frac{1}{2}} N^{\frac{\omega_1}{3}+C\eta} \langle w, \mathcal{B}w\rangle. \end{aligned} \tag{4.344}$$

Furthermore, since $1 \leq |i| \leq \ell^4 N^{\delta_3}$, we have that

$$\sum_{j}^{\mathcal{A}^c,(i)} \frac{1}{||i|^{\frac{3}{4}} - |j|^{\frac{3}{4}}|^{2-\eta}} \lesssim \frac{N^{\frac{\omega_1}{3}+C\eta}}{N^{\frac{3}{2}}} \sum_{j}^{\mathcal{A}^c,(i)} \frac{1}{(\widehat{z}_i - \widehat{z}_j)^2}. \tag{4.345}$$

By the rigidity (4.210), (4.211) and (4.213), we can replace $\widehat{z}_j$ by $\overline{\gamma}_j$ in the sum on the rhs. of (4.345) and so approximate it by an integral, then using that $\overline{\rho}_t(E) \lesssim \rho_{y,t}(E)$ in the cusp regime, i.e. $|E| \le \delta_*$, with δ_* defined in Definition 4.4.1, we conclude that

$$\frac{1}{N}\sum_j^{\mathcal{A}^c,(i)} \frac{1}{(\widehat{z}_i(t)-\widehat{z}_j(t))^2} \lesssim \int_{I_{i,y}(t)^c} \frac{\rho_{y,t}(E+\mathfrak{e}^+_{y,t})}{(\widehat{z}_i(t)-E)^2}\,\mathrm{d}E = -\mathcal{V}_i. \tag{4.346}$$

Hence, by (4.346), we conclude that

$$\begin{aligned}\sum_i \sum_j^{\mathcal{A}^c,(i)} \frac{w_i^2}{||i|^{\frac34}-|j|^{\frac34}|^{2-\eta}} &\lesssim \sum_{1\le |i|\le \ell^4 N^{\delta_3}} \sum_j^{\mathcal{A}^c,(i)} \frac{w_i^2}{||i|^{\frac34}-|j|^{\frac34}|^{2-\eta}} + N^{-200}\\ &\lesssim -N^{-\frac12}N^{\frac{\omega_1}{3}+C\eta}\sum_{|i|\le \ell^4 N^{\delta_3}} w_i^2\mathcal{V}_i + N^{-200}\\ &\lesssim -N^{-\frac12}N^{\frac{\omega_1}{3}+C\eta}\langle w,\mathcal{V}w\rangle + N^{-200}.\end{aligned} \tag{4.347}$$

Note that in the first inequality of (4.347) we used (4.342).

Summarizing (4.341), (4.344) and (4.347) and rewriting N^{-200} into an ℓ^p-norm using (4.338), we obtain

$$\|w\|_p^2 \le -N^{-\frac12}N^{\frac{\omega_1}{3}+C\eta}\langle w,\mathcal{L}w\rangle + \frac{1}{10}\|w\|_p^2.$$

Hence, using Hölder inequality, we have that

$$\begin{aligned}\partial_t\|w\|_2^2 = \langle w,\mathcal{L}w\rangle &\le -c_\eta N^{\frac12}N^{-\frac{\omega_1}{3}-C\eta}\|w\|_p^2\\ &\le -c_\eta N^{\frac12}N^{-\frac{\omega_1}{3}-C\eta}\|w\|_2^{\frac{6-3\eta}{2}}\|w\|_1^{-\frac{2-3\eta}{2}}\\ &\le -c_\eta N^{\frac12}N^{-\frac{\omega_1}{3}-C\eta}\|w\|_2^{\frac{6-3\eta}{2}}\|w_0\|_1^{-\frac{2-3\eta}{2}}.\end{aligned} \tag{4.348}$$

In the last inequality of (4.348) we used the ℓ^1-contraction of $\mathcal{U}^{\mathcal{L}}$. Integrating (4.348) back in time, it easily follows that

$$\|\mathcal{U}^{\mathcal{L}}(s,t)w_0\|_2 \le \left(\frac{N^{C\eta+\frac{\omega_1}{3}}}{c_\eta N^{\frac12}(t-s)}\right)^{1-3\eta}\|w_0\|_1, \tag{4.349}$$

proving (4.215). The same bound also holds for the transpose operator $(\mathcal{U}^{\mathcal{L}})^T$.

In order to prove (4.216) we follow Lemma 3.11 of [131]. Let $\chi(i) := \mathbf{1}_{\{|i|\le \ell^4 N^{\delta_5}\}}$, with $\delta_4 < \delta_5 < \frac{\delta_1}{2}$, and $v \in \mathbf{R}^{2N}$. Then, we have that

$$\langle \mathcal{U}^{\mathcal{L}}(0,t)w_0, v\rangle = \langle w_0, (\mathcal{U}^{\mathcal{L}})^T\chi v\rangle + \langle w_0, (\mathcal{U}^{\mathcal{L}})^T(1-\chi)v\rangle.$$

By Lemma 4.B.3 we have that

$$|\langle w_0, (\mathcal{U}^{\mathcal{L}})^T(1-\chi)v\rangle| \le N^{-100}\|w_0\|_2\|v\|_1. \tag{4.350}$$

By (4.215) and Cauchy-Schwarz inequality we have that

$$|\langle w_0, (\mathcal{U}^{\mathcal{L}})^T\chi v\rangle| \le \|w_0\|_2\|(\mathcal{U}^{\mathcal{L}})^T\chi v\|_2 \le \|w_0\|_2\left(\frac{N^{C\eta+\frac{\omega_1}{3}}}{c_\eta N^{\frac12}t}\right)^{1-3\eta}\|v\|_1. \tag{4.351}$$

Hence, combining (4.350) and (4.351), we conclude that

$$\|\mathcal{U}^{\mathcal{L}}(0,t)w_0\|_\infty \leq \left(\frac{N^{C\eta+\frac{\omega_1}{3}}}{c_\eta N^{\frac{1}{2}}t}\right)^{1-3\eta} \|w_0\|_2, \tag{4.352}$$

and so, by (4.349), that

$$\begin{aligned} \|\mathcal{U}^{\mathcal{L}}(0,t)w_0\|_\infty = \|\mathcal{U}^{\mathcal{L}}(t/2,t)\mathcal{U}^{\mathcal{L}}(0,t/2)w_0\|_\infty &\lesssim \left(\frac{N^{C\eta+\frac{\omega_1}{3}}}{c_\eta N^{\frac{1}{2}}t}\right)^{1-3\eta} \|\mathcal{U}^{\mathcal{L}}(0,t/2)w_0\|_2 \\ &\lesssim \left(\frac{N^{C\eta+\frac{\omega_1}{3}}}{c_\eta N^{\frac{1}{2}}t}\right)^{2(1-3\eta)} \|w_0\|_1, \end{aligned} \tag{4.353}$$

where in the first inequality we used that $\mathcal{U}^{\mathcal{L}}(0,t/2)w_0$ satisfies the hypothesis of Lemma 4.7.6, since $|(\mathcal{U}^{\mathcal{L}}(0,t/2)w_0)_i| \leq N^{-100}$ for $|i| \geq \ell^4 N^{2\delta_4}$ by the finite speed estimate of Lemma 4.B.3. Combining (4.352) and (4.353) then (4.216) follows by interpolation. □

Edge Universality for non-Hermitian Random Matrices 5

We consider large non-Hermitian real or complex random matrices X with independent, identically distributed centred entries. We prove that their local eigenvalue statistics near the spectral edge, the unit circle, coincide with those of the Ginibre ensemble, i.e. when the matrix elements of X are Gaussian. This result is the non-Hermitian counterpart of the universality of the Tracy-Widom distribution at the spectral edges of the Wigner ensemble.

Published as G. Cipolloni et al., *Edge universality for non-Hermitian random matrices*, Probability Theory and Related Fields, 1–28 (2020).

5.1 Introduction

Following Wigner's motivation from physics, most universality results on the local eigenvalue statistics for large random matrices concern the Hermitian case. In particular, the celebrated Wigner-Dyson statistics in the bulk spectrum [146], the Tracy-Widom statistics [202, 203] at the spectral edge and the Pearcey statistics [157, 204] at the possible cusps of the eigenvalue density profile all describe eigenvalue statistics of a large Hermitian random matrix. In the last decade there has been a spectacular progress in verifying Wigner's original vision, formalized as the Wigner-Dyson-Mehta conjecture, for Hermitian ensembles with increasing generality, see e.g. [6, 45, 84–86, 88, 118, 130, 134, 136, 155, 181, 193] for the bulk, [15, 41, 43, 115, 131, 133, 156, 186, 194] for the edge and more recently [57, 83, 109] at the cusps.

Much less is known about the spectral universality for non-Hermitian models. In the simplest case of the Ginibre ensemble, i.e. random matrices with i.i.d. standard Gaussian entries without any symmetry condition, explicit formulas for all correlation functions have been computed first for the complex case [102] and later for the more complicated real case [35, 121, 183] (with special cases solved earlier [76, 77, 137]). Beyond the explicitly computable Ginibre case only the method of *four moment matching* by Tao and Vu has been available. Their main universality result in [195] states that the local correlation functions of the eigenvalues of a random matrix X with i.i.d. matrix elements coincide with those of the Ginibre ensemble as long as the first four moments of the common distribution of the entries of X (almost) match the first four moments of the standard Gaussian. This result holds for both real and complex cases as well as throughout the spectrum, including the edge regime.

In the current paper we prove the edge universality for any $n \times n$ random matrix X with centred i.i.d. entries in the edge regime, in particular we remove the four moment matching condition from [195]. More precisely, under the normalization $\mathbf{E}\,|x_{ab}|^2 = \frac{1}{n}$, the spectrum of X converges to the unit disc with a uniform spectral density according to the *circular law* [18, 20, 33, 101, 103, 191]. The typical distance between nearest eigenvalues is of order $n^{-1/2}$. We pick a reference point z on the boundary of the limiting spectrum, $|z| = 1$, and rescale correlation functions by a factor of $n^{-1/2}$ to detect the correlation of individual eigenvalues. We show that these rescaled correlation functions converge to those of the Ginibre ensemble as $n \to \infty$. This result is the non-Hermitian analogue of the Tracy-Widom edge universality in the Hermitian case. A similar result is expected to hold in the bulk regime, i.e. for any reference point $|z| < 1$, but our method is currently restricted to the edge.

Investigating spectral statistics of non-Hermitian random matrices is considerably more challenging than Hermitian ones. We give two fundamental reasons for this: the first one is already present in the proof of the circular law on the global scale. The second one is specific to the most powerful existing method to prove universality of eigenvalue fluctuations.

The first issue a general one; it is well known that non-Hermitian, especially non-normal spectral analysis is difficult because, unlike in the Hermitian case, the resolvent $(X-z)^{-1}$ of a non-normal matrix is not effective to study eigenvalues near z. Indeed, $(X-z)^{-1}$ can be very large even if z is away from the spectrum, a fact that is closely related to the instability of the non-Hermitian eigenvalues under perturbations. The only useful expression to grasp non-Hermitian eigenvalues is Girko's celebrated formula, see (5.14) later, expressing linear statistics of eigenvalues of X in terms of the log-determinant of the symmetrized matrix

$$H^z = \begin{pmatrix} 0 & X-z \\ X^* - \overline{z} & 0 \end{pmatrix}. \tag{5.1}$$

Girko's formula is much more subtle and harder to analyse than the analogous expression for the Hermitian case involving the boundary value of the resolvent on the real line. In particular, it requires a good lower bound on the smallest singular value of $X-z$, a notorious difficulty behind the proof of the circular law. Furthermore, any conceivable universality proof would rely on a local version of the circular law as an a priori control. Local laws on optimal scale assert that the eigenvalue density on a scale $n^{-1/2+\epsilon}$ is deterministic with high probability, i.e. it is a law of large number type result and is not sufficiently refined to detect correlations of individual eigenvalues. The proof of the local circular law requires a careful analysis of H^z that has an additional structural instability due to its block symmetry. A specific estimate, tailored to Girko's formula, on the trace of the resolvent of $(H^z)^2$ was the main ingredient behind the proof of the local circular law on optimal scale [44, 46, 213], see also [195] under three moment matching condition. Very recently the optimal local circular law was even proven for ensembles with inhomogeneous variance profiles in the bulk [11] and at the edge [13], the latter result also gives an optimal control on the spectral radius. An optimal local law for H^z in the edge regime previously had not been available, even in the i.i.d. case.

The second major obstacle to prove universality of fluctuations of non-Hermitian eigenvalues is the lack of a good analogue of the Dyson Brownian motion. The essential ingredient behind the strongest universality results in the Hermitian case is the Dyson Brownian motion (DBM) [74], a system of coupled stochastic differential equations (SDE) that the eigenvalues of a natural stochastic flow of random matrices satisfy, see [90] for a pedagogical summary. The corresponding SDE in the non-Hermitian case involves not only eigenvalues but overlaps of eigenvectors as well, see e.g. [38, Appendix A]. Since overlaps themselves have strong correlation whose proofs are highly nontrivial even in the Ginibre case [38, 99], the analysis of this SDE is currently beyond reach.

Our proof of the edge universality circumvents DBM and it has two key ingredients. The first main input is an optimal local law for the resolvent of H^z both in *isotropic* and *averaged* sense, see (5.13) later, that allows for a concise and transparent comparison of the joint distribution of several resolvents of H^z with their Gaussian counterparts by following their evolution under the natural Ornstein-Uhlenbeck (OU). We are able to control this flow for a long time, similarly to an earlier proof of the Tracy-Widom law at the spectral edge of a Hermitian ensemble [135]. Note that the density of eigenvalues of H^z develops a cusp as $|z|$ passes through 1, the spectral radius of X. The optimal local law for very general Hermitian ensembles in the cusp regime has recently been proven [83], strengthening the non-optimal result in [6]. This optimality was essential in the proof of the universality of the Pearcey statistics for both the complex Hermitian [83] and real symmetric [57] matrices with a cusp in their density of states. The matrix H^z, however, does not satisfy the key *flatness* condition required [83] due its large zero blocks. A very delicate analysis of the underlying matrix Dyson equation was necessary to overcome the flatness condition and prove the optimal local law for H^z in [11, 13].

Our second key input is a lower tail estimate on the lowest singular value of $X - z$ when $|z| \approx 1$. A very mild regularity assumption on the distribution of the matrix elements of X, see (5.4) later, guarantees that there is no singular value below n^{-100}, say. Cruder bounds guarantee that there cannot be more than n^ϵ singular values below $n^{-3/4}$; note that this natural scaling reflects the cusp at zero in the density of states of H^z. Such information on the possible singular values in the regime $[n^{-100}, n^{-3/4}]$ is sufficient for the optimal local law since it is insensitive to n^ϵ-eigenvalues, but for universality every eigenvalue must be accounted for. We therefore need a stronger lower tail bound on the lowest eigenvalue λ_1 of $(X-z)(X-z)^*$. With supersymmetric methods we recently proved [61] a precise bound of the form

$$\mathbf{P}\Big(\lambda_1((X-z)(X-z)^*) \le \frac{x}{n^{3/2}}\Big) \lesssim \begin{cases} x + \sqrt{x}e^{-n(\Im z)^2}, & X \sim \mathrm{Gin}(\mathbf{R}) \\ x, & X \sim \mathrm{Gin}(\mathbf{C}), \end{cases} \tag{5.2}$$

modulo logarithmic corrections, for the Ginibre ensemble whenever $|z| = 1 + \mathcal{O}(n^{-1/2})$. Most importantly, (5.2) controls λ_1 on the optimal $n^{-3/2}$ scale and thus excluding singular values in the intermediate regime $[n^{-100}, n^{-3/4-\epsilon}]$ that was inaccessible with other methods. We extend this control to X with i.i.d. entries from the Ginibre ensemble with Green function comparison argument using again the optimal local law for H^z.

Notations and conventions

We introduce some notations we use throughout the paper. We write $\mathbf{H}$ for the upper half-plane $\mathbf{H} := \{z \in \mathbf{C} : \Im z > 0\}$, and for any $z \in \mathbf{C}$ we use the notation $\mathrm{d}z := 2^{-1}\mathrm{i}(\mathrm{d}z \wedge \mathrm{d}\overline{z})$ for the two dimensional volume form on $\mathbf{C}$. For any $2n \times 2n$ matrix A we use the notation $\langle A \rangle := (2n)^{-1} \operatorname{Tr} A$ to denote the normalized trace of A. For positive quantities f, g we write $f \lesssim g$ and $f \sim g$ if $f \leq Cg$ or $cg \leq f \leq Cg$, respectively, for some constants $c, C > 0$ which depends only on the constants appearing in (5.3). We denote vectors by bold-faced lower case Roman letters $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{C}^k$, for some $k \in \mathbf{N}$. Vector and matrix norms, $\|\boldsymbol{x}\|$ and $\|A\|$, indicate the usual Euclidean norm and the corresponding induced matrix norm. Moreover, for a vector $\boldsymbol{x} \in \mathbf{C}^k$, we use the notation $\mathrm{d}\boldsymbol{x} := \mathrm{d}x_1 \ldots \mathrm{d}x_k$.

We will use the concept of "with very high probability" meaning that for any fixed $D > 0$ the probability of the event is bigger than $1 - n^{-D}$ if $n \geq n_0(D)$. Moreover, we use the convention that $\xi > 0$ denotes an arbitrary small constant.

We use the convention that quantities without tilde refer to a general matrix with i.i.d. entries, whilst any quantity with tilde refers to the Ginibre ensemble, e.g. we use X, $\{\sigma_i\}_{i=1}^n$ to denote a non-Hermitian matrix with i.i.d. entries and its eigenvalues, respectively, and $\widetilde{X}$, $\{\widetilde{\sigma}_i\}_{i=1}^n$ to denote their Ginibre counterparts.

5.2 Model and main results

We consider real or complex i.i.d. matrices X, i.e. matrices whose entries are independent and identically distributed as $x_{ab} \stackrel{d}{=} n^{-1/2}\chi$ for a random variable χ. We formulate two assumptions on the random variable χ:

Assumption (5.A). *In the real case we assume that* $\mathbf{E}\,\chi = 0$ *and* $\mathbf{E}\,\chi^2 = 1$, *while in the complex case we assume* $\mathbf{E}\,\chi = \mathbf{E}\,\chi^2 = 0$ *and* $\mathbf{E}\,|\chi|^2 = 1$. *In addition, we assume the existence of high moments, i.e. that there exist constants* $C_p > 0$ *for each* $p \in \mathbf{N}$, *such that*

$$\mathbf{E}\,|\chi|^p \leq C_p. \tag{5.3}$$

Assumption (5.B). *There exist* $\alpha, \beta > 0$ *such that the probability density* $g : \mathbf{F} \to [0, \infty)$ *of the random variable* χ *satisfies*

$$g \in L^{1+\alpha}(\mathbf{F}), \qquad \|g\|_{1+\alpha} \leq n^{\beta}, \tag{5.4}$$

where $\mathbf{F} = \mathbf{R}, \mathbf{C}$ *in the real and complex case, respectively.*

Remark 5.2.1. *We remark that we use Assumption (5.B) only to control the probability of a very small singular value of* $X - z$. *Alternatively, one may use the statement*

$$\mathbf{P}(\operatorname{Spec}(H^z) \cap [-n^{-l}, n^{-l}] = \emptyset) \leq C_l n^{-l/2}, \tag{5.5}$$

for any $l \geq 1$, *uniformly in* $|z| \leq 2$, *that follows directly from [196, Theorem 3.2] without Assumption (5.B). Using* (5.5) *makes Assumption (5.B) superfluous in the entire paper, albeit at the expense of a quite sophisticated proof.*

We denote the eigenvalues of X by $\sigma_1, \dots, \sigma_n \in \mathbf{C}$, and define the *$k$-point correlation function* $p_k^{(n)}$ of X implicitly such that

$$\begin{aligned}\int_{\mathbf{C}^k} F(z_1, \dots, z_k) p_k^{(n)}(z_1, \dots, z_k) \, \mathrm{d}z_1 \dots \mathrm{d}z_k \\ = \binom{n}{k}^{-1} \mathbf{E} \sum_{i_1, \dots, i_k} F(\sigma_{i_1}, \dots, \sigma_{i_k}),\end{aligned} \tag{5.6}$$

for any smooth compactly supported test function $F : \mathbf{C}^k \to \mathbf{C}$, with $i_j \in \{1, \dots, n\}$ for $j \in \{1, \dots, k\}$ all distinct. For the important special case when χ follows a standard real or complex Gaussian distribution, we denote the k-point function of the *Ginibre matrix X* by $p_k^{(n,\mathrm{Gin}(\mathbf{F}))}$ for $\mathbf{F} = \mathbf{R}, \mathbf{C}$. The *circular law* implies that the 1-point function converges

$$\lim_{n\to\infty} p_1^{(n)}(z) = \frac{1}{\pi} \mathbf{1}(z \in \mathbf{D}) = \frac{1}{\pi} \mathbf{1}(|z| \le 1)$$

to the uniform distribution on the unit disk. On the scale $n^{-1/2}$ of individual eigenvalues the scaling limit of the k-point function has been explicitly computed in the case of complex and real Ginibre matrices, $X \sim \mathrm{Gin}(\mathbf{R}), \mathrm{Gin}(\mathbf{C})$, i.e. for any fixed $z_1, \dots, z_k, w_1, \dots, w_k \in \mathbf{C}$ there exist scaling limits $p_{z_1,\dots,z_k}^{(\infty)} = p_{z_1,\dots,z_k}^{(\infty,\mathrm{Gin}(\mathbf{F}))}$ for $\mathbf{F} = \mathbf{R}, \mathbf{C}$ such that

$$\lim_{n\to\infty} p_k^{(n,\mathrm{Gin}(\mathbf{F}))}\Big(z_1 + \frac{w_1}{n^{1/2}}, \dots, z_k + \frac{w_k}{n^{1/2}}\Big) = p_{z_1,\dots,z_k}^{(\infty,\mathrm{Gin}(\mathbf{F}))}(w_1, \dots, w_k). \tag{5.7}$$

Remark 5.2.2. *The k-point correlation function $p_{z_1,\dots,z_k}^{(\infty,\mathrm{Gin}(\mathbf{F}))}$ of the Ginibre ensemble in both the complex and real cases $\mathbf{F} = \mathbf{C}, \mathbf{R}$ is explicitly known; see [102] and [146] for the complex case, and [35, 76, 97] for the real case, where the appearance of $\sim n^{1/2}$ real eigenvalues causes a singularity in the density. In the complex case $p_{z_1,\dots,z_k}^{(\infty,\mathrm{Gin}(\mathbf{C}))}$ is determinantal, i.e. for any $w_1, \dots, w_k \in \mathbf{C}$ it holds*

$$p_{z_1,\dots,z_k}^{(\infty,\mathrm{Gin}(\mathbf{C}))}(w_1, \dots, w_k) = \det\Big(K_{z_i,z_j}^{(\infty,\mathrm{Gin}(\mathbf{C}))}(w_i, w_j)\Big)_{1\le i,j\le k}$$

where for any complex numbers z_1, z_2, w_1, w_2 the kernel $K_{z_1,z_2}^{(\infty,\mathrm{Gin}(\mathbf{C}))}(w_1, w_2)$ is defined by

(i) For $z_1 \ne z_2$, $K_{z_1,z_2}^{(\infty,\mathrm{Gin}(\mathbf{C}))}(w_1, w_2) = 0$.

(ii) For $z_1 = z_2$ and $|z_1| > 1$, $K_{z_1,z_2}^{(\infty,\mathrm{Gin}(\mathbf{C}))}(w_1, w_2) = 0$.

(iii) For $z_1 = z_2$ and $|z_1| < 1$,

$$K_{z_1,z_2}^{(\infty,\mathrm{Gin}(\mathbf{C}))}(w_1, w_2) = \frac{1}{\pi} e^{-\frac{|w_1|^2}{2} - \frac{|w_2|^2}{2} + w_1\overline{w_2}}.$$

(iv) For $z_1 = z_2$ and $|z_1| = 1$,

$$K_{z_1,z_2}^{(\infty,\mathrm{Gin}(\mathbf{C}))}(w_1, w_2) = \frac{1}{2\pi}\Big[1 + \mathrm{erf}\Big(-\sqrt{2}(z_1\overline{w_2} + w_1\overline{z_2})\Big)\Big] e^{-\frac{|w_1|^2}{2} - \frac{|w_2|^2}{2} + w_1\overline{w_2}},$$

where

$$\mathrm{erf}(z) := \frac{2}{\sqrt{\pi}} \int_{\gamma_z} e^{-t^2} \, \mathrm{d}t,$$

for any $z \in \mathbf{C}$, with γ_z any contour from 0 to z.

For the corresponding much more involved formulas for $p_k^{(\infty,\mathrm{Gin}(\mathbf{R}))}$ *we refer the reader to [35].*

Our main result is the universality of $p_{z_1,\ldots,z_k}^{(\infty,\mathrm{Gin}(\mathbf{R},\mathbf{C}))}$ at the edge. In particular we show, that the edge-scaling limit of $p_k^{(n)}$ agrees with the known scaling limit of the corresponding real or complex Ginibre ensemble.

Theorem 5.2.3 (Edge universality). *Let X be an i.i.d. $n \times n$ matrix, whose entries satisfy Assumption (5.A). Then, for any fixed integer $k \geq 1$, and complex spectral parameters $z_1, \ldots, z_k$ such that $|z_j|^2 = 1$, $j = 1, \ldots, k$, and for any compactly supported smooth function $F : \mathbf{C}^k \to \mathbf{C}$, we have the bound*

$$\int_{\mathbf{C}^k} F(\mathbf{w}) \left[p_k^{(n)} \left(\mathbf{z} + \frac{\mathbf{w}}{\sqrt{n}} \right) - p_{\mathbf{z}}^{(\infty,\mathrm{Gin}(\mathbf{F}))}(\mathbf{w}) \right] \mathrm{d}\mathbf{w} = \mathcal{O}(n^{-c}), \tag{5.8}$$

where the constant in $\mathcal{O}(\cdot)$ may depend on k and the C^{2k+1} norm of F, and $c > 0$ is a small constant depending on k.

5.2.1 Proof strategy

For the proof of Theorem 5.2.3 it is essential to study the linearized $2n \times 2n$ matrix H^z defined in (5.1) with eigenvalues $\lambda_1^z \leq \cdots \leq \lambda_{2n}^z$ and resolvent $G(w) = G^z(w) := (H^z - w)^{-1}$. We note that the block structure of H^z induces a spectrum symmetric around 0, i.e. $\lambda_i^z = -\lambda_{2n-i+1}^z$ for $i = 1, \ldots, n$. The resolvent becomes approximately deterministic as $n \to \infty$ and its limit can be found by solving the simple scalar equation

$$-\frac{1}{\widehat{m}^z} = w + \widehat{m}^z - \frac{|z|^2}{w + \widehat{m}^z}, \quad \widehat{m}^z(w) \in \mathbf{H}, \quad w \in \mathbf{H}, \tag{5.9}$$

which is a special case of the *matrix Dyson equation (MDE)*, see e.g. [5]. In the following we may often omit the z-dependence of $\widehat{m}^z$, $G^z(w)$, …, in the notation. We note that on the imaginary axis we have $\widehat{m}(\mathrm{i}\eta) = \mathrm{i}\Im\widehat{m}(\mathrm{i}\eta)$, and in the edge regime $|1 - |z|^2| \lesssim n^{-1/2}$ we have the scaling [13, Lemma 3.3]

$$\Im\widehat{m}(\mathrm{i}\eta) \sim \begin{Bmatrix} |1-|z|^2|^{1/2} + \eta^{1/3}, & |z| \leq 1, \\ \frac{\eta}{|1-|z|^2| + \eta^{2/3}}, & |z| > 1 \end{Bmatrix} \lesssim n^{-1/4} + \eta^{1/3}. \tag{5.10}$$

For $\eta > 0$ we define

$$u = u^z(\mathrm{i}\eta) := \frac{\Im\widehat{m}(\mathrm{i}\eta)}{\eta + \Im\widehat{m}(\mathrm{i}\eta)}, \qquad M = M^z(\mathrm{i}\eta) := \begin{pmatrix} \widehat{m}(\mathrm{i}\eta) & -zu(\mathrm{i}\eta) \\ -\overline{z}u(\mathrm{i}\eta) & \widehat{m}(\mathrm{i}\eta) \end{pmatrix}, \tag{5.11}$$

where M should be understood as a $2n \times 2n$ whose four $n \times n$ blocks are all multiples of the identity matrix, and we note that [13, Eq. (3.62)]

$$u(\mathrm{i}\eta) \lesssim 1, \quad \|M(\mathrm{i}\eta)\| \lesssim 1, \quad \|M'(\mathrm{i}\eta)\| \lesssim \frac{1}{\eta^{2/3}} \tag{5.12}$$

Throughout the proof we shall make use of the following optimal local law which is a direct consequence of [13, Theorem 5.2] (extending [11, Theorem 5.2] to the edge regime). Compared to [13] we require the local law simultaneously in all the spectral parameters z, η and for η slightly below the fluctuation scale $n^{-3/4}$. We defer the proofs for both extensions to Appendix 5.A.

Proposition 5.2.4 (Local law for H^z). *Let X be an i.i.d. $n \times n$ matrix, whose entries satisfy Assumption (5.A) and (5.B), and let H^z be as in (5.1). Then for any deterministic vectors $\boldsymbol{x}, \boldsymbol{y}$ and matrix R and any $\xi > 0$ the following holds true with very high probability: Simultaneously for any z with for $|1 - |z|| \lesssim n^{-1/2}$ and all η such that $n^{-1} \le \eta \le n^{100}$ we have the bounds*

$$
\begin{aligned}
|\langle \boldsymbol{x}, (G^z(\mathrm{i}\eta) - M^z(\mathrm{i}\eta))\boldsymbol{y}\rangle| &\le n^\xi \|\boldsymbol{x}\| \|\boldsymbol{y}\| \Big(\frac{1}{n^{1/2}\eta^{1/3}} + \frac{1}{n\eta} \Big), \\
|\langle R(G^z(\mathrm{i}\eta) - M^z(\mathrm{i}\eta))\rangle| &\le \frac{n^\xi \|R\|}{n\eta}.
\end{aligned}
\tag{5.13}
$$

For the application of Proposition 5.2.4 towards the proof of Theorem 5.2.3 the special case of R being the identity matrix, and $\boldsymbol{x}, \boldsymbol{y}$ being either the standard basis vectors, or the vectors $\mathbf{1}_\pm$ of zeros and ones defined later in (5.58).

The linearized matrix H^z can be related to the eigenvalues σ_i of X via Girko's Hermitization formula [103, 195]

$$
\begin{aligned}
\frac{1}{n} \sum_i f_{z_0}(\sigma_i) &= \frac{1}{4\pi n} \int_{\mathbf{C}} \Delta f_{z_0}(z) \log |\det H_z| \,\mathrm{d}z \\
&= -\frac{1}{4\pi n} \int_{\mathbf{C}} \Delta f_{z_0}(z) \int_0^\infty \Im \operatorname{Tr} G^z(\mathrm{i}\eta) \,\mathrm{d}\eta \,\mathrm{d}z
\end{aligned}
\tag{5.14}
$$

for rescaled test functions $f_{z_0}(z) := n f(\sqrt{n}(z - z_0))$, where $f : \mathbf{C} \to \mathbf{C}$ is smooth and compactly supported. When using (5.14) the small η regime requires additional bounds on the number of small eigenvalues λ_i^z of H^z, or equivalently small singular values of $X - z$. For very small η, say $\eta \le n^{-100}$, the absence of eigenvalues below η, can easily be ensured by Assumption (5.B). For η just below the critical scale of $n^{-3/4}$, however, we need to prove an additional bound on the number of eigenvalues, as stated below.

Proposition 5.2.5. *For any $n^{-1} \le \eta \le n^{-3/4}$ and $||z|^2 - 1| \lesssim n^{-1/2}$ we have the bound*

$$
\begin{aligned}
\mathbf{E}\,|\{i : |\lambda_i^z| \le \eta\}| \lesssim &\begin{cases} n^{3/2}\eta^2(1 + |\log(n\eta^{4/3})|), & X \text{ complex} \\ n^{3/4}\eta, & X \text{ real} \end{cases} \\
&+ \mathcal{O}\Big(\frac{n^\xi}{n^{5/2}\eta^3}\Big),
\end{aligned}
\tag{5.15}
$$

on the number of small eigenvalues, for any $\xi > 0$.

We remark that the precise asymptotics of (5.15) are of no importance for the proof of Theorem 5.2.3. Instead it would be sufficient to establish that for any $\epsilon > 0$ there exists $\delta > 0$ such that we have $\mathbf{E}|\{i : |\lambda_i^z| \le n^{-3/4-\epsilon}\}| \lesssim n^{-\delta}$.

The paper is organized as follows: In Section 5.3 we will prove Proposition 5.2.5 by a Green function comparison argument, using the analogous bound for the Gaussian case, as recently obtained in [61]. In Section 5.4 we will then present the proof of our main result, Theorem 5.2.3, which follows from combining the local law (5.13), Girko's Hermitization identity (5.14), the bound on small singular values (5.15) and another long-time Green function comparison argument.

5.3 Estimate on the lower tail of the smallest singular value of $X - z$

The main result of this section is an estimate of the lower tail of the density of the smallest $|\lambda_i^z|$ in Proposition 5.2.5. For this purpose we introduce the following flow

$$\mathrm{d}X_t = -\frac{1}{2}X_t\,\mathrm{d}t + \frac{\mathrm{d}B_t}{\sqrt{n}}, \tag{5.16}$$

with initial data $X_0 = X$, where B_t is the real or complex matrix valued standard Brownian motion, i.e. $B_t \in \mathbf{R}^{n\times n}$ or $B_t \in \mathbf{C}^{n\times n}$, accordingly with X being real or complex, where $(b_t)_{ab}$ in the real case, and $\sqrt{2}\Re[(b_t)_{ab}]$, $\sqrt{2}\Im[(b_t)_{ab}]$ in the complex case, are independent standard real Brownian motions for $a, b \in [n]$. The flow (5.16) induces a flow $\mathrm{d}\chi_t = -\chi_t\,\mathrm{d}t/2 + \mathrm{d}b_t$ on the entry distribution χ with solution

$$\chi_t = e^{-t/2}\chi + \int_0^t e^{-(t-s)/2}\,\mathrm{d}b_s, \quad \text{i.e.} \quad \chi_t \stackrel{d}{=} e^{-t/2}\chi + \sqrt{1-e^{-t}}g, \tag{5.17}$$

where $g \sim \mathcal{N}(0,1)$ is a standard real or complex Gaussian, independent of χ, with $\mathbf{E}\,g^2 = 0$ in the complex case. By linearity of cumulants we find

$$\kappa_{i,j}(\chi_t) = e^{-(i+j)t/2}\kappa_{i,j}(\chi) + \begin{cases} (1-e^{-t})\kappa_{i,j}(g), & i+j=2 \\ 0, & \text{else}, \end{cases} \tag{5.18}$$

where $\kappa_{i,j}(x)$ denotes the joint cumulant of i copies of x and j copies of $\overline{x}$, in particular $\kappa_{2,0}(x) = \kappa_{0,2}(x) = \kappa_{1,1}(x) = 1$ for $x = \chi, g$ in the real case, and $\kappa_{0,2}(x) = \kappa_{2,0}(x) = 0 \neq \kappa_{1,1}(x) = 1$ for $x = \chi, g$ in the complex case.

Thus (5.17) implies that, in distribution,

$$X_t \stackrel{d}{=} e^{-t/2}X_0 + \sqrt{1-e^{-t}}\widetilde{X}, \tag{5.19}$$

where $\widetilde{X}$ is a real or complex Ginibre matrix independent of $X_0 = X$. Then, we define the $2n\times 2n$ matrix $H_t = H_t^z$ as in (5.1) replacing X by X_t, and its resolvent $G_t(w) = G_t^z(w) := (H_t - w)^{-1}$, for any $w \in \mathbf{H}$. We remark that we defined the flow in (5.16) with initial data X and not H^z in order to preserve the shape of the self consistent density of states of the matrix H_t along the flow. In particular, by (5.16) it follows that H_t is the solution of the flow

$$\mathrm{d}H_t = -\frac{1}{2}(H_t + Z)\,\mathrm{d}t + \frac{\mathrm{d}\mathfrak{B}_t}{\sqrt{n}}, \quad H_0 = H = H^z \tag{5.20}$$

with

$$Z := \begin{pmatrix} 0 & zI \\ \overline{z}I & 0 \end{pmatrix}, \qquad \mathfrak{B}_t := \begin{pmatrix} 0 & B_t \\ B_t^* & 0 \end{pmatrix},$$

where I denotes the $n \times n$ identity matrix.

Proposition 5.3.1. *Let $R_t := \langle G_t(\mathrm{i}\eta)\rangle = \mathrm{i}\langle \Im G_t(\mathrm{i}\eta)\rangle$, then for any $n^{-1} \leq \eta \leq n^{-3/4}$ it holds that*

$$|\,\mathbf{E}[R_{t_2} - R_{t_1}]| \lesssim \frac{(e^{-3t_1/2} - e^{-3t_2/2})n^{\xi}}{n^{7/2}\eta^4}, \tag{5.21}$$

for any arbitrary small $\xi > 0$ and any $0 \leq t_1 < t_2 \leq +\infty$, with the convention that $e^{-\infty} = 0$.

Proof. Denote $W_t := H_t + Z$. By (5.20) and Ito's Lemma it follows that

$$\mathbf{E}\,\frac{\mathrm{d}R_t}{\mathrm{d}t} = \mathbf{E}\left[-\frac{1}{2}\sum_{\alpha} w_\alpha(t)\partial_\alpha R_t + \frac{1}{2}\sum_{\alpha,\beta}\kappa_t(\alpha,\beta)\partial_\alpha\partial_\beta R_t\right], \tag{5.22}$$

where $\alpha, \beta \in [2n]^2$ are double indices, $w_\alpha(t)$ are the entries of W_t and

$$\kappa_t(\alpha, \beta, , \dots) := \kappa(w_\alpha(t), w_\beta(t), \dots) \tag{5.23}$$

denotes the joint cumulant of $w_\alpha, w_\beta, \dots$, and $\partial_\alpha := \partial_{w_\alpha}$. By (5.18) and the independence of χ and g it follows that $\kappa_t(\alpha,\beta) = \kappa_0(\alpha,\beta)$ for all α, β and

$$\begin{aligned}&\kappa_t(\alpha, \beta_1, \dots, \beta_j) \\ &= \begin{cases} e^{-t\frac{j+1}{2}} n^{-\frac{j+1}{2}} \kappa_{l,k}(\chi) & \text{if } \alpha \notin [n]^2 \cup [n+1, 2n]^2,\ \beta_i \in \{\alpha, \alpha'\}\ \forall i \in [j] \\ 0 & \text{otherwise,}\end{cases}\end{aligned} \tag{5.24}$$

for $j > 1$, where for a double index $\alpha = (a,b)$, we use the notation $\alpha' := (b,a)$, and l, k with $l + k = j + 1$ denote the number of double indices among $\alpha, \beta_1, \dots, \beta_j$ which correspond to the upper-right, or respectively lower-left corner of the matrix H. In the sequel the value of $\kappa_{k,l}(\chi)$ is of no importance, but we note that Assumption (5.A) ensures the bound $|\kappa_{k,l}(\chi)| \lesssim \sum_{j \le k+l} C_j < \infty$ for any k, l, with C_j being the constants from Assumption (5.A).

We will use the cumulant expansion that holds for any smooth function f:

$$\mathbf{E}\, w_\alpha f(w) = \sum_{m=0}^{K} \sum_{\beta_1,\dots,\beta_m \in [2n]^2} \frac{\kappa(\alpha, \beta_1, \dots, \beta_m)}{m!}\,\mathbf{E}\,\partial_{\beta_1}\dots\partial_{\beta_m} f(w) + \Omega(K, f), \tag{5.25}$$

where the error term $\Omega(K, f)$ goes to zero as the expansion order K goes to infinity. In our application the error is negligible for, say, $K = 100$ since with each derivative we gain an additional factor of $n^{-1/2}$ and due to the independence (5.24) the sums of any order have effectively only n^2 terms. Applying (5.25) to (5.22) with $f = \partial_\alpha R_t$, the first order term is zero due to the assumption $\mathbf{E}\, x_\alpha = 0$, and the second order term cancels. The third order term is given by

$$\left|\sum_{\alpha\beta_1\beta_2} \kappa_t(\alpha, \beta_1, \beta_2)\,\mathbf{E}[\partial_\alpha\partial_{\beta_1}\partial_{\beta_2} R_t]\right| \lesssim e^{-3t/2}\frac{n^\xi}{n^{7/2}\eta^4}. \tag{5.26}$$

Proof of (5.26). It follows from the resolvent identity that $\partial_\alpha G = -G\Delta^\alpha G$, where Δ^α is the matrix of all zeros except for a 1 in the α-th entry[1]. Thus, neglecting minuses and irrelevant constant factors, for any fixed α, the sum (5.26) is given by a sum of terms of the form

$$\langle G_t \Delta^{\gamma_1} G_t \Delta^{\gamma_2} G_t \Delta^{\gamma_3} G_t\rangle, \qquad \gamma_1, \gamma_2, \gamma_3 \in \{\alpha, \alpha'\}.$$

[1] The matrix Δ^α is not to be confused with the Laplacian Δf in Girko's formula (5.14)

Hence, considering all possible choices of $\gamma_1, \gamma_2, \gamma_3$ and using independence to conclude that $\kappa_t(\alpha, \beta_1, \beta_2)$ can only be non-zero if $\beta_1, \beta_2 \in \{\alpha, \alpha'\}$ we arrive at

$$\begin{aligned} &\Big| \sum_{\alpha\beta_1\beta_2} \kappa_t(\alpha, \beta_1, \beta_2)\, \mathbf{E}[\partial_\alpha \partial_{\beta_1} \partial_{\beta_2} R_t] \Big| \\ &\lesssim e^{-3t/2} n^{-5/2} \Big(\Big| \sum_{abc} \Im\, \mathbf{E}\, G_{ca} G_{ba} G_{ba} G_{bc} \Big| + \Big| \sum_{abc} \Im\, \mathbf{E}\, G_{ca} G_{ba} G_{bb} G_{ac} \Big| \\ &\qquad + \Big| \sum_{abc} \Im\, \mathbf{E}\, G_{ca} G_{bb} G_{aa} G_{bc} \Big| \Big), \end{aligned} \tag{5.27}$$

where the sums are taken over $(a, b) \in [2n]^2 \setminus ([n]^2 \cup [n+1, 2n]^2)$ and $c \in [2n]$, and we dropped the time dependence of $G = G_t$ for notational convenience.

We estimate the three sums in (5.27) using that, by (5.10), (5.12), it follows

$$|G_{ab}| \lesssim n^\xi, \qquad |G_{aa}| \le \Im \widehat{m} + |(G - M)_{aa}| \lesssim n^{-1/4} + \eta^{1/3} + \frac{n^\xi}{n\eta} \lesssim \frac{n^\xi}{n\eta},$$

from Proposition 5.2.4, and Cauchy-Schwarz estimates by

$$\begin{aligned} \sum_{abc} |G_{ca} G_{ba} G_{ba} G_{bc}| &\le \sum_{ab} |G_{ba}|^2 \sqrt{\sum_c |G_{ca}|^2} \sqrt{\sum_c |G_{bc}|^2} \\ &= \sum_{ab} |G_{ba}|^2 \sqrt{(G^* G)_{aa}} \sqrt{(G G^*)_{bb}} \\ &= \frac{1}{\eta} \sum_{ab} |G_{ba}|^2 \sqrt{(\Im G)_{aa}} \sqrt{(\Im G)_{bb}} \lesssim \frac{n^\xi}{n\eta^2} \sum_b (G G^*)_{bb} \\ &= \frac{n^\xi}{n\eta^3} \sum_b (\Im G)_{bb} \lesssim \frac{n^{2\xi}}{n\eta^4}, \end{aligned}$$

and similarly

$$\begin{aligned} \sum_{abc} |G_{ca} G_{ba} G_{bb} G_{ac}| &\lesssim \frac{n^\xi}{n\eta^2} \sum_{ab} |G_{ba}| (\Im G)_{aa} \\ &\le \frac{n^\xi}{n^{1/2} \eta^{5/2}} \sum_a (\Im G)_{aa} \sqrt{(\Im G)_{aa}} \lesssim \frac{n^{5\xi/2}}{n\eta^4} \end{aligned}$$

and

$$\sum_{abc} |G_{ca} G_{bb} G_{aa} G_{bc}| \lesssim \frac{n^{2\xi}}{n^2 \eta^3} \sum_{ab} \sqrt{(\Im G)_{aa}} \sqrt{(\Im G)_{bb}} \lesssim \frac{n^{3\xi}}{n\eta^4}.$$

This concludes the proof of (5.26) by choosing ξ in Proposition 5.2.4 accordingly. □

Finally, in the cumulant expansion of (5.22) we are able to bound the terms of order at least four trivially. Indeed, for the fourth order, the trivial bound is e^{-2t} since the n^3 from the summation is compensated by the n^{-2} from the cumulants and the n^{-1} from the normalization of the trace. Morever, we can always perform at least two Ward-estimates on the first and last G with respect to the trace index. Thus we can estimate any fourth-order term by $e^{-2t}(n\eta)^{-2} \leq e^{-3t/2}n^{-7/2}\eta^{-4}$, and we note that the power-counting for higher order terms is even better than that. Whence we have shown that $\mathbf{E}\,|\,\mathrm{d}R_t/\,\mathrm{d}t| \lesssim e^{-3t/2}n^{-7/2}\eta^{-4}$ and the proof of Proposition 5.3.1 is complete after integrating (5.22) in t from t_1 to t_2. □

Let $\widetilde{X}$ be a real or complex $n \times n$ Ginibre matrix and let $\widetilde{H}^z$ be the linearized matrix defined as in (5.1) replacing X by $\widetilde{X}$. Let $\widetilde{\lambda}_i = \widetilde{\lambda}_i^z$, with $i \in \{1, \dots, 2n\}$, be the eigenvalues of $\widetilde{H}^z$. We define the non negative Hermitian matrix $\widetilde{Y} = \widetilde{Y}^z := (\widetilde{X} - z)(\widetilde{X} - z)^*$, then, by [61, Eq. (13c)-(14)] it follows that for any $\eta \leq n^{-3/4}$ we have

$$\mathbf{E}\,\mathrm{Tr}\,\big[\widetilde{Y} + \eta^2\big]^{-1} = \mathbf{E}\sum_{i=1}^{2n} \frac{1}{\widetilde{\lambda}_i^2 + \eta^2} \lesssim \begin{cases} n^{3/2}(1 + |\log(n\eta^{4/3})|), & \mathrm{Gin}(\mathbf{C}), \\ n^{3/4}\eta^{-1}, & \mathrm{Gin}(\mathbf{R}), \end{cases} \tag{5.28}$$

for $\widetilde{X}$ distributed according to the complex, or respective, real Ginibre ensemble.

Combining (5.28) and Proposition 5.3.1 we now present the proof of Proposition 5.2.5.

Proof of Proposition 5.2.5. Let $\lambda_i(t)$, with $i \in \{1, \dots, 2n\}$, be the eigenvalues of H_t for any $t \geq 0$. Note that $\lambda_i(0) = \lambda_i$, since $H_0 = H^z$. By (5.21), choosing $t_1 = 0$, $t_2 = +\infty$ it follows that

$$\begin{aligned} \mathbf{E}_{H_t}\,|\{i : |\lambda_i| \leq \eta\}| &\leq \eta \cdot \mathbf{E}_{H_t}\left(\Im \sum_{i=1}^{2n} \frac{1}{\lambda_i - \mathrm{i}\eta}\right) \\ &= \eta^2 \cdot \mathbf{E}_{H_\infty}\left(\sum_{i=1}^{2n} \frac{1}{\lambda_i^2 + \eta^2}\right) + \mathcal{O}\left(\frac{n^\xi}{n^{5/2}\eta^3}\right), \end{aligned} \tag{5.29}$$

for any $\xi > 0$. Since the distribution of H_∞ is the same as $\widetilde{H}^z$ it follows that

$$\mathbf{E}_{\widetilde{H}^z}\left(\sum_{i=1}^{2n} \frac{1}{\mu_i^2 + \eta^2}\right) = 2\,\mathbf{E}_{\widetilde{X}}\,\mathrm{Tr}\,\big[\widetilde{Y} + \eta^2\big]^{-1},$$

and combining (5.28) with (5.29), we immediately conclude the bound in (5.15). □

5.4 Edge universality for non-Hermitian random matrices

In this section we prove our main edge universality result, as stated in Theorem 5.2.3.

In the following of this section without loss of generality we can assume that the test function F is of the form

$$F(w_1, \dots, w_k) = f^{(1)}(w_1) \cdots f^{(k)}(w_k), \tag{5.30}$$

with $f^{(1)}, \dots, f^{(k)} : \mathbf{C} \to \mathbf{C}$ being smooth and compactly supported functions. Indeed, any smooth function F can be effectively approximated by its truncated Fourier series (multiplied by smooth cutoff function of product form); see also [195, Remark 3]. Using the effective decay of the Fourier coefficients of F controlled by its C^{2k+1} norm, a standard approximation argument shows that if (5.8) holds for F in the product form (5.30) with an error $\mathcal{O}(n^{-c(k)})$, then it also holds for a general smooth function with an error $\mathcal{O}(n^{-c})$, where the implicit constant in $\mathcal{O}(\cdot)$ depends on k and on the C^{2k+1}-norm of F, and the constant $c > 0$ depends on k.

To resolve eigenvalues on their natural scale we consider the rescaling $f_{z_0}(z) := nf(\sqrt{n}(z-z_0))$ and compare the linear statistics $n^{-1}\sum_i f_{z_0}(\sigma_i)$ and $n^{-1}\sum_i f_{z_0}(\widetilde{\sigma}_i)$, with $\sigma_i, \widetilde{\sigma}_i$ being the eigenvalues of X and of the comparison Ginibre ensemble $\widetilde{X}$, respectively. For convenience we may normalize both linear statistics by their deterministic approximation from the local law (5.13) which, according to (5.14) is given by

$$\frac{1}{n}\sum_i f_{z_0}(\sigma_i) \approx \frac{1}{\pi}\int_{\mathbf{D}} f_{z_0}(z)\,\mathrm{d}z, \tag{5.31}$$

where $\mathbf{D}$ denotes the unit disk of the complex plane.

Proposition 5.4.1. *Let $k \in \mathbf{N}$ and $z_1, \dots, z_k \in \mathbf{C}$ be such that $|z_j|^2 = 1$ for all $j \in [k]$, and let $f^{(1)}, \dots, f^{(k)}$ be smooth compactly supported test functions. Denote the eigenvalues of an i.i.d. matrix X satisfying Assumptions (5.A)–(5.B) and a corresponding real or complex Ginibre matrix $\widetilde{X}$ by $\{\sigma_i\}_{i=1}^n$, $\{\widetilde{\sigma}_i\}_{i=1}^n$. Then we have the bound*

$$\begin{aligned}\mathbf{E}\Bigg[\prod_{j=1}^k &\left(\frac{1}{n}\sum_{i=1}^n f_{z_j}^{(j)}(\sigma_i) - \frac{1}{\pi}\int_{\mathbf{D}} f_{z_j}^{(j)}(z)\,\mathrm{d}z\right) \\ &- \prod_{j=1}^k \left(\frac{1}{n}\sum_{i=1}^n f_{z_j}^{(j)}(\widetilde{\sigma}_i) - \frac{1}{\pi}\int_{\mathbf{D}} f_{z_j}^{(j)}(z)\,\mathrm{d}z\right)\Bigg] = \mathcal{O}(n^{-c(k)}),\end{aligned} \tag{5.32}$$

for some small constant $c(k) > 0$, where the implicit multiplicative constant in $\mathcal{O}(\cdot)$ depends on the norms $\|\Delta f^{(j)}\|_1$, $j = 1, 2, \dots, k$.

Proof of Theorem 5.2.3. Theorem 5.2.3 follows directly from Proposition 5.4.1 by the definition of the k-point correlation function in (5.6), the exclusion-inclusion principle and the bound

$$\left|\frac{1}{\pi}\int_{\mathbf{D}} f_{z_0}(z)\,\mathrm{d}z\right| \lesssim 1. \qquad \square$$

The remainder of this section is devoted to the proof of Proposition 5.4.1. We now fix some $k \in \mathbf{N}$ and some $z_1, \dots, z_k, f^{(1)}, \dots, f^{(k)}$ as in Proposition 5.4.1. All subsequent estimates in this section, also if not explicitly stated, hold true uniformly for any z in an order $n^{-1/2}$-neighborhood of $z_1, \dots, z_k$. In order to prove (5.32), we use Girko's formula (5.14) to write

$$\frac{1}{n}\sum_{i=1}^n f_{z_j}^{(j)}(\sigma_i) - \frac{1}{\pi}\int_{\mathbf{D}} f_{z_j}^{(j)}(z)\,\mathrm{d}z = I_1^{(j)} + I_2^{(j)} + I_3^{(j)} + I_4^{(j)}, \tag{5.33}$$

where

$$\begin{aligned}
I_1^{(j)} &:= \frac{1}{4\pi n}\int_{\mathbf{C}} \Delta f_{z_j}^{(j)}(z)\log|\det(H^z - \mathrm{i}T)|\,\mathrm{d}z \\
I_2^{(j)} &:= -\frac{1}{2\pi}\int_{\mathbf{C}} \Delta f_{z_j}^{(j)}(z)\int_0^{\eta_0} \left[\langle \Im G^z(\mathrm{i}\eta)\rangle - \Im \widehat{m}^z(\mathrm{i}\eta)\right]\mathrm{d}\eta\,\mathrm{d}z \\
I_3^{(j)} &:= -\frac{1}{2\pi}\int_{\mathbf{C}} \Delta f_{z_j}^{(j)}(z)\int_{\eta_0}^{T} \left[\langle \Im G^z(\mathrm{i}\eta)\rangle - \Im \widehat{m}^z(\mathrm{i}\eta)\right]\mathrm{d}\eta\,\mathrm{d}z \\
I_4^{(j)} &:= +\frac{1}{2\pi}\int_{\mathbf{C}} \Delta f_{z_j}^{(j)}(z)\int_T^{+\infty} \left(\Im \widehat{m}^z(\mathrm{i}\eta) - \frac{1}{\eta+1}\right)\mathrm{d}\eta\,\mathrm{d}z,
\end{aligned}$$

with $\eta_0 := n^{-3/4-\delta}$, for some small fixed $\delta > 0$, and for some very large $T > 0$, say $T := n^{100}$. We define $\widetilde{I}_1^{(j)}, \widetilde{I}_2^{(j)}, \widetilde{I}_3^{(j)}, \widetilde{I}_4^{(j)}$ analogously for the Ginibre ensemble by replacing H^z by $\widetilde{H}^z$ and G^z by $\widetilde{G}^z$.

Proof of Proposition 5.4.1. The first step in the proof of Proposition 5.4.1 is the reduction to a corresponding statement about the I_3-part in (5.33), as summarized in the following lemma.

Lemma 5.4.2. *Let $k \geq 1$, let $I_3^{(1)}, \dots, I_3^{(k)}$ be the integrals defined in (5.33), with $\eta_0 = n^{-3/4-\delta}$, for some small fixed $\delta > 0$, and let $\widetilde{I}_3^{(1)}, \dots, \widetilde{I}_3^{(k)}$ be defined as in (5.33) replacing m^z with $\widetilde{m}^z$. Then,*

$$\begin{aligned}\mathbf{E}\Bigg[\prod_{j=1}^{k}\left(\frac{1}{n}\sum_{i=1}^{n} f_{z_j}^{(j)}(\sigma_i) - \frac{1}{\pi}\int_{\mathbf{D}} f_{z_j}^{(j)}(z)\,\mathrm{d}z\right) - \prod_{j=1}^{k}\left(\frac{1}{n}\sum_{i=1}^{n} f_{z_j}^{(j)}(\widetilde{\sigma}_i) - \frac{1}{\pi}\int_{\mathbf{D}} f_{z_j}^{(j)}(z)\,\mathrm{d}z\right)\Bigg] \\ = \mathbf{E}\left[\prod_{j=1}^{k} I_3^{(j)} - \prod_{j=1}^{k} \widetilde{I}_3^{(j)}\right] + \mathcal{O}\left(n^{-c_2(k,\delta)}\right),\end{aligned} \tag{5.34}$$

for some small constant $c_2(k,\delta) > 0$.

In order to conclude the proof of Proposition 5.4.1, due to Lemma 5.4.2, it only remains to prove that

$$\mathbf{E}\left[\prod_{j=1}^{k} I_3^{(j)} - \prod_{j=1}^{k} \widetilde{I}_3^{(j)}\right] = \mathcal{O}\left(n^{-c(k)}\right), \tag{5.35}$$

for any fixed k with some small constant $c(k) > 0$, where we recall the definition of I_3 and the corresponding $\widetilde{I}_3$ for Ginibre from (5.33). The proof of (5.35) is similar to the Green function comparison proof in Proposition 5.3.1 but more involved due to the fact that we compare products of resolvents and that we have an additional η-integration. Here we define the observable

$$Z_t := \prod_{j\in[k]} I_3^{(j)}(t) := \prod_{j\in[k]}\left(-\frac{1}{2\pi}\int_{\mathbf{C}} \Delta f_{z_j}^{(j)}(z)\int_{\eta_0}^{T} \Im\langle G_t^z(\mathrm{i}\eta) - M^z(\mathrm{i}\eta)\rangle\,\mathrm{d}\eta\,\mathrm{d}z\right), \tag{5.36}$$

where we recall that $G_t^z(w) := (H_t^z - w)^{-1}$ with $H_t^z = H_t$ as in (5.20).

Lemma 5.4.3. *For any $n^{-1} \leq \eta_0 \leq n^{-3/4}$ and $T = n^{100}$ and any small $\xi > 0$ it holds that*

$$|\mathbf{E}[Z_{t_2} - Z_{t_1}]| \lesssim \left(e^{-3t_0/2} - e^{-3t_1/2}\right)\frac{n^{\xi}}{n^{5/2}\eta_0^3}\prod_j \left\|\Delta f^{(j)}\right\|_1 \tag{5.37}$$

uniformly in $0 \leq t_1 < t_2 \leq +\infty$ with the convention that $e^{-\infty} = 0$.

Since $Z_0 = \prod_j I_3^{(j)}$ and $Z_\infty = \prod_j \widetilde{I}_3^{(j)}$, the proof of Proposition 5.4.1 follows directly from (5.35), modulo the proofs of Lemmata 5.4.2–5.4.3 that will be given in the next two subsections. $\square$

5.4.1 Proof of Lemma 5.4.2

In order to estimate the probability that there exists an eigenvalue of H^z very close to zero, we use the following proposition that has been proven in [11, Prop. 5.7] adapting the proof of [34, Lemma 4.12].

Proposition 5.4.4. *Under Assumption (5.B) there exists a constant $C > 0$, depending only on α, such that*

$$\mathbf{P}\left(\min_{i\in[2n]} |\lambda_i^z| \leq \frac{u}{n}\right) \leq C u^{\frac{2\alpha}{1+\alpha}} n^{\beta+1}, \tag{5.38}$$

for all $u > 0$ and $z \in \mathbf{C}$.

In the following lemma we prove a very high probability bound for $I_1^{(j)}, I_2^{(j)}, I_3^{(j)}, I_4^{(j)}$. The same bounds hold true for $\widetilde{I}_1^{(j)}, \widetilde{I}_2^{(j)}, \widetilde{I}_3^{(j)}, \widetilde{I}_4^{(j)}$ as well. These bounds in the bulk regime were already proven in [11, Proof of Theorem 2.5], the current edge regime is analogous, so we only provide a sketch of the proof for completeness.

Lemma 5.4.5. *For any $j \in [k]$ the bounds*

$$|I_1^{(j)}| \leq \frac{n^{1+\xi}\|\Delta f^{(j)}\|_1}{T^2}, \quad |I_2^{(j)}| + |I_3^{(j)}| \leq n^{\xi}\|\Delta f^{(j)}\|_1, \quad |I_4^{(j)}| \leq \frac{n\|\Delta f^{(j)}\|_1}{T}, \tag{5.39}$$

hold with very high probability for any $\xi > 0$. The bounds analogous to (5.39) *also hold for $\widetilde{I}_l^{(j)}$.*

Proof. For notational convenience we do not carry the j-dependence of $I_l^{(j)}$ and $f^{(j)}$, and the dependence of $\lambda_i, H, G, M, \widehat{m}$ on z within this proof. Using that

$$\log|\det(H - \mathrm{i}T)| = 2n\log T + \sum_{j\in[n]} \log\left(1 + \frac{\lambda_j^2}{T^2}\right),$$

we easily estimate $|I_1|$ as follows

$$\begin{aligned} |I_1| &= \left|\frac{1}{4\pi n}\int_{\mathbf{C}} \Delta f_{z_j}(z)\log|\det(H - \mathrm{i}T)|\,\mathrm{d}z\right| \\ &\lesssim \frac{1}{n}\int_{\mathbf{C}} |\Delta f_{z_j}(z)|\frac{\operatorname{Tr} H^2}{T^2}\,\mathrm{d}z \lesssim \frac{n^{1+\xi}\|\Delta f\|_1}{T^2}, \end{aligned}$$

for any $\xi > 0$ with very high probability owing to the high moment bound (5.3). By (5.9) it follows that $|\Im\widehat{m}^z(\mathrm{i}\eta) - (\eta+1)^{-1}| \sim \eta^{-2}$ for large η, proving also the bound on I_4 in (5.39). The bound for I_3 follows immediately from the averaged local law in (5.13).

For the I_2 estimate we split the η-integral of $\Im m^z(\mathrm{i}\eta) - \Im\widehat{m}^z(\mathrm{i}\eta)$ in I_2 as follows

$$\begin{aligned} &\int_0^{\eta_0} \Im\langle G^z(\mathrm{i}\eta) - M^z(\mathrm{i}\eta)\rangle\,\mathrm{d}\eta \\ &= \frac{1}{n}\sum_{|\lambda_i|<n^{-l}} \log\left(1 + \frac{\eta_0^2}{\lambda_i^2}\right) + \frac{1}{n}\sum_{|\lambda_i|\geq n^{-l}} \log\left(1 + \frac{\eta_0^2}{\lambda_i^2}\right) - \int_0^{\eta_0} \Im\widehat{m}^z(\mathrm{i}\eta)\,\mathrm{d}\eta, \end{aligned} \tag{5.40}$$

where $l \in \mathbf{N}$ is a large fixed integer. Using (5.10) we find that the third term in (5.40) is bounded by $n^{-1-\delta}$. Choosing l large enough, it follows, as in [11, Eq. (5.35)], using the bound (5.38) that

$$\frac{1}{n} \sum_{|\lambda_i| < n^{-l}} \log\left(1 + \frac{\eta_0^2}{\lambda_i^2}\right) \le n^{-1+\xi}, \tag{5.41}$$

with very high probability for any $\xi > 0$. Alternatively, this bound also follows from (5.5) without Assumption (5.B), circumventing Proposition 5.4.4, see Remark 5.2.1. For the second term in (5.40) we define $\eta_1 := n^{-3/4+\xi}$ with some very small $\xi > 0$ and using $\log(1+x) \le x$ we write

$$\begin{aligned}
\sum_{|\lambda_i| \ge n^{-l}} \log\left(1 + \frac{\eta_0^2}{\lambda_i^2}\right) &= \sum_{n^{-l} \le |\lambda_i| \le n^{\delta/2}\eta_0} \log\left(1 + \frac{\eta_0^2}{\lambda_i^2}\right) + \eta_0^2 \sum_{|\lambda_i| \ge n^{\delta/2}\eta_0} \frac{1}{\lambda_i^2} \\
&\lesssim |\{i : |\lambda_i| < n^{\delta/2}\eta_0\}| \cdot \log n + \eta_0^2 \sum_{|\lambda_i| \ge n^{\delta/2}\eta_0} \frac{1}{\lambda_i^2} \\
&\lesssim (\log n) n^{4\xi/3} + \frac{\eta_0^2 n^{\delta+2\xi}}{\eta_1} \sum_{|\lambda_i| \ge n^{\delta/2}\eta_0} \frac{\eta_1}{\lambda_i^2 + \eta_1^2} \\
&\lesssim (\log n) n^{4\xi/3} + n^{1-\delta}\eta_1 \langle \Im G^z(\mathrm{i}\eta_1)\rangle \le n^{2\xi} + n^{-\delta+2\xi}
\end{aligned} \tag{5.42}$$

by the averaged local law in (5.13), and $\langle \Im M^z(\mathrm{i}\eta_1)\rangle \lesssim \eta_1^{1/3}$ from (5.10). Here from the second to third line in (5.42) we used that

$$|\{i : |\lambda_i| \le n^{\delta/2}\eta_0\}| \le \sum_i \frac{\eta_1^2}{\lambda_i^2 + \eta_1^2} = n\eta_1 \langle \Im G^z(\mathrm{i}\eta_1)\rangle \le n^{4\xi/3}, \tag{5.43}$$

again by the local law. By redefining ξ, this concludes the high probability bound on I_2 in (5.39), and thereby the proof of the lemma. □

In the following lemma we prove an improved bound for $I_2^{(j)}$, compared with (5.39), which holds true only in expectation. The main input of the following lemma is the stronger lower tail estimate on λ_i, in the regime $|\lambda_i| \ge n^{-l}$, from (5.15) instead of (5.43).

Lemma 5.4.6. *Let $I_2^{(j)}$ be defined in* (5.33)*, then*

$$\mathbf{E}\, |I_2^{(j)}| \lesssim n^{-\delta/3} \|\Delta f^{(j)}\|_1, \tag{5.44}$$

for any $j \in \{1, \dots, k\}$.

Proof. We split the η-integral of $\Im m^z(\mathrm{i}\eta) - \Im \widehat{m}^z(\mathrm{i}\eta)$ as in (5.40). The third term in the r.h.s. of (5.40) is of order $n^{-1-4\delta/3}$. Then, we estimate the first term in the r.h.s. of (5.40) as

$$\begin{aligned}
\mathbf{E}\left[\frac{1}{n} \sum_{|\lambda_i| < n^{-l}} \log\left(1 + \frac{\eta_0^2}{\lambda_i^2}\right)\right] &\le \mathbf{E}\left[\log\left(1 + \frac{\eta_0^2}{\lambda_1^2}\right) \mathbf{1}(\lambda_1 \le n^{-l})\right] \\
&\lesssim \mathbf{E}[|\log \lambda_1| \mathbf{1}(\lambda_1 \le n^{-l})] \\
&= \int_{l \log n}^{+\infty} \mathbf{P}(\lambda_1 \le e^{-t})\, \mathrm{d}t \lesssim n^{\beta+1+\frac{2\alpha}{1+\alpha}} e^{-\frac{2\alpha l}{1+\alpha}},
\end{aligned} \tag{5.45}$$

where in the last inequality we use (5.38) with $u = e^{-t}n$. Note that by (5.15) it follows that

$$\mathbf{E}\left|\{i : |\lambda_i| \le n^{\delta/2}\eta_0\}\right| \lesssim n^{-\delta/2}. \tag{5.46}$$

Hence, by (5.46), using similar computations to (5.42), we conclude that

$$\mathbf{E}\left[\frac{1}{n}\sum_{|\lambda_i|\ge n^{-l}} \log\left(1+\frac{\eta_0^2}{\lambda_i^2}\right)\right] \lesssim \frac{\log n}{n^{1+\delta/2}}. \tag{5.47}$$

Note that the only difference to prove (5.47) respect to (5.42) is that the first term in the first line of the r.h.s. of (5.42) is estimated using (5.46) instead of (5.43). Finally, choosing $l \ge \alpha^{-1}(3+\beta)(1+\alpha)+2$, and combining (5.45), (5.47) we conclude (5.44). □

Equipped with Lemmata 5.4.5–5.4.6, we now present the proof of Lemma 5.4.2.

Proof of Lemma 5.4.2. Using the definitions for $I_1^{(j)}, I_2^{(j)}, I_3^{(j)}, I_4^{(j)}$ in (5.33), and similar definitions for $\widetilde{I}_1^{(j)}, \widetilde{I}_2^{(j)}, \widetilde{I}_3^{(j)}, \widetilde{I}_4^{(j)}$, we conclude that

$$\begin{aligned}
&\mathbf{E}\left[\prod_{j=1}^k \left(\frac{1}{n}\sum_{i=1}^n f_{z_j}^{(j)}(\sigma_i) - \frac{1}{\pi}\int_{\mathbf{D}} f_{z_j}^{(j)}(z)\,\mathrm{d}z\right) - \prod_{j=1}^k \left(\frac{1}{n}\sum_{i=1}^n f_{z_j}^{(j)}(\widetilde{\sigma}_i) - \frac{1}{\pi}\int_{\mathbf{D}} f_{z_j}^{(j)}(z)\,\mathrm{d}z\right)\right] \\
&\quad = \mathbf{E}\left[\prod_{j=1}^k \left(I_1^{(j)}+I_2^{(j)}+I_3^{(j)}+I_4^{(j)}\right) - \prod_{j=1}^k \left(\widetilde{I}_1^{(j)}+\widetilde{I}_2^{(j)}+\widetilde{I}_3^{(j)}+\widetilde{I}_4^{(j)}\right)\right] \\
&\quad = \mathbf{E}\left[\prod_{j=1}^k I_3^{(j)} - \prod_{j=1}^k \widetilde{I}_3^{(j)}\right] + \sum_{\substack{j_1+j_2+j_3+j_4=k,\\ j_i\ge 0,\, j_3<k}} \mathbf{E}\prod_{\substack{i_l=1,\\ l=1,2,3,4}}^{j_l} I_1^{(i_1)}I_2^{(i_2)}I_3^{(i_3)}I_4^{(i_4)} \\
&\qquad - \sum_{\substack{j_1+j_2+j_3+j_4=k,\\ j_i\ge 0,\, j_3<k}} \mathbf{E}\prod_{\substack{i_l=1,\\ l=1,2,3,4}}^{j_l} \widetilde{I}_1^{(i_1)}\widetilde{I}_2^{(i_2)}\widetilde{I}_3^{(i_3)}\widetilde{I}_4^{(i_4)}.
\end{aligned}$$

Then, if $j_2 \ge 1$, by Lemma 5.4.5 and Lemma 5.4.6, using that $T = n^{100}$ in the definition of $I_1^{(j)}, \dots, I_4^{(j)}$ in (5.33), it follows that

$$\mathbf{E}\prod_{\substack{i_l=1,\\ l=1,2,3,4}}^{j_l} I_1^{(i_1)}I_2^{(i_2)}I_3^{(i_3)}I_4^{(i_4)} \lesssim \frac{n^{j_1+j_4}n^{(k-j_4-1)\xi}\prod_{j=1}^k \|\Delta f^{(j)}\|_1}{n^{\delta/3}T^{2j_1+j_4}} \le n^{-c_2(k,\delta)},$$

for any $j_1, j_3, j_4 \ge 0$, and a small constant $c(2k,\delta) > 0$ which only depends on k, δ. If, instead, $j_2 = 0$, then at least one among j_1 and j_4 is not zero, since $0 \le j_3 \le k-1$ and $j_1+j_2+j_3+j_4 = k$. Assume $j_1 \ge 1$, the case $j_4 \ge 1$ is completely analogous, then

$$\mathbf{E}\prod_{\substack{i_l=1,\\ l=1,2,3,4}}^{j_l} I_1^{(i_1)}I_2^{(i_2)}I_3^{(i_3)}I_4^{(i_4)} \lesssim \frac{n^{j_1+j_4}n^{(k-j_4)\xi}\prod_{j=1}^k \|\Delta f^{(j)}\|_1}{T^{2j_1+j_4}} \le n^{-c_2(k,\delta)}.$$

Since similar bounds hold true for $\widetilde{I}_1^{(i_1)}, \widetilde{I}_2^{(i_2)}, \widetilde{I}_3^{(i_3)}, \widetilde{I}_4^{(i_4)}$ as well, the above inequalities conclude the proof of (5.34). □

5.4.2 Proof of Lemma 5.4.3

We begin with a lemma generalizing the bound in (5.39) to derivatives of $I_3^{(j)}$.

Lemma 5.4.7. *Assume $n^{-1} \le \eta_0 \le n^{-3/4}$ and fix $l \ge 0$, $j \in [k]$ and a double index $\alpha = (a, b)$ such that $a \neq b$. Then, for any choice of $\gamma_i \in \{\alpha, \alpha'\}$ and any $\xi > 0$ we have the bounds*

$$|\partial_\gamma^l I_3^{(j)}(t)| \lesssim \left\| \Delta f^{(j)} \right\|_1 n^\xi \left(\frac{1}{(n\eta_0)^{\min\{l,2\}}} + \mathbf{1}(a \equiv b + n \pmod{2n}) \right), \tag{5.48}$$

where $\partial_\gamma^l := \partial_{\gamma_1} \dots \partial_{\gamma_l}$, with very high probability uniformly in $t \ge 0$.

Proof. We omit the t- and z-dependence of G_t^z, $\widehat{m}^z$ within this proof since all bounds hold uniformly in $t \ge 0$ and $|z - z_j| \lesssim n^{-1/2}$. We also omit the η-argument from these functions, but the η-dependence of all estimates will explicitly be indicated. Note that the $l = 0$ case was already proven in (5.39). We now separately consider the remaining cases $l = 1$ and $l \ge 2$. For notational simplicity we neglect the n^ξ multiplicative error factors (with arbitrarily small exponents $\xi > 0$) applications of the local law (5.13) within the proof. In particular we will repeatedly use (5.13) in the form

$$\begin{aligned} |G_{ba}| &\lesssim \begin{cases} 1, & a \equiv b + n \pmod{2n}, \\ \psi, & a \not\equiv b + n \pmod{2n}, \end{cases} \qquad G_{bb} = \widehat{m} + \mathcal{O}(\psi), \\ |\widehat{m}| &\lesssim \min\{1, \eta^{1/3} + n^{-1/4}\}, \end{aligned} \tag{5.49}$$

where we defined the parameter

$$\psi := \frac{1}{n\eta} + \frac{1}{n^{1/2}\eta^{1/3}}.$$

Case $l = 1$

This follows directly from

$$\begin{aligned} \left| \int_{\eta_0}^T \langle G \Delta^{ab} G \rangle \, \mathrm{d}\eta \right| &= \left| \frac{1}{n} \int_{\eta_0}^T G_{ba}^2 \, \mathrm{d}\eta \right| = \frac{|G(\mathrm{i}T)_{ab} - G(\mathrm{i}\eta_0)_{ab}|}{n} \\ &\lesssim \frac{1}{n^2\eta_0} + \frac{1}{n} \mathbf{1}(a \equiv b + n \pmod{2n}), \end{aligned}$$

where in the last step we used $\|G(\mathrm{i}T)\| \le T^{-1} = n^{-100}$ and (5.49). Since this bound is uniform in z we may bound the remaining integral by $n\|\Delta f^{(j)}\|_1$, proving (5.48).

Case $l \ge 2$

For the case $l \ge 2$ there are many assignments of γ_i's to consider, e.g.

$$\begin{aligned} \langle G\Delta^{ab} G \Delta^{ab} G \rangle &= \frac{1}{n} \sum_c G_{ca} G_{ba} G_{bc}, \quad \langle G\Delta^{ab} G \Delta^{ba} G \rangle = \frac{1}{n} \sum_c G_{ca} G_{bb} G_{ac}, \\ \langle G\Delta^{ab} G \Delta^{ba} G \Delta^{ab} G \rangle &= \frac{1}{n} \sum_c G_{ca} G_{bb} G_{aa} G_{bc}, \\ \langle G\Delta^{ab} G \Delta^{ba} G \Delta^{ba} G \rangle &= \frac{1}{n} \sum_c G_{ca} G_{bb} G_{ab} G_{ac} \end{aligned}$$

but all are of the form that there are two G-factors carrying the independent summation index c. In the case that $a \equiv b+n \pmod{2n}$ we simply bound all remaining G-factors by 1 using (5.49) and use a simple Cauchy-Schwarz inequality to obtain

$$|\partial_\gamma^l I_3^{(j)}| \lesssim \int_{\mathbf{C}} |\Delta f_{z_j}^{(j)}(z)| \frac{1}{n} \int_{\eta_0}^T \sum_c \Big(|G_{cb}|^2 + |G_{ca}|^2 \Big) \, \mathrm{d}\eta \, \mathrm{d}z. \tag{5.50}$$

Now it follows from the Ward-identity

$$GG^* = G^* G = \frac{\Im G}{\eta} \tag{5.51}$$

and the very crude bound $|G_{aa}| \lesssim 1$ from (5.49) and $|\widehat{m}| \lesssim 1$, that

$$\int_{\eta_0}^T \sum_c \Big(|G_{cb}|^2 + |G_{ca}|^2 \Big) \, \mathrm{d}\eta = \int_{\eta_0}^T \frac{|(\Im G)_{aa}| + |(\Im G)_{bb}|}{\eta} \, \mathrm{d}\eta \lesssim \int_{\eta_0}^T \frac{1}{\eta} \, \mathrm{d}\eta \lesssim \log n.$$

By estimating the remaining z-integral in (5.50) by $n \left\| \Delta f^{(j)} \right\|$ the claimed bound in (5.48) for $a = b+n \pmod{2n}$ follows.

In the case $a \not\equiv b+n \pmod{2n}$ we can use (5.49) to gain a factor of ψ for some G_{ab} or $G_{bb} - \widehat{m}$ in all assignments except for the one in which all but two G-factors are diagonal, and those G_{aa}, G_{bb}-factors are replaced by $\widehat{m}$. For example, we would expand

$$G_{ca} G_{bb} G_{aa} G_{bc} = \widehat{m}^2 G_{ca} G_{bc} + \widehat{m} G_{ca} G_{bc} \mathcal{O}(\psi) + G_{ca} G_{bc} \mathcal{O}(\psi^2),$$

where in all but the first term we gained at least a factor of ψ. Using Cauchy-Schwarz as before we thus have the bound

$$\begin{aligned} |\partial_\gamma^l I_3^{(j)}| \lesssim \int_{\mathbf{C}} \frac{|\Delta f_{z_j}^{(j)}(z)|}{n} \Big(& \int_{\eta_0}^T \psi \sum_c \Big(|G_{cb}|^2 + |G_{ca}|^2 \Big) \, \mathrm{d}\eta \\ & + \left| \int_{\eta_0}^T (\widehat{m})^{l-1} (G^2)_{aa} \, \mathrm{d}\eta \right| + \left| \int_{\eta_0}^T (\widehat{m})^{l-1} (G^2)_{ab} \, \mathrm{d}\eta \right| \Big) \, \mathrm{d}z, \end{aligned} \tag{5.52}$$

where, strictly speaking, the second and third terms are only present for even, or respectively odd, l. For the first term in (5.52) we again proceed by applying the Ward identity (5.51), and (5.49) to obtain the bound

$$\begin{aligned} \int_{\eta_0}^T \psi \sum_c \Big(|G_{cb}|^2 + |G_{ca}|^2 \Big) \, \mathrm{d}\eta &= \int_{\eta_0}^T \psi \frac{|(\Im G)_{aa}| + (\Im G)_{bb}|}{\eta} \, \mathrm{d}\eta \\ &\lesssim \int_{\eta_0}^T \frac{\psi(\psi + \eta^{1/3})}{\eta} \, \mathrm{d}\eta \lesssim \frac{\log n}{(n\eta_0)^2}. \end{aligned}$$

For the second and third terms in (5.52) we use $\mathrm{i}G^2 = G'$, where prime denotes ∂_η, and integration by parts, $|\widehat{m}'| \lesssim \eta^{-2/3}$ from (5.12), and (5.49) to obtain the bounds

$$\begin{aligned} \left| \int_{\eta_0}^T (\widehat{m})^{l-1} (G^2)_{aa} \, \mathrm{d}\eta \right| &\lesssim \left| \int_{\eta_0}^T \widehat{m}' (\widehat{m})^{l-2} G_{aa} \, \mathrm{d}\eta \right| \\ &\qquad + |(\widehat{m}(\mathrm{i}\eta_0))^{l-1} G(\mathrm{i}\eta_0)_{aa}| + |(\widehat{m}(\mathrm{i}T))^{l-1} G(\mathrm{i}T)_{aa}| \\ &\lesssim \left| \int_{\eta_0}^T \widehat{m}' (\widehat{m})^{l-1} \, \mathrm{d}\eta \right| + \int_{\eta_0}^T |\widehat{m}'| \psi \, \mathrm{d}\eta + \frac{1}{n^{1/4}(n\eta_0)} \\ &\lesssim \frac{\log n}{n^{1/4}(n\eta_0)} \end{aligned}$$

and

$$\begin{aligned}\left|\int_{\eta_0}^{T}(\widehat{m})^{l-1}(G^2)_{ab}\,\mathrm{d}\eta\right| &\lesssim \left|\int_{\eta_0}^{T}\widehat{m}'(\widehat{m})^{l-2}G_{ab}\,\mathrm{d}\eta\right| \\ &\quad + |(\widehat{m}(\mathrm{i}\eta_0))^{l-1}G(\mathrm{i}\eta_0)_{ab}| + |(\widehat{m}(\mathrm{i}T))^{l-1}G(\mathrm{i}T)_{ab}| \\ &\lesssim \int_{\eta_0}^{T}|\widehat{m}'|\psi\,\mathrm{d}\eta + \frac{1}{n^{1/4}(n\eta_0)} \lesssim \frac{\log n}{n^{1/4}(n\eta_0)}.\end{aligned}$$

In the explicit deterministic term we performed an integration and estimated

$$\left|\int_{\eta_0}^{T}\widehat{m}'(\widehat{m})^{l-1}\,\mathrm{d}\eta\right| \lesssim |\widehat{m}(\mathrm{i}\eta_0)|^l + |\widehat{m}(\mathrm{i}T)|^l \lesssim n^{-l/4} + n^{-100} \le n^{-1/2}.$$

The claim (5.48) for $l \ge 2$ and $a \not\equiv b + n \pmod{2n}$ now follows from estimating the remaining z-integral in (5.52) by $n\left\|\Delta f^{(j)}\right\|_1$. □

Proof of Lemma 5.4.3. By (5.20) and Ito's Lemma it follows that

$$\mathbf{E}\,\frac{\mathrm{d}Z_t}{\mathrm{d}t} = \mathbf{E}\left[-\frac{1}{2}\sum_{\alpha} w_\alpha(t)\partial_\alpha Z_t + \frac{1}{2}\sum_{\alpha,\beta}\kappa_t(\alpha,\beta)\partial_\alpha\partial_\beta Z_t\right], \tag{5.53}$$

where we recall the definition of κ_t in (5.23). In fact, the point-wise estimate from Lemma 5.4.7 gives a sufficiently strong bound for most terms in the cumulant expansion, the few remaining terms will be computed more carefully.

In the cumulant expansion (5.25) of (5.53) the second order terms cancel exactly and we now separately estimate the third-, fourth- and higher order terms.

Order three terms

For the third order, when computing $\partial_\alpha\partial_{\beta_1}\partial_{\beta_2}Z_t$ through the Leibniz rule we have to consider all possible assignments of derivatives $\partial_\alpha, \partial_{\beta_1}, \partial_{\beta_2}$ to the factors $I_3^{(1)}, \dots, I_3^{(k)}$. Since the particular functions $f^{(j)}$ and complex parameters z_j play no role in the argument, there is no loss in generality in considering only the assignments

$$\begin{gathered}\Big(\partial_{\alpha,\beta_1,\beta_2}I_3^{(1)}\Big)\prod_{j>1}I_3^{(j)},\quad \Big(\partial_{\alpha,\beta_1}I_3^{(1)}\Big)\Big(\partial_{\beta_2}I_3^{(2)}\Big)\prod_{j>2}I_3^{(j)}, \\ \Big(\partial_\alpha I_3^{(1)}\Big)\Big(\partial_{\beta_1}I_3^{(2)}\Big)\Big(\partial_{\beta_2}I_3^{(3)}\Big)\prod_{j>3}I_3^{(j)}\end{gathered} \tag{5.54}$$

for the second and third term of which we obtain a bound of

$$\begin{aligned}&n^{\xi-3/2}e^{-3t/2}\Big(\sum_{a\equiv b+n}\prod_j\left\|\Delta f^{(j)}\right\|_1 + \sum_{a\not\equiv b+n}\prod_j\left\|\Delta f^{(j)}\right\|_1\frac{1}{(n\eta_0)^3}\Big) \\ &\lesssim \frac{n^\xi e^{-3t/2}}{n^{5/2}\eta_0^3}\prod_j\left\|\Delta f^{(j)}\right\|_1\end{aligned}$$

using Lemma 5.4.7 and the cumulant scaling (5.24). Note that the condition $a \neq b$ in the lemma is ensured by the fact that for $a = b$ the cumulants $\kappa_t(\alpha, \beta_1, \dots)$ vanish.

The first term in (5.54) requires an additional argument. We write out all possible index allocations and claim that ultimately we obtain the same bound, as for the other two terms in (5.54), i.e.

$$\begin{aligned}\left|\sum_{\alpha\beta_1\beta_2}\kappa_t(\alpha,\beta_1,\beta_2)\partial_\alpha\partial_{\beta_1}\partial_{\beta_2}I_3^{(1)}\right| &\lesssim \frac{e^{-3t/2}}{n^{3/2}}\int_{\mathbf{C}}\frac{|\Delta f_{z_1}^{(1)}|}{n}J_3\,\mathrm{d}z\\ &\lesssim \frac{n^\xi e^{-3t/2}}{n^{5/2}\eta_0^3}\left\|\Delta f^{(1)}\right\|_1\end{aligned}\tag{5.55}$$

where

$$\begin{aligned}J_3 := &\left|\int_{\eta_0}^T\sum_{ab}(G^2)_{ab}G_{ab}G_{ab}\,\mathrm{d}\eta\right| + \left|\int_{\eta_0}^T\sum_{ab}(G^2)_{aa}G_{bb}G_{ab}\,\mathrm{d}\eta\right|\\ &+\left|\int_{\eta_0}^T\sum_{ab}(G^2)_{ab}G_{aa}G_{bb}\,\mathrm{d}\eta\right|.\end{aligned}\tag{5.56}$$

Proof of (5.55). Compared to the previous bound in Lemma 5.4.7 we now exploit the a,b summation via the isotropic structure of the bound in the local law (5.59). We have the simple bounds

$$\begin{aligned}\frac{|\langle \boldsymbol{x},\Im G\boldsymbol{x}\rangle|}{\|\boldsymbol{x}\|^2} &\lesssim |\widehat{m}| + n^\xi\psi \lesssim n\eta\psi^2,\\ |\langle \boldsymbol{x},G^2\boldsymbol{y}\rangle| &\le \frac{1}{\eta}\sqrt{\langle \boldsymbol{x},\Im G\boldsymbol{x}\rangle\langle \boldsymbol{y},\Im G\boldsymbol{y}\rangle}\lesssim n^\xi\|\boldsymbol{x}\|\|\boldsymbol{y}\|n\psi^2\end{aligned}\tag{5.57}$$

as a consequence of the Ward identity (5.51) and using (5.13) and (5.10). For the first term in (5.56) we can thus use (5.57) and (5.51) to obtain

$$\begin{aligned}\left|\int_{\eta_0}^T\sum_{ab}(G^2)_{ab}G_{ab}G_{ab}\,\mathrm{d}\eta\right| &\lesssim n^\xi\int_{\eta_0}^T n\psi^2\sum_{ab}|G_{ab}|^2\,\mathrm{d}\eta\\ &\lesssim n^\xi\int_{\eta_0}^T n\psi^2\sum_a\frac{(\Im G)_{aa}}{\eta}\,\mathrm{d}\eta\\ &\lesssim n^\xi\int_{\eta_0}^T n^3\psi^4\,\mathrm{d}\eta\lesssim\frac{n^\xi}{n\eta_0^3}.\end{aligned}$$

For the second term in (5.56) we split $G_{bb}=\widehat{m}+\mathcal{O}(\psi)$ and bound it by

$$\begin{aligned}&\left|\int_{\eta_0}^T\sum_{ab}(G^2)_{aa}G_{bb}G_{ab}\,\mathrm{d}\eta\right|\\ &\lesssim n^\xi\int_{\eta_0}^T\psi\sum_{ab}|(G^2)_{aa}G_{ab}|\,\mathrm{d}\eta + \left|\int_{\eta_0}^T\widehat{m}\sum_a(G^2)_{aa}\langle e_a,G\mathbf{1}_{s(a)}\rangle\,\mathrm{d}\eta\right|\\ &\lesssim n^\xi\int_{\eta_0}^T n^{3/2}\psi^2\left(\psi\sum_b\sqrt{\frac{(\Im G)_{bb}}{\eta}}+\sqrt{\frac{\langle\mathbf{1}_+,\Im G\mathbf{1}_+\rangle+\langle\mathbf{1}_-,\Im G\mathbf{1}_-\rangle}{\eta}}\right)\mathrm{d}\eta\\ &\lesssim n^\xi\int_{\eta_0}^T\left(n^3\psi^4+n^{5/2}\psi^3\right)\mathrm{d}\eta\lesssim\frac{n^\xi}{n\eta_0^3}\end{aligned}$$

where e_a denotes the a-th standard basis vector,

$$\mathbf{1}_+ := (1, \ldots, 1, 0, \ldots, 0), \quad \mathbf{1}_- := (0, \ldots, 0, 1, \ldots, 1) \tag{5.58}$$

are vectors of n ones and zeros, respectively, of norm $\|\mathbf{1}_\pm\| = \sqrt{n}$ and $s(a) := -$ for $a \leq n$, and $s(a) := +$ for $a > n$. Here in the second step we used a Cauchy-Schwarz inequality for the a-summation in both integrals after estimating the G^2-terms using (5.57). Finally, for the third term in (5.56) we split both $G_{aa} = \widehat{m} + \mathcal{O}(\psi)$ and $G_{bb} = \widehat{m} + \mathcal{O}(\psi)$ to estimate

$$\begin{aligned}
&\left| \int_{\eta_0}^T \sum_{ab} (G^2)_{ab} G_{aa} G_{bb} \, \mathrm{d}\eta \right| \\
&\quad \lesssim n^\xi \int_{\eta_0}^T n^3 \psi^4 \, \mathrm{d}\eta + \sum_a \int_{\eta_0}^T |\widehat{m} \langle e_a, G^2 \mathbf{1}_{s(a)} \rangle \psi| \, \mathrm{d}\eta + \int_{\eta_0}^T |\widehat{m}^2 \langle \mathbf{1}_+, G^2 \mathbf{1}_- \rangle| \, \mathrm{d}\eta \\
&\quad \lesssim \frac{n^\xi}{n \eta_0^3} + n^\xi \int_{\eta_0}^T n^{5/2} \psi^3 \, \mathrm{d}\eta + n^\xi \int_{\eta_0}^T \frac{n^2 \psi^2}{1 + \eta^2} \, \mathrm{d}\eta \lesssim \frac{n^\xi}{n \eta_0^3},
\end{aligned}$$

using (5.57). In the last integral we used that $|\widehat{m}| \lesssim (1 + \eta)^{-1}$ to ensure the integrability in the large η-regime. Inserting these estimates on (5.56) into (5.55) and estimating the remaining integral by $n \left\| \Delta f^{(1)} \right\|_1$ completes the proof of (5.55). □

Order four terms

For the fourth-order Leibniz rule we have to consider the assignments

$$\begin{aligned}
&\left(\partial_{\alpha, \beta_1, \beta_2, \beta_3} I_3^{(1)} \right) \prod_{j > 1} I_3^{(j)}, \quad \left(\partial_{\alpha, \beta_1, \beta_2} I_3^{(1)} \right) \left(\partial_{\beta_3} I_3^{(2)} \right) \prod_{j > 2} I_3^{(j)}, \\
&\left(\partial_{\alpha, \beta_1} I_3^{(1)} \right) \left(\partial_{\beta_2, \beta_3} I_3^{(2)} \right) \prod_{j > 2} I_3^{(j)}, \quad \left(\partial_{\alpha, \beta_1} I_3^{(1)} \right) \left(\partial_{\beta_2} I_3^{(2)} \right) \left(\partial_{\beta_3} I_3^{(3)} \right) \prod_{j > 3} I_3^{(j)}, \\
&\left(\partial_{\alpha, \beta_1} I_3^{(1)} \right) \left(\partial_{\beta_2} I_3^{(2)} \right) \left(\partial_{\beta_2} I_3^{(3)} \right) \left(\partial_{\beta_3} I_3^{(4)} \right) \prod_{j > 4} I_3^{(j)},
\end{aligned}$$

for all of which we obtain a bound of

$$\frac{n^\xi e^{-2t}}{n^2 \eta_0^2} \prod_j \left\| \Delta f^{(j)} \right\|_1,$$

again using Lemma 5.4.7 and (5.24).

Higher order terms

For terms order at least 5, there is no need to additionally gain from any of the factors of I_3 and we simply bound all those, and their derivatives, by n^ξ using Lemma 5.4.7. This results in a bound of $n^{\xi - (l-4)/2} e^{-lt/2} \prod_j \left\| \Delta f^{(j)} \right\|_1$ for the terms of order l.

By combining the estimates on the terms of order three, four and higher order derivatives, and integrating in t we obtain the bound (5.37). This completes the proof of Lemma 5.4.3. □

5.A Extension of the local law

Proof of Proposition 5.2.4. The statement follows directly from [13, Theorem 5.2] if $\eta \geq \eta_0 := n^{-3/4+\epsilon}$. For smaller η_1, using $\partial_\eta G(\mathrm{i}\eta) = \mathrm{i}G^2(\mathrm{i}\eta)$, we write

$$\begin{aligned} \langle \boldsymbol{x}, [G(\mathrm{i}\eta_1) - M(\mathrm{i}\eta_1)]\boldsymbol{y}\rangle &= \langle \boldsymbol{x}, [G(\mathrm{i}\eta_0) - M(\mathrm{i}\eta_0)]\boldsymbol{y}\rangle \\ &\quad + \mathrm{i}\int_{\eta_0}^{\eta_1} \langle \boldsymbol{x}, [G^2(\mathrm{i}\eta) - M'(\mathrm{i}\eta)]\boldsymbol{y}\rangle \,\mathrm{d}\eta \end{aligned} \tag{5.59}$$

and estimate the first term using the local law by $n^{-1/4+\xi}$. For the second term we bound

$$\begin{aligned} |\langle \boldsymbol{x}, G^2\boldsymbol{y}\rangle| &\leq \sqrt{\langle \boldsymbol{x}, G^*G\boldsymbol{x}\rangle\langle \boldsymbol{y}, G^*G\boldsymbol{y}\rangle} = \frac{1}{\eta}\sqrt{\langle \boldsymbol{x}, \Im G\boldsymbol{x}\rangle\langle \boldsymbol{y}, \Im G\boldsymbol{y}\rangle}, \\ |\langle \boldsymbol{x}, M'\boldsymbol{y}\rangle| &\lesssim \|\boldsymbol{x}\|\|\boldsymbol{y}\|\frac{1}{\eta^{2/3}} \end{aligned}$$

from $\|M'\| \lesssim (\Im\widehat{m})^{-2}$ and (5.10), and use monotonicity of $\eta \mapsto \eta\langle \boldsymbol{x}, \Im G(\mathrm{i}\eta)\boldsymbol{x}\rangle$ in the form

$$\Im\langle \boldsymbol{x}, G(\mathrm{i}\eta)\boldsymbol{x}\rangle \leq \frac{\eta_0}{\eta}\langle \boldsymbol{x}, \Im G(\mathrm{i}\eta_0)\boldsymbol{x}\rangle \prec \|\boldsymbol{x}\|^2\Big(\frac{\eta_0^{4/3}}{\eta} + \frac{\eta_0^{2/3}}{\eta n^{1/2}}\Big) \lesssim \|\boldsymbol{x}\|^2\frac{n^{4\epsilon/3}}{n\eta}.$$

After integration we thus obtain a bound of $\|\boldsymbol{x}\|\|\boldsymbol{y}\|n^{4\epsilon/3}/(n\eta_1)$ which proves the first bound in (5.13). The second, averaged, bound in (5.13) follows directly from the first one since below the scale $\eta \leq n^{-3/4}$ there is no additional gain from the averaging, as compared to the isotropic bound.

In order to conclude the local law simultaneously in all z, η we use a standard *grid argument*. To do so, we choose a regular grid of z's and η's at a distance of, say, n^{-3} and use Lipschitz continuity (with Lipschitz constant n^2) of $(\eta, z) \mapsto G^z(\mathrm{i}\eta)$ and a union bound over the exceptional events at each grid point. □

Central limit theorem for linear eigenvalue statistics of non-Hermitian random matrices 6

We consider large non-Hermitian random matrices X with complex, independent, identically distributed centred entries and show that the linear statistics of their eigenvalues are asymptotically Gaussian for test functions having $2 + \epsilon$ derivatives. Previously this result was known only for a few special cases; either the test functions were required to be analytic [162], or the distribution of the matrix elements needed to be Gaussian [164], or at least match the Gaussian up to the first four moments [195], [126]. We find the exact dependence of the limiting variance on the fourth cumulant that was not known before. The proof relies on two novel ingredients: (i) a local law for a product of two *resolvents of the Hermitisation of X with different spectral parameters and (ii) a coupling of several weakly dependent Dyson Brownian Motions. These methods are also the key inputs for our analogous results on the linear eigenvalue statistics of* real *matrices X that are presented in the companion paper [60].*

Published as G. Cipolloni et al., *Central limit theorem for linear eigenvalue statistics of non-Hermitian random matrices*, Accepted to Communications on Pure and Applied Mathematics (2020), `arXiv:1912.04100`

6.1 Introduction

Eigenvalues of random matrices form a strongly correlated point process. One manifestation of this fact is the unusually small fluctuation of their linear statistics making the eigenvalue process distinctly different from a Poisson point process. Suppose that the $n \times n$ random matrix X has i.i.d. entries of zero mean and variance $1/n$. The empirical density of the eigenvalues $\{\sigma_i\}_{i=1}^n$ converges to a limit distribution; it is the uniform distribution on the unit disk in the non-Hermitian case *(circular law)* and the semicircular density in the Hermitian case *(Wigner semicircle law)*. For test functions f defined on the spectrum one may consider the fluctuation of the linear statistics and one expects that

$$L_n(f) := \sum_{i=1}^n f(\sigma_i) - \mathbf{E}\sum_{i=1}^n f(\sigma_i) \sim \mathcal{N}(0, V_f) \tag{6.1}$$

converges to a centred normal distribution as $n \to \infty$. The variance V_f is expected to depend only on the second and fourth moments of the single entry distribution. Note that, unlike in the usual central limit theorem, there is no $1/\sqrt{n}$ rescaling in (6.1) which is a quantitative indication of a strong correlation. The main result of the current paper is the proof of (6.1) for non-Hermitian random matrices with complex i.i.d. entries and for general test functions f. We give an explicit formula for V_f that involves the fourth cumulant of X as well, disproving a conjecture by Chafaï [51]. By polarisation, from (6.1) it also follows that the limiting joint distribution of $(L_n(f_1), L_n(f_2), \dots, L_n(f_k))$ for a fixed number of test functions is jointly Gaussian.

We remark that another manifestation of the strong eigenvalue correlation is the repulsion between neighbouring eigenvalues. For Gaussian ensembles the local repulsion is directly seen from the well-known determinantal structure of the joint distribution of all eigenvalues; both in the non-Hermitian *Ginibre* case and in the Hermitian *GUE/GOE* case. In the spirit of *Wigner-Dyson-Mehta universality* of the local correlation functions [146] level repulsion should also hold for random matrices with general distributions. While for the Hermitian case the universality has been rigorously established for a large class of random matrices (see e.g. [90] for a recent monograph), the analogous result for the non-Hermitian case is still open in the bulk spectrum (see, however, [59] for the edge regime and [195] for entry distributions whose first four moments match the Gaussian).

These two manifestations of the eigenvalue correlations cannot be deduced from each other, however the proofs often share common tools. For n-independent test functions f, (6.1) apparently involves understanding the eigenvalues only on the macroscopic scales, while the level repulsion is expressly a property on the microscopic scale of individual eigenvalues. However the suppression of the usual $\sqrt{n}$ fluctuation is due to delicate correlations on all scales, so (6.1) also requires understanding local scales.

Hermitian random matrices are much easier to handle, hence fluctuation results of the type (6.1) have been gradually obtained for more and more general matrix ensembles as well as for broader classes of test functions, see, e.g. [19, 117, 124, 143, 169] and [187] for the weakest regularity conditions on f. Considering n-dependent test functions, Gaussian fluctuations have been detected even on mesoscopic scales [47, 48, 72, 110, 112, 114, 128, 138].

Non-Hermitian random matrices pose serious challenges, mainly because their eigenvalues are potentially very unstable. When X has i.i.d. centred Gaussian entries with variance $1/n$ (this is called the *Ginibre ensemble*), the explicit determinantal formulas for the correlation functions may be used to compute the distribution of the linear statistics $L_n(f)$. Forrester in [95] proved (6.1) for complex Ginibre ensemble and radially symmetric f and obtained the variance $V_f = (4\pi)^{-1}\int_{\mathbf{D}}|\nabla f|^2 \,\mathrm{d}^2 z$ where $\mathbf{D}$ is the unit disk. He also gave a heuristic argument based on Coulomb gas theory for general f and his calculations predicted an additional boundary term $\frac{1}{2}\|f\|^2_{\dot{H}^{1/2}(\partial\mathbf{D})}$ in the variance V_f. Rider considered test functions f depending only on the angle [161] when $f \notin H^1(\mathbf{D})$ and accordingly V_f grows with $\log n$ (similar growth is proved for $f = \log$ in [150]). Finally, Rider and Virág in [164] have rigorously verified Forrester's prediction for general $f \in C^1(\mathbf{D})$ using a cumulant formula for determinantal processes found first by Costin and Lebowitz [64] and extended by

Soshnikov [185]. They also presented a *Gaussian free field (GFF)* interpretation of the result that we extend in Section 6.2.1.

The first result beyond the explicitly computable Gaussian case is due to Rider and Silverstein [162, Theorem 1.1] who proved (6.1) for X with i.i.d. complex matrix elements and for test functions f that are analytic on a large disk. Analyticity allowed them to use contour integration and thus deduce the result from analysing the resolvent at spectral parameters far away from the actual spectrum. The domain of analyticity was optimized in [152], where extensions to elliptic ensembles were also proven. Polynomial test functions via the alternative moment method were considered by Nourdin and Peccati in [151]. The analytic method of [162] was recently extended by Coston and O'Rourke [65] to fluctuations of linear statistics for *products* of i.i.d. matrices. However, these method fail for a larger class of test functions.

Since the first four moments of the matrix elements fully determine the limiting eigenvalue statistics, Tao and Vu were able to compare the fluctuation of the local eigenvalue density for a general non-Gaussian X with that of a Ginibre matrix [195, Corollary 10] assuming the first four moments of X match those of the complex Ginibre ensemble. This method was extended by Kopel [126, Corollary 1] to general smooth test functions with an additional study on the real eigenvalues when X is real (see also the work of Simm for polynomial statistics of the real eigenvalues [179]).

Our result removes the limitations of both previous approaches: we allow general test functions and general distribution for the matrix elements without constraints on matching moments. We remark that the dependence of the variance V_f on the fourth cumulant of the single matrix entry escaped all previous works. The Ginibre ensemble with its vanishing fourth cumulant clearly cannot catch this dependence. Interestingly, even though the fourth cumulant in general is not zero in the work Rider and Silverstein [162], it is multiplied by a functional of f that happens to vanish for analytic functions (see (6.9), (6.11) and Remark 6.2.4 later). Hence this result did not detect the precise role of the fourth cumulant either. This may have motivated the conjecture [51] that the variance does not depend on the fourth cumulant at all.

In order to focus on the main new ideas, in this paper we consider the problem only for X with genuinely complex entries. Our method also works for real matrices where the real axis in the spectrum plays a special role that modifies the exact formula for the expectation and the variance V_f in (6.1). This leads to some additional technical complications that we have resolved in a separate work [60] which contains the real version of our main Theorem 6.2.1.

Finally, we remark that the problem of fluctuations of linear statistics has been considered for β-log-gases in one and two dimensions; these are closely related to the eigenvalues of the Hermitian, resp. non-Hermitian Gaussian matrices for classical values $\beta = 1, 2, 4$ and for quadratic potential. In fact, in two dimensions the logarithmic interaction also corresponds to the Coulomb gas from statistical physics. Results analogous to (6.1) in one dimension were obtained e.g. in [1, 26, 27, 37, 114, 117, 127, 170]. In two dimensions similar results have been established both in the macroscopic [132] and in the mesoscopic [25] regimes.

We now outline the main ideas in our approach. We use Girko's formula [103] in the form given in [195] to express linear eigenvalue statistics of X in terms of resolvents of a family of $2n \times 2n$ Hermitian matrices

$$H^z := \begin{pmatrix} 0 & X - z \\ X^* - \overline{z} & 0 \end{pmatrix} \tag{6.2}$$

parametrized by $z \in \mathbf{C}$. This formula asserts that

$$\sum_{\sigma \in \mathrm{Spec}(X)} f(\sigma) = -\frac{1}{4\pi} \int_{\mathbf{C}} \Delta f(z) \int_0^\infty \Im \operatorname{Tr} G^z(i\eta)\, \mathrm{d}\eta\, \mathrm{d}^2 z \tag{6.3}$$

for any smooth, compactly supported test function f (the apparent divergence of the η-integral at infinity can easily be removed, see (6.28)). Here we set $G^z(w) := (H^z - w)^{-1}$ to be the resolvent of H^z. We have thus transformed our problem to a Hermitian one and all tools and results developed for Hermitian ensembles in the recent years are available.

Utilizing Girko's formula requires a good understanding of the resolvent of H^z along the imaginary axis for all $\eta > 0$. On very small scales $\eta \ll n^{-1}$, there are no eigenvalues thus $\Im \operatorname{Tr} G^z(\mathrm{i}\eta)$ is negligible. All other scales $\eta \gtrsim n^{-1}$ need to be controlled carefully since *a priori* they could all contribute to the fluctuation of $L_n(f)$, even though *a posteriori* we find that the entire variance comes from scales $\eta \sim 1$.

In the mesoscopic regime $\eta \gg n^{-1}$, *local laws* from [11, 13] accurately describe the leading order deterministic behaviour of $\frac{1}{n} \operatorname{Tr} G^z(\mathrm{i}\eta)$ and even the matrix elements $G^z_{ab}(\mathrm{i}\eta)$; now we need to identify the next order fluctuating term in the local law. In other words we need to prove a central limit theorem for the traces of resolvents G^z. In fact, based upon (6.3), for the higher k-th moments of $L_n(f)$ we need the joint distribution of $\operatorname{Tr} G^{z_l}(\mathrm{i}\eta)$ for different spectral parameters $z_1, z_2, \ldots, z_k$. This is one of our main technical achievements. Note that the asymptotic joint Gaussianity of traces of Wigner resolvents $\operatorname{Tr}(H - w_1)^{-1}, \operatorname{Tr}(H - w_2)^{-1}, \ldots$ at different spectral parameters has been obtained in [111, 112]. However, the method of this result is not applicable since the role of the spectral parameter z in (6.2) is very different from w; it is in an off-diagonal position thus these resolvents do not commute and they are not in the spectral resolution of a single matrix.

The microscopic regime, $\eta \sim n^{-1}$, is much more involved than the mesoscopic one. Local laws and their fluctuations are not sufficient, we need to trace the effect of the individual eigenvalues $0 \le \lambda_1^z \le \lambda_2^z, \ldots$ of H^z near zero (the spectrum of H^z is symmetric, we may focus on the positive eigenvalues). Moreover, we need their *joint* distribution for different z parameters which, for arbitrary z's, is not known even in the Ginibre case. We prove, however, that λ_1^z and $\lambda_1^{z'}$ are asymptotically independent if z and z' are far away, say $|z - z'| \ge n^{-1/100}$. A similar result holds simultaneously for several small eigenvalues. Notice that due to the z-integration in (6.3), when the k-th moment of $L_n(f)$ is computed, the integration variables $z_1, z_2, \ldots, z_k$ are typically far away from each other. The resulting independence of the spectra of $H^{z_1}, H^{z_2}, \ldots$ near zero ensures that the microscopic regime eventually does not contribute to the fluctuation of $L_n(f)$.

The proof of the independence of λ_1^z and $\lambda_1^{z'}$ relies on the analysis of the *Dyson Brownian motion* (DBM) developed in the recent years [90] for the proof of the Wigner-Dyson-Mehta universality conjecture for Wigner matrices. The key mechanism is the fast local equilibration of the eigenvalues $\boldsymbol{\lambda}^z(t) := \{\lambda_i^z(t)\}$ along the stochastic flow generated by adding a small time-dependent Gaussian component to the original matrix. This Gaussian component can then be removed by the *Green function comparison theorem* (GFT). One of the main technical results of [54] (motivated by the analogous analysis in [129] for Wigner matrices that relied on coupling and homogenisation ideas introduced first in [42]) asserts that for any fixed z the DBM process $\boldsymbol{\lambda}^z(t)$ can be pathwise approximated by a similar DBM with a different initial condition by *exactly* coupling the driving Brownian motions in their DBMs. We extend this idea to simultaneously trailing $\boldsymbol{\lambda}^z(t)$ and $\boldsymbol{\lambda}^{z'}(t)$ by their independent Ginibre counterparts. The evolutions of $\boldsymbol{\lambda}^z(t)$ and $\boldsymbol{\lambda}^{z'}(t)$ are not independent since their driving Brownian motions are correlated; the correlation is given by the eigenfunction overlap $\langle u_i^z, u_j^{z'}\rangle\langle v_j^{z'}, v_i^z\rangle$ where $w_i^z = (u_i^z, v_i^z) \in \mathbf{C}^n \times \mathbf{C}^n$ denotes the eigenvector of H^z belonging to λ_i^z. However, this overlap turns out to be small if z and z' are far away and i is not too big. Thus the analysis of the microscopic regime has two ingredients: (i) extending the coupling idea to driving Brownian motions whose distributions are not identical but close to each other; and (ii) proving the smallness of the overlap.

While (i) can be achieved by relatively minor modifications to the proofs in [54], (ii) requires to develop a new type of local law. Indeed, the overlap can be estimated in terms of traces of products of resolvents, $\operatorname{Tr} G^z(\mathrm{i}\eta)G^{z'}(\mathrm{i}\eta')$ with $\eta, \eta' \sim n^{-1+\epsilon}$ in the mesoscopic regime. Customary local laws, however, do not apply to a quantity involving *products* of resolvents. In fact, even the leading deterministic term needs to be identified by solving a new type of deterministic Dyson equation. We first show the stability of this new equation using the lower bound on $|z - z'|$. Then we prove the necessary high probability bound for the error term in the Dyson equation by a diagrammatic cumulant expansion adapted to the new situation of product of resolvents. The key novelty is to extract the effect that G^z and $G^{z'}$ are weakly correlated when z and z' are far away from each other.

We close this section with an important remark concerning the proofs for Hermitian versus non-Hermitian matrices. Similarly to Girko's formula (6.3), the linear eigenvalue statistics for Hermitian matrices are also expressed by an integral of the resolvents over all spectral parameters. However, in the corresponding Helffer-Sjöstrand formula, sufficient regularity of f directly neutralizes the potentially singular behaviour of the resolvent near the real axis, giving rise to CLT results even with suboptimal control on the resolvent in the mesoscopic regime. A similar trade-off in (6.3) is not apparent; it is unclear if and how the integration in z could help regularize the η integral. This is a fundamental difference between CLTs for Hermitian and non-Hermitian ensembles that explains the abundance of Hermitian results in contrast to the scarcity of available non-Hermitian CLTs.

Acknowledgement

L.E. would like to thank Nathanaël Berestycki, and D.S. would like to thank Nina Holden for valuable discussions on the Gaussian free field.

Notations and conventions

We introduce some notations we use throughout the paper. For integers $k \in \mathbf{N}$ we use the notation $[k] := \{1, \dots, k\}$. We write $\mathbf{H}$ for the upper half-plane $\mathbf{H} := \{z \in \mathbf{C} \mid \Im z > 0\}$, $\mathbf{D} \subset \mathbf{C}$ for the open unit disk, and for any $z \in \mathbf{C}$ we use the notation $\mathrm{d}^2 z := 2^{-1}\mathrm{i}(\mathrm{d}z \wedge \mathrm{d}\overline{z})$ for the two dimensional volume form on $\mathbf{C}$. For positive quantities f, g we write $f \lesssim g$ and $f \sim g$ if $f \leq Cg$ or $cg \leq f \leq Cg$, respectively, for some constants $c, C > 0$ which depend only on the constants appearing in (6.4). For any two positive real numbers $\omega_*, \omega^* \in \mathbf{R}_+$ by $\omega_* \ll \omega^*$ we denote that $\omega_* \leq c\omega^*$ for some small constant $0 < c \leq 1/100$. We denote vectors by bold-faced lower case Roman letters $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{C}^k$, for some $k \in \mathbf{N}$. Vector and matrix norms, $\|\boldsymbol{x}\|$ and $\|A\|$, indicate the usual Euclidean norm and the corresponding induced matrix norm. For any $2n \times 2n$ matrix A we use the notation $\langle A \rangle := (2n)^{-1} \operatorname{Tr} A$ to denote the normalized trace of A. Moreover, for vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{C}^n$ and matrices $A, B \in \mathbf{C}^{2n\times 2n}$ we define

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum \overline{x}_i y_i, \qquad \langle A, B \rangle := \langle A^* B \rangle.$$

We will use the concept of "with very high probability" meaning that for any fixed $D > 0$ the probability of the event is bigger than $1 - n^{-D}$ if $n \geq n_0(D)$. Moreover, we use the convention that $\xi > 0$ denotes an arbitrary small constant which is independent of n.

6.2 Main results

We consider *complex i.i.d. matrices* X, i.e. $n \times n$ matrices whose entries are independent and identically distributed as $x_{ab} \overset{d}{=} n^{-1/2}\chi$ for some complex random variable χ, satisfying the following:

Assumption (6.A). *We assume that* $\mathbf{E}\,\chi = \mathbf{E}\,\chi^2 = 0$ *and* $\mathbf{E}|\chi|^2 = 1$. *In addition we assume the existence of high moments, i.e. that there exist constants* $C_p > 0$, *for any* $p \in \mathbf{N}$, *such that*

$$\mathbf{E}|\chi|^p \leq C_p. \tag{6.4}$$

The *circular law* [18, 20, 33, 101, 103, 105, 154, 191] asserts that the empirical distribution of eigenvalues $\{\sigma_i\}_{i=1}^n$ of a complex i.i.d. matrix X converges to the uniform distribution on the unit disk $\mathbf{D}$, i.e.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n f(\sigma_i) = \frac{1}{\pi} \int_{\mathbf{D}} f(z)\, \mathrm{d}^2 z, \tag{6.5}$$

with very high probability for any continuous bounded function f. Our main result is a central limit theorem for the centred *linear statistics*

$$L_n(f) := \sum_{i=1}^n f(\sigma_i) - \mathbf{E} \sum_{i=1}^n f(\sigma_i) \tag{6.6}$$

for general complex i.i.d. matrices and generic test functions f.

In order to state the result we introduce some notations and certain Sobolev spaces. We fix some open bounded $\Omega \subset \mathbf{C}$ containing the closed unit disk $\overline{\mathbf{D}} \subset \Omega$ and having a piecewise C^1-boundary, or, more generally, any boundary satisfying the *cone property* (see e.g. [141, Section 8.7]). We consider test functions $f \in H_0^{2+\delta}(\Omega)$ in the Sobolev space $H_0^{2+\delta}(\Omega)$ which is defined as the completion of the smooth compactly supported functions $C_c^\infty(\Omega)$ under the norm

$$\|f\|_{H^{2+\delta}(\Omega)} := \|(1+|\xi|)^{2+\delta} \widehat{f}(\xi)\|_{L^2(\Omega)}$$

and we note that by Sobolev embedding such functions are continuously differentiable, and vanish at the boundary of Ω. For notational convenience we identify $f \in H_0^{2+\delta}(\Omega)$ with its extension to all of $\mathbf{C}$ obtained from setting $f \equiv 0$ in $\mathbf{C} \setminus \Omega$. We note that our results can trivially be extended to bounded test functions with non-compact support since due to [13, Theorem 2.1], with high probability, all eigenvalues satisfy $|\sigma_i| \leq 1 + \epsilon$ and therefore non-compactly supported test functions can simply be smoothly cut-off. For h defined on the boundary of the unit disk $\partial\mathbf{D}$ we define its Fourier transform

$$\widehat{h}(k) = \frac{1}{2\pi}\int_0^{2\pi} h(e^{\mathrm{i}\theta})e^{-\mathrm{i}\theta k}\,\mathrm{d}\theta, \qquad k\in\mathbf{Z}. \tag{6.7}$$

For $f,g\in H_0^{2+\delta}(\Omega)$ we define the homogeneous semi-inner products

$$\langle g,f\rangle_{\dot{H}^{1/2}(\partial\mathbf{D})} := \sum_{k\in\mathbf{Z}}|k|\widehat{f}(k)\overline{\widehat{g}(k)}, \qquad \|f\|^2_{\dot{H}^{1/2}(\partial\mathbf{D})} := \langle f,f\rangle_{\dot{H}^{1/2}(\partial\mathbf{D})}, \tag{6.8}$$

where, with a slight abuse of notation, we identified f and g with their restrictions to $\partial\mathbf{D}$.

Theorem 6.2.1 (Central Limit Theorem for linear statistics)**.** *Let X be a complex $n\times n$ i.i.d. matrix satisfying Assumption (6.A) with eigenvalues $\{\sigma_i\}_{i=1}^n$, and denote the fourth* cumulant *of χ by $\kappa_4 := \mathbf{E}|\chi|^4-2$. Fix $\delta>0$, an open complex domain Ω with $\overline{\mathbf{D}}\subset\Omega\subset\mathbf{C}$ and a complex valued test function $f\in H_0^{2+\delta}(\Omega)$. The centred linear statistics $L_n(f)$, defined in* (6.6), *converges*

$$L_n(f)\Longrightarrow L(f),$$

to a centred complex Gaussian random variable $L(f)$ with variance $\mathbf{E}|L(f)|^2 = C(f,f) =: V_f$ and $\mathbf{E}\,L(f)^2 = C(\overline{f},f)$, where

$$\begin{aligned} C\,(g,f) := {}& \frac{1}{4\pi}\langle\nabla g,\nabla f\rangle_{L^2(\mathbf{D})} + \frac{1}{2}\langle g,f\rangle_{\dot{H}^{1/2}(\partial\mathbf{D})} \\ &+ \kappa_4\left(\frac{1}{\pi}\int_{\mathbf{D}}\overline{g(z)}\,\mathrm{d}^2 z - \frac{1}{2\pi}\int_0^{2\pi}\overline{g(e^{\mathrm{i}\theta})}\,\mathrm{d}\theta\right) \\ &\times\left(\frac{1}{\pi}\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z - \frac{1}{2\pi}\int_0^{2\pi} f(e^{\mathrm{i}\theta})\,\mathrm{d}\theta\right). \end{aligned} \tag{6.9}$$

More precisely, any finite moment of $L_n(f)$ converges at a rate $n^{-c(k)}$, for some small $c(k)>0$, i.e.

$$\mathbf{E}\,L_n(f)^k\overline{L_n(f)}^l = \mathbf{E}\,L(f)^k\overline{L(f)}^l + \mathcal{O}\left(n^{-c(k+l)}\right). \tag{6.10}$$

Moreover, the expectation in (6.6) *is given by*

$$\mathbf{E}\sum_{i=1}^n f(\sigma_i) = \frac{n}{\pi}\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z - \frac{\kappa_4}{\pi}\int_{\mathbf{D}} f(z)(2|z|^2-1)\,\mathrm{d}^2 z + \mathcal{O}\left(n^{-c}\right) \tag{6.11}$$

for some small constant $c>0$. The implicit constants in the error terms in (6.10)–(6.11) *depend on the $H^{2+\delta}$-norm of f and C_p from* (6.4).

Remark 6.2.2 (V_f is strictly positive)**.** *The variance $V_f = \mathbf{E}|L(f)|^2$ in Theorem 6.2.1 is strictly positive. Indeed, by the Cauchy-Schwarz inequality it follows that*

$$\left|\frac{1}{\pi}\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z - \frac{1}{2\pi}\int_0^{2\pi} f(e^{i\theta})\,d\theta\right|^2 \le \frac{1}{8\pi}\int_{\mathbf{D}}|\nabla f|^2\,\mathrm{d}^2 z.$$

Hence, since $\kappa_4\ge -1$ in (6.9), *this shows that*

$$V_f \ge \frac{1}{8\pi}\int_{\mathbf{D}}|\nabla f|^2\,\mathrm{d}^2 z + \frac{1}{2}\|f\|_{\dot{H}^{1/2}(\partial\mathbf{D})} > 0.$$

By polarisation, a multivariate Central Limit Theorem readily follows from Theorem 6.2.1:

Corollary 6.2.3. *Let X be an $n \times n$ i.i.d. complex matrix satisfying Assumption (6.A), and let $L_n(f)$ be defined in* (6.6). *For a fixed open bounded complex domain Ω with $\overline{\mathbf{D}} \subset \Omega \subset \mathbf{C}$, $\delta > 0$, $p \in \mathbf{N}$ and for any finite collection of test functions $f^{(1)}, \dots, f^{(p)} \in H_0^{2+\delta}(\Omega)$ the vector*

$$(L_n(f^{(1)}), \dots, L_n(f^{(p)})) \Longrightarrow (L(f^{(1)}), \dots, L(f^{(p)})), \tag{6.12}$$

converges to a multivariate complex Gaussian of zero expectation $\mathbf{E}\, L(f) = 0$ and covariance $\mathbf{E}\, L(f)\overline{L(g)} = \mathbf{E}\, L(f)L(\overline{g}) = C(f,g)$ with C as in (6.9). *Moreover, for any mixed k-moments we have an effective convergence rate of order $n^{-c(k)}$, as in* (6.10)

Remark 6.2.4. *We may compare Theorem 6.2.1 with the previous results in [164, Theorem 1] and [162, Theorem 1.1]:*

1. *Note that for a single $f\colon \mathbf{C} \to \mathbf{R}$ in the Ginibre case, i.e. $\kappa_4 = 0$, Theorem 6.2.1 implies [164, Theorem 1] with $\sigma_f^2 + \widetilde{\sigma}_f^2 = C(f,f)$, using the notation therein and with $C(f,f)$ defined in* (6.9).

2. *If additionally f is complex analytic in a neighbourhood of $\overline{\mathbf{D}}$, using the notation $\partial := \partial_z$, the expressions in* (6.9),(6.11) *of Theorem 6.2.1 simplify to*

$$\mathbf{E} \sum_{i=1}^{n} f(\sigma_i) = n f(0) + \mathcal{O}\left(n^{-\delta'}\right), \quad C(f,g) = \frac{1}{\pi} \int_{\mathbf{D}} \partial f(z) \overline{\partial g(z)}\, \mathrm{d}^2 z, \tag{6.13}$$

where we used that for any f, g complex analytic in a neighbourhood of $\overline{\mathbf{D}}$ we have

$$\frac{1}{2\pi} \int_{\mathbf{D}} \langle \nabla g, \nabla f \rangle\, \mathrm{d}^2 z = \frac{1}{\pi} \int_{\mathbf{D}} \partial f(z) \overline{\partial g(z)}\, \mathrm{d}^2 z = \sum_{k \in \mathbf{Z}} |k| \widehat{f \restriction_{\partial \mathbf{D}}}(k) \overline{\widehat{g \restriction_{\partial \mathbf{D}}}(k)}, \tag{6.14}$$

and that

$$\frac{1}{\pi} \int_{\mathbf{D}} f(z)\, \mathrm{d}^2 z = \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\theta})\, \mathrm{d}\theta = f(0).$$

The second equality in (6.14) *follows by writing f and g in Fourier series. The result in* (6.13) *exactly agrees with [162, Theorem 1.1].*

Remark 6.2.5 (Mesoscopic regime). *We formulated our result for* macroscopic *linear statistics, i.e. for test functions f that are independent of n. One may also consider* mesoscopic *linear statistics as well when $f(\sigma)$ is replaced with $\varphi(n^a(\sigma - z_0))$ for some fixed scale $a > 0$, reference point $z_0 \in \mathbf{D}$ and function $\varphi \in H^{2+\delta}(\mathbf{C})$. Our proof can directly handle this situation as well for any small $a \leq 1/500$[1], say, since all our error terms are effective as a small power of $1/n$. For $a > 0$ the leading term to the variance V_f comes solely from the $\|\nabla f\|^2$ term in* (6.9), *in particular the effect of the fourth cumulant is negligible.*

[1] The upper bound $1/500$ for a is a crude overestimate, we did not optimise it along the proof. The actual value of a comes from the fact that it has to be smaller than ω_d (see of Proposition 6.3.5) and from Lemma 6.7.9 (which is the main input of Proposition 6.3.5) it follows that $\omega_d \leq 1/100$.

6.2.1 Connection to the Gaussian free field

It has been observed in [164] that for the special case $\kappa_4 = 0$ the limiting random field $L(f)$ can be viewed as a variant of the *Gaussian free field* [178]. The Gaussian free field on some bounded domain $\Omega \subset \mathbf{C}$ can formally be defined as a *Gaussian Hilbert space* of random variables $h(f)$ indexed by functions in the homogeneous Sobolev space $f \in \dot{H}^1_0(\Omega)$ such that the map $f \mapsto h(f)$ is linear and

$$\mathbf{E}\, h(f) = 0, \quad \mathbf{E}\, \overline{h(f)} h(g) = \langle f, g \rangle_{\dot{H}^1(\Omega)}. \tag{6.15}$$

Here for $\Omega \subset \mathbf{C}$ we defined the homogeneous Sobolev space $\dot{H}^1_0(\Omega)$ as the completion of smooth compactly supported function $C_c^\infty(\Omega)$ with respect to the semi-inner product

$$\langle g, f \rangle_{\dot{H}^1(\Omega)} := \langle \nabla g, \nabla f \rangle_{L^2(\Omega)}, \qquad \|f\|^2_{\dot{H}^1(\Omega)} := \langle f, f \rangle_{\dot{H}^1(\Omega)}.$$

By the Poincaré inequality the space $\dot{H}^1_0(\Omega)$ is in fact a Hilbert space and as a vector space coincides with the usual Sobolev space $H^1_0(\Omega)$ with an equivalent norm but a different scalar product.

Since $\overline{\mathbf{D}} \subset \Omega$, the Sobolev space $\dot{H}^1_0(\Omega)$ can be orthogonally decomposed as

$$\dot{H}^1_0(\Omega) = \dot{H}^1_0(\mathbf{D}) \oplus \dot{H}^1_0(\overline{\mathbf{D}}^c) \oplus \dot{H}^1_0((\partial\mathbf{D})^c)^\perp,$$

where the complements are understood as the complements within Ω. The orthogonal complement $\dot{H}^1_0((\partial\mathbf{D})^c)^\perp$ is (see e.g. [178, Thm. 2.17]) given by the closed subspace of functions which are harmonic in $\mathbf{D} \cup \overline{\mathbf{D}}^c = (\partial\mathbf{D})^c$, i.e. away from the unit circle. For closed subspaces $S \subset \dot{H}^1_0(\Omega)$ we denote the orthogonal projection onto S by P_S. Then by orthogonality and conformal symmetry it follows [164, Lemma 3.1][2] that

$$\begin{aligned} \left\| P_{\dot{H}^1_0(\mathbf{D})} f + P_{\dot{H}^1_0((\partial\mathbf{D})^c)^\perp} f \right\|^2_{\dot{H}^1(\Omega)} &= \|f\|^2_{\dot{H}^1(\mathbf{D})} + \|P_{\dot{H}^1_0((\partial\mathbf{D})^c)^\perp} f\|^2_{\dot{H}^1(\mathbf{D})} \\ &= \|f\|^2_{\dot{H}^1(\mathbf{D})} + 2\pi \|f\|^2_{\dot{H}^{1/2}(\partial\mathbf{D})}, \end{aligned} \tag{6.16}$$

where we canonically identify $f \in \dot{H}^1_0(\Omega)$ with its restriction to $\mathbf{D}$. If $\kappa_4 = 0$, then the rhs. of (6.16) is precisely $4\pi C(f,f)$ and therefore $L(f)$ can be interpreted [164, Corollary 1.2] as the projection

$$L = (4\pi)^{-1/2} P h, \qquad P := \left(P_{\dot{H}^1_0(\mathbf{D})} + P_{\dot{H}^1_0((\partial\mathbf{D})^c)^\perp} \right) \tag{6.17}$$

of the Gaussian free field h onto $\dot{H}^1_0(\mathbf{D}) \oplus \dot{H}^1_0((\partial\mathbf{D})^c)^\perp$, i.e. the Gaussian free field conditioned to be harmonic in $\mathbf{D}^c$. The projection (6.17) is defined via duality, i.e. $(Ph)(f) := h(Pf)$ so that indeed

$$\mathbf{E} \left| \left[\frac{1}{\sqrt{4\pi}} P h \right] (f) \right|^2 = \frac{1}{4\pi} \left(\|f\|^2_{\dot{H}^1(\mathbf{D})} + 2\pi \|f\|^2_{\dot{H}^{1/2}(\partial\mathbf{D})} \right) = C(f,f) = \mathbf{E} |L(f)|^2.$$

[2] In Eq. (3.1), and in the last displayed equation of the proof of Lemma 3.1 factors of 2 are missing. In the notation of [164] the correct equations read

$$\frac{1}{2} \|P_H f\|^2_{H^1(\mathbf{C})} = \|P_H f\|^2_{H^1(\mathbf{U})} = 2\pi \|f\|^2_{H^{1/2}(\partial\mathbf{U})} \quad \text{and} \quad \langle g_1, g_2 \rangle_{H^1(\mathbf{U})} = 2\pi \langle g_1, g_2 \rangle_{H^{1/2}(\partial\mathbf{U})}.$$

If $\kappa_4 > 0$, then L can be interpreted as the sum

$$L = \frac{1}{\sqrt{4\pi}} Ph + \sqrt{\kappa_4}\Big(\langle\cdot\rangle_{\mathbf{D}} - \langle\cdot\rangle_{\partial\mathbf{D}}\Big)\Xi \tag{6.18}$$

of the Gaussian free field Ph conditioned to be harmonic in $\mathbf{D}^c$, and an independent standard real Gaussian Ξ multiplied by difference of the averaging functionals $\langle\cdot\rangle_{\mathbf{D}}$, $\langle\cdot\rangle_{\partial\mathbf{D}}$ on $\mathbf{D}$ and $\partial\mathbf{D}$. For $\kappa_4 < 0$ there seems to be no direct interpretation of L similar to (6.18).

6.3 Proof strategy

For the proof of Theorem 6.2.1 we study the $2n \times 2n$ matrix H^z defined in (6.2), that is the Hermitisation of $X - z$. Denote by $\{\lambda_{\pm i}^z\}_{i=1}^n$ the eigenvalues of H^z labelled in an increasing order (we omit the index $i = 0$ for notational convenience). As a consequence of the block structure of H^z its spectrum is symmetric with respect to zero, i.e. $\lambda_{-i}^z = -\lambda_i^z$ for any $i \in [n]$.

Let $G(w) = G^z(w) := (H^z - w)^{-1}$ denote the resolvent of H^z with $\eta = \Im w \neq 0$. It is well known (e.g. see [11, 13]) that G^z becomes approximately deterministic, as $n \to \infty$, and its limit is expressed via the unique solution of the scalar equation

$$-\frac{1}{m^z} = w + m^z - \frac{|z|^2}{w + m^z}, \quad \eta\Im m^z(w) > 0, \quad \eta = \Im w \neq 0, \tag{6.19}$$

which is a special case of the *matrix Dyson equation* (MDE), see e.g. [5]. We note that on the imaginary axis $m^z(\mathrm{i}\eta) = \mathrm{i}\Im m^z(\mathrm{i}\eta)$. To find the limit of G^z we define a $2n \times 2n$ block-matrix

$$M^z(w) := \begin{pmatrix} m^z(w) & -zu^z(w) \\ -\bar{z}u^z(w) & m^z(w) \end{pmatrix}, \quad u^z(w) := \frac{m^z(w)}{w + m^z(w)}, \tag{6.20}$$

where each block is understood to be a scalar multiple of the $n \times n$ identity matrix. We note that m, u, M are uniformly bounded in z, w, i.e.

$$\|M^z(w)\| + |m^z(w)| + |u^z(w)| \lesssim 1. \tag{6.21}$$

Indeed, taking the imaginary part of (6.19) we have (dropping z, w)

$$\beta_* \Im m = (1 - \beta_*)\Im w, \qquad \beta_* := 1 - |m|^2 - |u|^2|z|^2, \tag{6.22}$$

which implies

$$|m|^2 + |u|^2|z|^2 < 1, \tag{6.23}$$

as $\Im m$ and $\Im w$ have the same sign. Note that (6.23) saturates if $\Im w \to 0$ and $\Re w$ is in the support of the *self-consistent density of states*, $\rho^z(E) := \pi^{-1}\Im m^z(E + \mathrm{i}0)$. Moreover, (6.19) is equivalent to $u = -m^2 + u^2|z|^2$, thus $|u| < 1$ and (6.21) follows.

For our analysis the derivative $m'(w)$ in the w-variable plays a central role and we note that by taking the derivative of (6.19) we obtain

$$m' = \frac{1 - \beta}{\beta}, \qquad \beta := 1 - m^2 - u^2|z|^2. \tag{6.24}$$

On the imaginary axis, $w = \mathrm{i}\eta$, where by taking the real part of (6.19) it follows that $\Re m(\mathrm{i}\eta) = 0$, we can use [13, Eq. (3.13)]

$$\Im m(\mathrm{i}\eta) \sim \begin{cases} \eta^{1/3} + |1-|z|^2|^{1/2} & \text{if} \quad |z| \le 1, \\ \frac{\eta}{|z|^2-1+\eta^{2/3}} & \text{if} \quad |z| > 1, \end{cases} \qquad \eta \lesssim 1 \tag{6.25}$$

to obtain asymptotics for

$$\beta_* \sim \frac{\eta}{\Im m}, \quad \beta = \beta_* + 2(\Im m)^2, \qquad \eta \lesssim 1. \tag{6.26}$$

The optimal local law from Theorem [11, Theorem 5.2] and [13, Theorem 5.2][3], which for the application in Girko's formula (6.3) is only needed on the imaginary axis, asserts that $G^z \approx M^z$ in the following sense:

Theorem 6.3.1 (Optimal local law for G)**.** *The resolvent G^z is very well approximated by the deterministic matrix M^z in the sense*

$$|\langle (G^z(\mathrm{i}\eta) - M^z(\mathrm{i}\eta))A\rangle| \le \frac{\|A\| n^\xi}{n\eta}, \qquad |\langle \boldsymbol{x}, (G^z(\mathrm{i}\eta) - M^z(\mathrm{i}\eta))\boldsymbol{y}\rangle| \le \frac{\|\boldsymbol{x}\|\|\boldsymbol{y}\| n^\xi}{\sqrt{n\eta}}, \tag{6.27}$$

with very high probability, uniformly for $\eta > 0$ and for any deterministic matrices and vectors $A, \boldsymbol{x}, \boldsymbol{y}$.

The matrix H^z can be related to the linear statistics of eigenvalues σ_i of X via the precise (regularised) version of Girko's Hermitisation formula (6.3)

$$\begin{aligned} L_n(f) &= \frac{1}{4\pi}\int_{\mathbf{C}} \Delta f(z) \Big[\log|\det(H^z - \mathrm{i}T)| - \mathbf{E}\log|\det(H^z - \mathrm{i}T)| \Big] \,\mathrm{d}^2 z \\ &\quad - \frac{n}{2\pi \mathrm{i}} \int_{\mathbf{C}} \Delta f \left[\left(\int_0^{\eta_0} + \int_{\eta_0}^{\eta_c} + \int_{\eta_c}^{T} \right) [\langle G^z(\mathrm{i}\eta) - \mathbf{E}\, G^z(\mathrm{i}\eta)\rangle] \,\mathrm{d}\eta \right] \mathrm{d}^2 z \\ &=: J_T + I_0^{\eta_0} + I_{\eta_0}^{\eta_c} + I_{\eta_c}^T, \end{aligned} \tag{6.28}$$

for

$$\eta_0 := n^{-1-\delta_0}, \quad \eta_c := n^{-1+\delta_1}, \tag{6.29}$$

and some very large $T > 0$, say $T = n^{100}$. Note that in (6.28) we used that $\langle G^z(\mathrm{i}\eta)\rangle = \mathrm{i}\langle \Im G^z(\mathrm{i}\eta)\rangle$ by spectral symmetry. The test function $f \colon \mathbf{C} \to \mathbf{C}$ is in $H^{2+\delta}$ and it is compactly supported. J_T in (6.28) consists of the first line in the rhs., whilst $I_0^{\eta_0}, I_{\eta_0}^{\eta_c}, I_{\eta_c}^T$ corresponds to the three different η-regimes in the second line of the rhs. of (6.28).

Remark 6.3.2. *We remark that in* (6.28) *we split the η-regimes in a different way compared to [59, Eq. (32)]. We also use a different notation to identify the η-scales: here we use the notation $J_T, I_0^{\eta_0}, I_{\eta_0}^{\eta_c}, I_{\eta_c}^T$, whilst in [59, Eq. (32)] we used the notation I_1, I_2, I_3, I_4.*

[3] The local laws in [11, Theorem 5.2] and [13, Theorem 5.2] have been proven for $\eta \ge \eta_f(z)$, with $\eta_f(z)$ being the fluctuation scale defined in [13, Eq. (5.2)], but they can be easily extend to any $\eta > 0$ by a standard argument, see [59, Appendix A].

The different regimes in (6.28) will be treated using different techniques. More precisely, the integral J_T is easily estimated as in [13, Proof of Theorem 2.3], which uses similar computations to [11, Proof of Theorem 2.5]. The term $I_0^{\eta_0}$ is estimated using the fact that with high probability there are no eigenvalues in the regime $[0, \eta_0]$; this follows by [196, Theorem 3.2]. Alternatively (see Remark 6.4.2 and Remark 6.4.5 later), the contribution of the regime $I_0^{\eta_0}$ can be estimated without resorting to the quite sophisticated proof of [196, Theorem 3.2] if the entries of X satisfy the additional assumption (6.37). More precisely, this can be achieved using [11, Proposition 5.7] (which follows adapting the proof of [34, Lemma 4.12]) to bound the very small regime $[0, n^{-l}]$, for some large $l \in \mathbf{N}$, and then using [61, Corollary 4] to bound the regime $[n^{-l}, \eta_0]$.

The main novel work is done for the integrals $I_{\eta_0}^{\eta_c}$ and $I_{\eta_c}^{T}$. The main contribution to $L_n(f)$ comes from the mesoscopic regime in $I_{\eta_c}^{T}$, which is analysed using the following Central Limit Theorem for resolvents.

Proposition 6.3.3 (CLT for resolvents). *Let $\epsilon, \xi > 0$ be arbitrary. Then for $z_1, \dots, z_p \in \mathbf{C}$ and $\eta_1, \dots, \eta_p \ge n^{\xi - 1} \max_{i \ne j} |z_i - z_j|^{-2}$, denoting the pairings on $[p]$ by Π_p, we have*

$$\begin{aligned}\mathbf{E} \prod_{i \in [p]} \langle G_i - \mathbf{E}\, G_i \rangle &= \sum_{P \in \Pi_p} \prod_{\{i,j\} \in P} \mathbf{E} \langle G_i - \mathbf{E}\, G_i \rangle \langle G_j - \mathbf{E}\, G_j \rangle + \mathcal{O}(\Psi) \\ &= \frac{1}{n^p} \sum_{P \in \Pi_p} \prod_{\{i,j\} \in P} \frac{V_{i,j} + \kappa_4 U_i U_j}{2} + \mathcal{O}(\Psi),\end{aligned} \tag{6.30}$$

where $G_i = G^{z_i}(\mathrm{i}\eta_i)$,

$$\Psi := \frac{n^{\epsilon}}{(n\eta_*)^{1/2}} \frac{1}{\min_{i \ne j} |z_i - z_j|^4} \prod_{i \in [p]} \frac{1}{|1 - |z_i|| n \eta_i}, \tag{6.31}$$

$\eta_* := \min_i \eta_i$, and $V_{i,j} = V_{i,j}(z_i, z_j, \eta_i, \eta_j)$ and $U_i = U_i(z_i, \eta_i)$ are defined as

$$\begin{aligned} V_{i,j} &:= \frac{1}{2} \partial_{\eta_i} \partial_{\eta_j} \log \left[1 + (u_i u_j |z_i||z_j|)^2 - m_i^2 m_j^2 - 2 u_i u_j \Re z_i \overline{z_j}\right], \\ U_i &:= \frac{\mathrm{i}}{\sqrt{2}} \partial_{\eta_i} m_i^2, \end{aligned} \tag{6.32}$$

with $m_i = m^{z_i}(\mathrm{i}\eta_i)$ and $u_i = u^{z_i}(\mathrm{i}\eta_i)$.

Moreover, the expectation of G is given by

$$\langle \mathbf{E}\, G \rangle = \langle M \rangle - \frac{\mathrm{i}\kappa_4}{4n} \partial_\eta (m^4) + \mathcal{O}\Big(\frac{1}{|1 - |z|| n^{3/2} (1 + \eta)} + \frac{1}{|1 - |z|| (n\eta)^2} \Big). \tag{6.33}$$

Remark 6.3.4. *In Section 6.4 we will apply this proposition in the regime where $\min_{i \ne j} |z_i - z_j|$ is quite large, i.e. it is at least $n^{-\delta}$, for some small $\delta > 0$, hence we did not optimise the estimates for the opposite regime. However, using the more precise [60, Lemma 6.1] instead of Lemma 6.6.1 within the proof, one can immediately strengthen Proposition 6.3.3 on two accounts. First, the condition on $\eta_* = \min \eta_i$ can be relaxed to*

$$\eta_* \gtrsim n^{\xi - 1} \Big(\min_{i \ne j} |z_i - z_j|^2 + \eta_* \Big)^{-1}.$$

Second, the denominator $\min_{i \ne j} |z_i - z_j|^4$ in (6.31) can be improved to

$$\Big(\min_{i \ne j} |z_i - z_j|^2 + \eta_* \Big)^2.$$

In order to show that the contribution of $I_{\eta_0}^{\eta_c}$ to $L_n(f)$ is negligible, in Proposition 6.3.5 we prove that $\langle G^{z_1}(\mathrm{i}\eta_1)\rangle$ and $\langle G^{z_2}(\mathrm{i}\eta_2)\rangle$ are asymptotically independent if z_1, z_2 are far enough from each other, they are well inside $\mathbf{D}$, and $\eta_0 \le \eta_1, \eta_2 \le \eta_c$.

Proposition 6.3.5 (Independence of resolvents with small imaginary part)**.** *Fix $p \in \mathbf{N}$. For any sufficiently small $\omega_d, \omega_h, \omega_f > 0$ such that $\omega_h \ll \omega_f$, there exist $\omega, \widehat{\omega}, \delta_0, \delta_1 > 0$ such that $\omega_h \ll \delta_m \ll \widehat{\omega} \ll \omega \ll \omega_f$, for $m = 0, 1$, such that for any $|z_l| \le 1 - n^{-\omega_h}$, $|z_l - z_m| \ge n^{-\omega_d}$, with $l, m \in [p]$, $l \neq m$, it holds*

$$\mathbf{E}\prod_{l=1}^{p}\langle G^{z_l}(\mathrm{i}\eta_l)\rangle = \prod_{l=1}^{p}\mathbf{E}\langle G^{z_l}(\mathrm{i}\eta_l)\rangle + \mathcal{O}\left(\frac{n^{p(\omega_h+\delta_0)+\delta_1}}{n^{\omega}} + \frac{n^{\omega_f+3\delta_0}}{\sqrt{n}}\right), \tag{6.34}$$

for any $\eta_1, \dots, \eta_p \in [n^{-1-\delta_0}, n^{-1+\delta_1}]$.

The paper is organised as follows: In Section 6.4 we conclude Theorem 6.2.1 by combining Propositions 6.3.3 and 6.3.5. In Section 6.5 we prove a local law for G_1AG_2, for a deterministic matrix A. In Section 6.6, using the result in Section 6.5 as an input, we prove Proposition 6.3.3, the Central Limit Theorem for resolvents. In Section 6.7 we prove Proposition 6.3.5 using the fact that the correlation among small eigenvalues of H^{z_1}, H^{z_2} is "small", if z_1, z_2 are far from each other, as a consequence of the local law in Section 6.5.

6.4 Central limit theorem for linear statistics

In this section, using Proposition 6.3.3–6.3.5 as inputs, we prove our main result Theorem 6.2.1.

6.4.1 Preliminary reductions in Girko's formula

In this section we prove that the main contribution to $L_n(f)$ in (6.28) comes from the regime $I_{\eta_c}^{T}$. This is made rigorous in the following lemma.

Lemma 6.4.1. *Fix $p \in \mathbf{N}$ and some bounded open $\overline{\mathbf{D}} \subset \Omega \subset \mathbf{C}$, and for any $l \in [p]$ let $f^{(l)} \in H_0^{2+\delta}(\Omega)$. Then*

$$\mathbf{E}\prod_{l=1}^{p} L_n(f^{(l)}) = \mathbf{E}\prod_{l=1}^{p} I_{\eta_c}^{T}(f^{(l)}) + \mathcal{O}\left(n^{-c(p)}\right), \tag{6.35}$$

for some small $c(p) > 0$, with $L_n(f^{(l)})$ and $I_{\eta_c}^{T}(f^{(l)})$ defined in (6.28)*. The constant in $\mathcal{O}(\cdot)$ may depend on p and on the L^2-norm of $\Delta f^{(1)}, \dots, \Delta f^{(p)}$.*

Remark 6.4.2. *In the remainder of this section we need to ensure that with high probability the matrix H^z, defined in* (6.2)*, does not have eigenvalues very close to zero, i.e. that*

$$\mathbf{P}\left(\mathrm{Spec}(H^z) \cap \left[-n^{-l}, n^{-l}\right] \neq \emptyset\right) \le C_l n^{-l/2}, \tag{6.36}$$

for any $l \ge 2$ uniformly in $|z| \le 1$. The bound (6.36) *directly follows from [196, Theorem 3.2]. Alternatively,* (6.36) *follows by [11, Proposition 5.7] (which follows adapting the proof of [34, Lemma 4.12]), without recurring to the quite sophisticated proof of [196, Theorem 3.2], under the* *additional assumption that there exist $\alpha, \beta > 0$ such that the random variable χ has a density $g : \mathbf{C} \to [0, \infty)$ which satisfies*

$$g \in L^{1+\alpha}(\mathbf{C}), \qquad \|g\|_{L^{1+\alpha}(\mathbf{C})} \le n^{\beta}. \tag{6.37}$$

We start proving *a priori* bounds for the integrals defined in (6.28).

Lemma 6.4.3. *Fix some bounded open $\overline{\mathbf{D}} \subset \Omega \subset \mathbf{C}$ and let $f \in H_0^{2+\delta}(\Omega)$. Then for any $\xi > 0$ the bounds*

$$|J_T| \le \frac{n^{1+\xi}\|\Delta f\|_{L^1(\Omega)}}{T^2}, \qquad |I_0^{\eta_0}| + \left|I_{\eta_0}^{\eta_c}\right| + |I_{\eta_c}^T| \le n^{\xi}\|\Delta f\|_{L^2(\Omega)}|\Omega|^{1/2}, \tag{6.38}$$

hold with very high probability, where $|\Omega|$ denotes the Lebesgue measure of the set Ω.

Proof. The proof of the bound for J_T is identical to [13, Proof of Theorem 2.3] and so omitted.

The bound for $I_0^{\eta_0}, I_{\eta_0}^{\eta_c}, I_{\eta_c}^T$ relies on the local law of Theorem 6.3.1. More precisely, by Theorem 6.3.1 and (6.33) of Proposition 6.3.3 it follows that

$$|\langle G^z - \mathbf{E}\, G^z\rangle| \le \frac{n^{\xi}}{n\eta}, \tag{6.39}$$

with very high probability uniformly in $\eta > 0$ and $|z| \le C$ for some large $C > 0$. First of all we remove the regime $[0, n^{-l}]$ by [196, Theorem 3.2], i.e. its contribution is smaller than n^{-l}, for some large $l \in \mathbf{N}$, with very high probability. Alternatively, this can be achieved by [11, Proposition 5.7] under the additional assumption (6.37) in Remark 6.4.2. Then for any $a, b \ge n^{-l}$, by (6.39), we have

$$n\left|\int_{\Omega} \mathrm{d}^2 z \Delta f(z) \int_a^b \mathrm{d}\eta \big[\langle G(\mathrm{i}\eta) - \mathbf{E}\, G(\mathrm{i}\eta)\rangle\big]\right| \lesssim n^{\xi}|\Omega|^{1/2}\|\Delta f\|_{L^2(\Omega)}, \tag{6.40}$$

with very high probability. This concludes the proof of the second bound in (6.38). □

We have a better bound for $I_0^{\eta_0}$, $I_{\eta_0}^{\eta_c}$ which holds true in expectation.

Lemma 6.4.4. *Fix some bounded open $\overline{\mathbf{D}} \subset \Omega \subset \mathbf{C}$ and let $f \in H_0^{2+\delta}(\Omega)$. Then there exists $\delta' > 0$ such that*

$$\mathbf{E}|I_0^{\eta_0}| + \mathbf{E}\left|I_{\eta_0}^{\eta_c}\right| \le n^{-\delta'}\|\Delta f\|_{L^2(\Omega)}. \tag{6.41}$$

Proof of Lemma 6.4.1. Lemma 6.4.1 readily follows (see e.g. [59, Lemma 4.2]) combining Lemma 6.4.3 and Lemma 6.4.4. □

We conclude this section with the proof of Lemma 6.4.4.

Proof of Lemma 6.4.4. The bound for $\mathbf{E}|I_0^{\eta_0}|$ immediately follows by [196, Theorem 3.2] (see also Remark 6.4.5 for an alternative proof).

By the local law outside the spectrum, given in the second part of [13, Theorem 5.2], it follows that for $0 < \gamma < 1/2$ we have

$$|\langle G^z(\mathrm{i}\eta) - M^z(\mathrm{i}\eta)\rangle| \le \frac{n^{\xi}}{n^{1+\gamma/3}\eta}, \tag{6.42}$$

uniformly for all $|z|^2 \geq 1+(n^\gamma\eta)^{2/3}+n^{(\gamma-1)/2}$, $\eta > 0$, and $|z| \leq 1+\tau^*$, for some $\tau^* \sim 1$. We remark that the local law (6.42) was initially proven only for η above the fluctuation scale $\eta_f(z)$, which is defined in [13, Eq. (5.2)], but it can be easily extend to any $\eta > 0$ using the monotonicity of the function $\eta \mapsto \eta\langle \Im G(i\eta)\rangle$ and the fact that

$$\left|n^\xi \eta_f(z)\langle M^z(\mathrm{i}n^\xi\eta_f(z))\rangle\right| + |\eta\langle M^z(\mathrm{i}\eta)\rangle| \lesssim n^{2\xi}\frac{\eta_f(z)^2}{|z|^2-1}, \tag{6.43}$$

uniformly in $\eta > 0$, since $\Im M^z(\mathrm{i}\eta) = \Im m^z(\mathrm{i}\eta)I$ by (6.20), with I the $2n \times 2n$ identity matrix, and $\Im m^z(\mathrm{i}\eta) \leq \eta(|z|^2-1)^{-1}$ by [13, Eq. (3.13)]. Note that we assumed the additional term $n^{(\gamma-1)/2}$ in the lower bound for $|z|^2$ compared with [13, Theorem 5.2] in order to ensure that the rhs. in (6.43), divided by η, is smaller than the error term in (6.42).

Next, in order to bound $\mathbf{E}|I_{\eta_0}^{\eta_c}|$, we consider

$$\begin{aligned}\mathbf{E}|I_{\eta_0}^{\eta_c}|^2 &= -\frac{n^2}{4\pi^2}\int_{\mathbf{C}} \mathrm{d}^2 z_1(\Delta f)(z_1)\int_{\mathbf{C}} \mathrm{d}^2 z_2(\Delta \overline{f})(z_2)\int_{\eta_0}^{\eta_c}\mathrm{d}\eta_1\int_{\eta_0}^{\eta_c}\mathrm{d}\eta_2\, F \\ F = F(z_1,z_2,\eta_1,\eta_2) &:= \mathbf{E}\left[\langle G^{z_1}(\mathrm{i}\eta_1)-\mathbf{E}\,G^{z_1}(\mathrm{i}\eta_1)\rangle\langle G^{z_2}(\mathrm{i}\eta_2)-\mathbf{E}\,G^{z_2}(\mathrm{i}\eta_2)\rangle\right].\end{aligned} \tag{6.44}$$

By (6.40) it follows that the regimes $1-n^{-2\omega_h} \leq |z_l|^2 \leq 1+n^{-2\omega_h}$, with $l=1,2$, and $|z_1-z_2| \leq n^{-\omega_d}$ in (6.44), with ω_h, ω_d defined in Proposition 6.3.5, are bounded by $n^{-2\omega_h+\xi}$ and $n^{-\omega_d/2+\xi}$, respectively. Moreover, the contribution from the regime $|z_l| \geq 1+n^{-2\omega_h}$ is also bounded by $n^{-2\omega_h+\xi}$ using (6.42) with $\gamma \leq 1-3\omega_h-2\delta_1$, say $\gamma = 1/4$. After collecting these error terms we conclude that

$$\begin{aligned}\mathbf{E}|I_{\eta_0}^{\eta_c}|^2 = \frac{n^2}{4\pi^2}&\int_{|z_1|\leq 1-n^{-\omega_h}} \mathrm{d}^2 z_1 \Delta f(z_1)\int_{\substack{|z_2|\leq 1-n^{-\omega_h},\\ |z_2-z_1|\geq n^{-\omega_d}}} \mathrm{d}^2 z_2 \Delta\overline{f(z_2)} \\ &\times \int_{\eta_0}^{\eta_c}\mathrm{d}\eta_1\int_{\eta_0}^{\eta_c}\mathrm{d}\eta_2 F + \mathcal{O}\left(\frac{n^\xi}{n^{\omega_h}}+\frac{n^\xi}{n^{\omega_d/2}}\right).\end{aligned} \tag{6.45}$$

We remark that the implicit constant in $\mathcal{O}(\cdot)$ in (6.45) and in the remainder of the proof may depend on $\|\Delta f\|_{L^2(\Omega)}$.

Then by Proposition 6.3.5 it follows that

$$\mathbf{E}\left[\langle G^{z_1}(\mathrm{i}\eta_1)-\mathbf{E}\langle G^{z_1}(\mathrm{i}\eta_1)\rangle\langle G^{z_2}(\mathrm{i}\eta_2)-\mathbf{E}\,G^{z_2}(\mathrm{i}\eta_2)\rangle\right] = \mathcal{O}\left(\frac{n^{c(\omega_h+\delta_0)+\delta_1}}{n^\omega}\right), \tag{6.46}$$

with $\omega_h \ll \delta_0 \ll \omega$. Hence, plugging (6.46) into (6.45) it follows that

$$\mathbf{E}|I_{\eta_0}^{\eta_c}|^2 = \mathcal{O}\left(\frac{n^{c(\omega_h+\delta_0)+2\delta_1}}{n^\omega}\right). \tag{6.47}$$

This concludes the proof under the assumption $\omega_h \ll \delta_m \ll \omega$, with $m = 0,1$, of Proposition 6.3.5 (see Section 6.7.2.3 later for a summary on all the scales involved in the proof of Proposition 6.3.5). □

Remark 6.4.5 (Alternative proof of the bound for $\mathbf{E}|I_0^{\eta_0}|$)**.** *Under the additional assumption* (6.37) *in Remark 6.4.2, we can prove the same bound for* $\mathbf{E}|I_0^{\eta_0}|$ *in* (6.41) *without relying on the fairly sophisticated proof of [196, Theorem 3.2].*

In order to bound $\mathbf{E}|I_0^{\eta_0}|$ *we first remove the regime* $\eta \in [0, n^{-l}]$ *as in the proof of Lemma 6.4.3. Then, using* (6.40) *to bound the integral over the regime* $|1-|z|^2| \le 1+n^{-2\omega_h}$*, with* ω_h *defined in Proposition 6.3.5, and* (6.42) *for the regime* $|z|^2 \ge 1-n^{-2\omega_h}$*, we conclude that*

$$\mathbf{E}|I_0^{\eta_0}| = \mathbf{E}\,\frac{n}{2\pi}\int_{|z|\le 1-n^{-2\omega_h}} |\Delta f| \left|\int_0^{\eta_0} \langle G^z - \mathbf{E}\,G^z\rangle \, \mathrm{d}\eta\right| \mathrm{d}^2 z + \mathcal{O}\left(\frac{n^\xi}{n^{\omega_h}}\right). \tag{6.48}$$

By universality of the smallest eigenvalue of H^z *(which directly follows by Proposition 6.7.13 for any fixed* $|z|^2 \le 1-n^{-2\omega_h}$*; see also [54]), and the bound in [61, Corollary 2.4] we have that*

$$\mathbf{P}\left(\lambda_1^z \le \eta_0\right) \le n^{-\delta_0/4},$$

with $\eta_0 = n^{-1-\delta_0}$ *and* $\omega_h \ll \delta_0$*. This concludes the bound in* (6.41) *for* $I_0^{\eta_0}$ *following exactly the same proof of [59, Lemma 4.6], by* (6.48)*. We warn the reader that in [61, Corollary 2.4]* λ_1 *denotes the smallest eigenvalue of* $(X-z)(X-z)^*$*, whilst here* λ_1^z *denotes the smallest (positive) eigenvalue of* H^z*.*

6.4.2 Computation of the expectation in Theorem 6.2.1

In this section we compute the expectation $\mathbf{E}\sum_i f(\sigma_i)$ in (6.11) using the computation of $\mathbf{E}\langle G\rangle$ in (6.33) of Proposition 6.3.3 as an input. More precisely, we prove the following lemma. Note that (6.49) proves (6.11) in Theorem 6.2.1.

Lemma 6.4.6. *Fix some bounded open* $\overline{\mathbf{D}} \subset \Omega \subset \mathbf{C}$ *and let* $f \in H_0^{2+\delta}(\Omega)$*, and let* $\kappa_4 := n^2[\mathbf{E}|x_{11}|^4 - 2(\mathbf{E}|x_{11}|^2)]$*, then*

$$\mathbf{E}\sum_{i=1}^n f(\sigma_i) = \frac{n}{\pi}\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z - \frac{\kappa_4}{\pi}\int_{\mathbf{D}} f(z)(2|z|^2-1)\,\mathrm{d}^2 z + \mathcal{O}\left(n^{-\delta'}\right), \tag{6.49}$$

for some small $\delta' > 0$.

Proof. By the circular law (e.g. see [11, Eq. (2.7)], [13, Theorem 2.3]) it immediately follows that

$$\sum_{i=1}^n f(\sigma_i) - \frac{n}{\pi}\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z = \mathcal{O}(n^\xi), \tag{6.50}$$

with very high probability. Hence, in order to prove (6.49) we need to identify the sub-leading term in the expectation of (6.50), which is not present in the Ginibre case since $\kappa_4 = 0$.

First of all by Lemma 6.4.1 it follows that the main contribution in Girko's formula comes from $I_{\eta_c}^T$. Since the error term in (6.33) is not affordable for $1-|z|$ very close to zero, we remove the regime $|1-|z|^2| \le n^{-2\nu}$ in the z-integral by (6.40) at the expense of an error term $n^{-\nu+\xi}$, for some very small $\nu > 0$ we will choose shortly. The regime $|1-|z|^2| \ge n^{-2\nu}$, instead, is computed using (6.33). Hence, collecting these error terms we conclude that there exists $\delta' > 0$ such that

$$\begin{aligned}
&\mathbf{E}\sum_i f(\sigma_i) - \frac{n}{\pi}\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z \\
&\quad = -\frac{n}{2\pi\mathrm{i}}\int_{|1-|z|^2|\geq n^{-2\nu}} \mathrm{d}^2 z\Delta f \int_{\eta_c}^{T} \mathrm{d}\eta\, \mathbf{E}\langle G - M\rangle + \mathcal{O}\left(n^{-\delta'} + n^{-\nu+\xi}\right) \\
&\quad = \frac{\kappa_4}{8\pi}\int \mathrm{d}^2 z\Delta f \int_0^{\infty} \mathrm{d}\eta\, \partial_\eta(m^4) + \mathcal{O}\left(n^{-\delta'} + \frac{n^{2\nu}}{n\eta_c} + n^{2\nu}\eta_c + n^{-\nu+\xi}\right) \\
&\quad = -\frac{\kappa_4}{\pi}\int_{\mathbf{D}} f(z)(2|z|^2-1)\,\mathrm{d}^2 z + \mathcal{O}\left(n^{-\delta'} + \frac{n^{2\nu}}{n\eta_c} + n^{2\nu}\eta_c + n^{-\nu+\xi}\right),
\end{aligned} \tag{6.51}$$

with $\eta_c = n^{-1+\delta_1}$ defined in (6.29). To go from the second to the third line we used (6.33), and then we added back the regimes $\eta \in [0, \eta_c]$ and $\eta \geq T$, and the regime $|1-|z|^2| \leq n^{-2\nu}$ in the z-integration at the price of a negligible error. In particular, in the η-integration we used that $|\partial_\eta(m^4)| \lesssim n^{2\nu}$ in the regime $\eta \in [0, \eta_c]$, by (6.24)–(6.26), and that using $|m| \leq \eta^{-1}$ we have $|\partial_\eta(m^4)| \lesssim \eta^{-5}$ by (6.24), in the regime $\eta \geq T$. Choosing $\nu, \delta' > 0$ so that $\nu \ll \delta_1 \ll \delta'$ we conclude the proof of (6.49). □

6.4.3 Computation of the second and higher moments in Theorem 6.2.1

In this section we conclude the proof of Theorem 6.2.1, i.e. we compute

$$\begin{aligned}
\mathbf{E}\prod_{i\in[p]} L_n(f^{(i)}) &= \mathbf{E}\prod_{i\in[p]} I_{\eta_c}^{T}(f^{(i)}) + \mathcal{O}(n^{-c(p)}) \\
&= \mathbf{E}\prod_{i\in[p]}\left[-\frac{n}{2\pi\mathrm{i}}\int_{\mathbf{C}} \Delta f^{(i)}(z)\int_{\eta_c}^{T} \langle G^z(\mathrm{i}\eta) - \mathbf{E}\,G^z(\mathrm{i}\eta)\rangle\,\mathrm{d}\eta\,\mathrm{d}^2 z\right] \\
&\quad + \mathcal{O}(n^{-c(p)})
\end{aligned} \tag{6.52}$$

to leading order using (6.30).

Lemma 6.4.7. *Let $f^{(i)}$ be as in Theorem 6.2.1 and set $f^{(i)} = f$ or $f^{(i)} = \overline{f}$ for any $i \in [p]$, and recall that Π_p denotes the set of pairings on $[p]$. Then*

$$\begin{aligned}
&\mathbf{E}\prod_{i\in[p]}\left[-\frac{n}{2\pi\mathrm{i}}\int_{\mathbf{C}} \Delta f^{(i)}(z)\int_{\eta_c}^{T} \langle G^z(\mathrm{i}\eta) - \mathbf{E}\,G^z(\mathrm{i}\eta)\rangle\,\mathrm{d}\eta\,\mathrm{d}^2 z\right] \\
&= \sum_{P\in\Pi_p}\prod_{\{i,j\}\in P}\left[-\int_{\mathbf{C}} \mathrm{d}^2 z_i \Delta f^{(i)} \int_{\mathbf{C}} \mathrm{d}^2 z_j \Delta f^{(j)} \int_0^{\infty} \mathrm{d}\eta_i \int_0^{\infty} \mathrm{d}\eta_j \frac{V_{i,j} + \kappa_4 U_i U_j}{8\pi^2}\right] \\
&\quad + \mathcal{O}(n^{-c(p)}),
\end{aligned} \tag{6.53}$$

for some small $c(p) > 0$, where $V_{i,j}$ and U_i are as in (6.32). The implicit constant in $\mathcal{O}(\cdot)$ may depend on p.

Proof. In order to prove the lemma we have to check that the integral of the error term in (6.30) is at most of size $n^{-c(p)}$, and that the integral of $V_{i,j} + \kappa_4 U_i U_j$ for $\eta_i \leq \eta_c$ or $\eta_i \geq T$ is similarly negligible. In the remainder of the proof we assume that p is even, since the terms with p odd are of lower order by (6.30).

Note that by the explicit form of m_i, u_i in (6.19)–(6.20), by the definition of $V_{i,j}, U_i, U_j$ in (6.32), the fact that by $-m_i^2 + |z_i|^2 u_i^2 = u_i$ we have

$$V_{i,j} = \frac{1}{2}\partial_{\eta_i}\partial_{\eta_j} \log\Big(1 - u_i u_j\Big[1 - |z_i - z_j|^2 + (1-u_i)|z_i|^2 + (1-u_j)|z_j|^2\Big]\Big),$$

and using $|\partial_{\eta_i} m_i| \le [\Im m^{z_i}(\mathrm{i}\eta_i) + \eta_i]^{-2}$ by (6.24)–(6.26), we conclude (see also (6.109)–(6.110) later)

$$|V_{i,j}| \lesssim \frac{[(\Im m^{z_i}(\mathrm{i}\eta_i)+\eta_i)(\Im m^{z_j}(\mathrm{i}\eta_j)+\eta_j)]^{-2}}{[|z_i - z_j|^2 + (\eta_i+\eta_j)(\min\{\Im m^{z_i}, \Im m^{z_j}\}^2)]^2}, \quad |U_i| \lesssim \frac{1}{\Im m^{z_i}(\mathrm{i}\eta_i)^2 + \eta_i^3}. \tag{6.54}$$

Using the bound (6.40) to remove the regime $Z_i := \{|1-|z_i|^2| \le n^{-2\nu}\}$ for any $i \in [p]$, for some small $\nu > 0$, we conclude that the lhs. of (6.53) is equal to

$$\frac{(-n)^p}{(2\pi\mathrm{i})^p}\prod_{i\in[p]}\int_{Z_i^c} \mathrm{d}^2 z_i \Delta f^{(i)}(z_i)\, \mathbf{E}\prod_{i\in[p]}\int_{\eta_c}^{T} \langle G^{z_i}(\mathrm{i}\eta_i) - \mathbf{E}\, G^{z_i}(\mathrm{i}\eta_i)\rangle\, \mathrm{d}\eta_i + \mathcal{O}\left(\frac{n^{p\xi}}{n^{\nu}}\right), \tag{6.55}$$

for any very small $\xi > 0$. Additionally, since the error term Ψ defined in (6.31) behaves badly for small $|z_i - z_j|$, we remove the regime

$$\widehat{Z}_i := \bigcup_{j<i}\{z_i : |z_i - z_j| \le n^{-2\nu}\}$$

in each z_i-integral in (6.55) using (6.40), and, denoting $f^{(i)} = f^{(i)}(z_i)$, get

$$\frac{(-n)^p}{(2\pi\mathrm{i})^p}\prod_{i\in[p]}\int_{Z_i^c\cap\widehat{Z}_i^c} \mathrm{d}^2 z_i \Delta f^{(i)}\, \mathbf{E}\prod_{i\in[p]}\int_{\eta_c}^{T} \langle G^{z_i}(\mathrm{i}\eta_i) - \mathbf{E}\, G^{z_i}(\mathrm{i}\eta_i)\rangle\, \mathrm{d}\eta_i + \mathcal{O}\left(\frac{n^{p\xi}}{n^{\nu}}\right). \tag{6.56}$$

Plugging (6.30) into (6.56), and using the first bound in (6.38) to remove the regime $\eta_i \ge T$ for the lhs. of (6.53) we get

$$\begin{aligned} &\frac{1}{(2\pi\mathrm{i})^p}\prod_{i\in[p]}\int_{Z_i^c\cap\widehat{Z}_i^c} \mathrm{d}^2 z_i \Delta f^{(i)} \sum_{P\in\Pi_p}\prod_{\{i,j\}\in P}\int_0^\infty\int_0^\infty -\frac{V_{i,j}+\kappa_4 U_i U_j}{8\pi^2}\,\mathrm{d}\eta_j\,\mathrm{d}\eta_i \\ &\qquad + \mathcal{O}\left(\frac{n^{p\xi}}{n^{\nu}} + \frac{n^{20\nu p+\delta_1}}{n} + \frac{n^{\xi p + 2p\nu}}{n^{\delta_1/2}}\right), \end{aligned} \tag{6.57}$$

where $\eta_c = n^{-1+\delta_1}$, the second last error term comes from adding back the regimes $\eta_i \in [0,\eta_c]$ using that

$$|V_{i,j}| \le \frac{n^{20\nu}}{(1+\eta_i^2)(1+\eta_j^2)}, \qquad |U_i| \le \frac{n^{4\nu}}{1+\eta_i^3},$$

for $z_i \in Z_i^c \cap \widehat{Z}_i^c$ and $z_j \in Z_j^c \cap \widehat{Z}_j^c$ by (6.54). The last error term in (6.57) comes from the integral of Ψ, with Ψ defined in (6.31). Finally, we perform the η-integrations using the explicit formulas (6.58) and (6.59) below. After that, we add back the domains Z_i and $\widehat{Z_i}$ for $i \in [p]$ at a negligible error, since these domains have volume of order $n^{-2\nu}$, $\Delta f^{(i)} \in L^2$, and the logarithmic singularities from (6.58) are integrable. This concludes (6.53) choosing ν so that $\nu \ll \delta_1 \ll 1$. □

In the next three sub-sections we compute the integrals in (6.53) for any i, j's. To make our notation simpler we use only the indices $1, 2$, i.e. we compute the integral of $V_{1,2}$ and $U_1 U_2$.

6.4.3.1 Computation of the (η_1, η_2)-integrals

Using the relations in (6.32) we explicitly compute the (η_1, η_2)-integral of $V_{1,2}$:

$$
\begin{aligned}
-\int_0^\infty \int_0^\infty V_{1,2}\, \mathrm{d}\eta_1\, \mathrm{d}\eta_2 &= -\frac{1}{2}\log A \restriction_{\substack{\eta_1=0,\\ \eta_2=0}} \\
&= \Theta(z_1,z_2) := \frac{1}{2}\begin{cases} -\log|z_1-z_2|^2, & |z_1|,|z_2|\le 1,\\ \log|z_l|^2 - \log|z_1-z_2|^2, & |z_m|\le 1, |z_l|>1,\\ \log|z_1z_2|^2 - \log|1-z_1\overline{z}_2|^2, & |z_1|,|z_2|>1, \end{cases}
\end{aligned} \tag{6.58}
$$

with $A(\eta_1, \eta_2, z_1, z_2)$ defined by

$$
A(\eta_1,\eta_2,z_1,z_2) := 1 + (u_1u_2|z_1||z_2|)^2 - m_1^2 m_2^2 - 2u_1u_2\Re z_1\overline{z}_2.
$$

Then the η_i-integral of U_i, for $i \in \{1,2\}$, is given by

$$
\int_0^\infty U_i\, \mathrm{d}\eta_i = \frac{\mathrm{i}}{\sqrt{2}}(1-|z_i|^2). \tag{6.59}
$$

Before proceeding we rewrite $\Theta(z_1, z_2)$ as

$$
\begin{aligned}
2\Theta(z_1,z_2) &= -\log|z_1-z_2|^2 + \log|z_1|^2 \mathbf{1}(|z_1|>1) + \log|z_2|^2\mathbf{1}(|z_2|>1) \\
&\quad + \left[\log|z_1-z_2|^2 - \log|1-z_1\overline{z}_2|^2\right]\mathbf{1}(|z_1|,|z_2|>1).
\end{aligned}
$$

In the remainder of this section we use the notations

$$
\mathrm{d}z := \mathrm{d}z + \mathrm{i}\,\mathrm{d}y, \quad \mathrm{d}\overline{z} := \mathrm{d}x - \mathrm{i}\,\mathrm{d}y, \qquad \partial_z := \frac{\partial_x - \mathrm{i}\partial_y}{2}, \quad \partial_{\overline{z}} := \frac{\partial_x + \mathrm{i}\partial_y}{2},
$$

and $\partial_l := \partial_{z_l}$, $\overline{\partial}_l := \partial_{\overline{z}_l}$. With this notation $\Delta_{z_l} = 4\partial_{z_l}\partial_{\overline{z}_l}$.

We split the computation of the leading term in the rhs. of (6.53) into two parts: the integral of $V_{1,2}$, and the the integral of U_1U_2.

6.4.3.2 Computation of the (z_1, z_2)-integral of $V_{1,2}$

In this section we compute the integral of $V_{1,2}$ in (6.53). To make our notation easier in the remainder of this section we use the notation f and g, instead of $f^{(1)}$, $f^{(2)}$, with f in Theorem 6.2.1 and $g = f$ or $g = \overline{f}$.

Lemma 6.4.8. *Let $V_{1,2}$ be defined in (6.32), then*

$$
\begin{aligned}
&-\frac{1}{8\pi^2}\int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 \Delta f(z_1)\Delta\overline{g(z_2)} \int_0^\infty \mathrm{d}\eta_1 \int_0^\infty \mathrm{d}\eta_2 V_{1,2} \\
&\qquad = \frac{1}{4\pi}\int_{\mathbf{D}} \langle \nabla g, \nabla f\rangle\, \mathrm{d}^2 z + \frac{1}{2}\sum_{m\in\mathbf{Z}} |m| \widehat{f\restriction_{\partial \mathbf{D}}}(m)\overline{\widehat{g}\restriction_{\partial\mathbf{D}}(m)}.
\end{aligned} \tag{6.60}
$$

Note that the rhs. of (6.60) gives exactly the first two terms in (6.9).

Using the expression of $V_{1,2}$ in (6.32) and the computation of its (η_1, η_2)-integral in (6.58), we have that

$$
\begin{aligned}
&-\frac{1}{8\pi^2}\int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 \Delta f(z_1)\Delta\overline{g(z_2)} \int_0^\infty \mathrm{d}\eta_1 \int_0^\infty \mathrm{d}\eta_2 V_{1,2} \\
&\qquad = \frac{2}{\pi^2}\int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 \partial_1\overline{\partial}_1 f(z_1)\partial_2\overline{\partial}_2\overline{g(z_2)}\Theta(z_1,z_2),
\end{aligned} \tag{6.61}
$$

with $\Theta(z_1, z_2)$ is defined in the rhs. of (6.58).

We compute the r.h.s. of (6.61) as stated in Lemma 6.4.9. The proof of this lemma is postponed to Appendix 6.A.

Lemma 6.4.9. *Let* $\Theta(z_1, z_2)$ *be defined in* (6.58)*, then we have that*

$$\begin{aligned}\frac{2}{\pi^2}\int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 \partial_1 \overline{\partial}_1 f(z_1) \partial_2 \overline{\partial}_2 \overline{g(z_2)} \Theta(z_1, z_2) &= \frac{1}{4\pi}\int_{\mathbf{D}} \langle \nabla g, \nabla f \rangle \,\mathrm{d}^2 z \\ &+ \lim_{\epsilon \to 0} \left[\frac{1}{2\pi^2} \int_{|z_1|\geq 1} \mathrm{d}^2 z_1 \int_{\substack{|1-z_1\overline{z}_2|\geq \epsilon, \\ |z_2|\geq 1}} \mathrm{d}^2 z_2 \, \partial_1 f(z_1) \overline{\partial}_2 \overline{g(z_2)} \frac{1}{(1-\overline{z}_1 z_2)^2} \right. \\ &\left. + \frac{1}{2\pi^2} \int_{|z_1|\geq 1} \mathrm{d}^2 z_1 \int_{\substack{|1-z_1\overline{z}_2|\geq \epsilon, \\ |z_2|\geq 1}} \mathrm{d}^2 z_2 \, \overline{\partial}_1 f(z_1) \partial_2 \overline{g(z_2)} \frac{1}{(1-z_1 \overline{z}_2)^2} \right].\end{aligned} \tag{6.62}$$

Proof of Lemma 6.4.8. By Lemma 6.4.9 it follows that to prove Lemma 6.4.8 it is enough to compute the last two lines in the rhs. of (6.62).

Note that using the change of variables $\overline{z}_1 \to 1/\overline{z}_1$, $z_2 \to 1/z_2$ the integral in the rhs. of (6.62) is equal to the same integral on the domain $|z_1|, |z_2| \leq 1$, $|1 - z_1\overline{z}_2| \geq \epsilon$. By a standard density argument, using that $f, g \in H_0^{2+\delta}$, it is enough to compute the limit in (6.62) only for polynomials, hence, from now on, we consider polynomials f, g of the form

$$f(z_1) = \sum_{k,l\geq 0} z_1^k \overline{z}_1^l a_{kl}, \qquad g(z_2) = \sum_{k,l\geq 0} z_2^k \overline{z}_2^l b_{kl}, \tag{6.63}$$

for some coefficients $a_{kl}, b_{kl} \in \mathbf{C}$. We remark that the summations in (6.63) are finite since f and g are polynomials. Then, using that

$$\lim_{\epsilon\to 0} \int_{|z_1|\leq 1} \int_{\substack{|1-z_1\overline{z}_2|\geq \epsilon, \\ |z_2|\leq 1}} z_1^\alpha \overline{z}_1^\beta z_2^{\alpha'} \overline{z}_2^{\beta'} \,\mathrm{d}^2 z_1 \,\mathrm{d}^2 z_2 = \frac{\pi^2}{(\alpha+1)(\alpha'+1)} \delta_{\alpha,\beta}\delta_{\alpha',\beta'},$$

we compute the limit in the rhs. of (6.62) as follows

$$\begin{aligned}\lim_{\epsilon\to 0} \sum_{k,l,k',l',m\geq 0} \frac{1}{2\pi^2} \int \int_{|z_1|\leq 1} \int_{\substack{|1-z_1\overline{z}_2|\geq \epsilon, \\ |z_2|\leq 1}} \mathrm{d}^2 z_1 \,\mathrm{d}^2 z_2 \, m a_{kl} \overline{b_{k'l'}} \\ \times \left[kk' z_1^{k-1} \overline{z}_1^{l+m-1} z_2^{l'+m-1} \overline{z}_2^{k'-1} + ll' z_1^{k+m-1} \overline{z}_1^{l-1} z_2^{k'+m-1} \overline{z}_2^{l'-1} \right] \\ = \frac{1}{2} \sum_{\substack{k,l,k',l', \\ m\geq 0}} m a_{kl} \overline{b_{k'l'}} \left[\delta_{k,l+m}\delta_{k',l'+m} + \delta_{k,l-m}\delta_{k',l'-m} \right] \\ = \frac{1}{2} \sum_{\substack{k,l,k',l'\geq 0, \\ m\in\mathbf{Z}}} |m| a_{kl} \overline{b_{k'l'}} \delta_{k,l+m}\delta_{k',l'+m}.\end{aligned} \tag{6.64}$$

On the other hand

$$\sum_{m\in\mathbf{Z}} |m| \widehat{f \restriction_{\partial\mathbf{D}}}(m) \widehat{g \restriction_{\partial\mathbf{D}}}(m) = \sum_{m\in\mathbf{Z}} |m| \sum_{k,l,k',l'\geq 0} a_{kl} \overline{b_{k'l'}} \delta_{m,k-l}\delta_{m,k'-l'}, \tag{6.65}$$

where

$$\widehat{f \restriction_{\partial\mathbf{D}}}(k) := \frac{1}{2\pi}\int_0^{2\pi} f \restriction_{\partial\mathbf{D}} (e^{i\theta}) e^{-ik\theta} \,\mathrm{d}\theta, \quad f \restriction_{\partial\mathbf{D}} (e^{i\theta_j}) = \sum_{k\in\mathbf{Z}} \widehat{f \restriction_{\partial\mathbf{D}}}(k) e^{i\theta_j k}.$$

Finally, combining (6.62) and (6.64)–(6.65), we conclude the proof of (6.60).

□

6.4.3.3 Computation of the (z_1, z_2)-integral of U_1U_2

In order to conclude the proof of Theorem 6.2.1, in this section we compute the integral of U_1U_2 in (6.53). Similarly to the previous section, we use the notation f and g, instead of $f^{(1)}$, $f^{(2)}$, with f in Theorem 6.2.1 and $g = f$ or $g = \overline{f}$.

Lemma 6.4.10. *Let* $\kappa_4 = n^2[\mathbf{E}|x_{11}|^2 - 2(\mathbf{E}|x_{11}|^2)]$, *and let* U_1, U_2 *be defined in* (6.32), *then*

$$\begin{aligned} &-\frac{\kappa_4}{8\pi^2}\int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 \Delta f(z_1)\Delta\overline{g(z_2)} \int_0^\infty \mathrm{d}\eta_1 \int_0^\infty \mathrm{d}\eta_2 U_1 U_2 \\ &\quad = \kappa_4 \left(\frac{1}{\pi}\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z - \widehat{f\restriction_{\partial\mathbf{D}}}(0)\right)\left(\frac{1}{\pi}\int_{\mathbf{D}} \overline{g(z)}\,\mathrm{d}^2 z - \overline{\widehat{g\restriction_{\partial\mathbf{D}}}(0)}\right). \end{aligned} \tag{6.66}$$

Proof of Theorem 6.2.1. Theorem 6.2.1 readily follows combining Lemma 6.4.7, Lemma 6.4.8 and Lemma 6.4.10. □

Proof of Lemma 6.4.10. First of all, we recall the following formulas of integration by parts

$$\int_{\mathbf{D}} \partial_z f(z,\overline{z})\,\mathrm{d}^2 z = \frac{\mathrm{i}}{2}\int_{\partial\mathbf{D}} f(z,\overline{z})\,\mathrm{d}\overline{z}, \qquad \int_{\mathbf{D}} \partial_{\overline{z}} f(z,\overline{z})\,\mathrm{d}^2 z = -\frac{\mathrm{i}}{2}\int_{\partial\mathbf{D}} f(z,\overline{z})\,\mathrm{d}z. \tag{6.67}$$

Then, using the computation of the η-integral of W in (6.59), and integration by parts (6.67) twice, we conclude that

$$\begin{aligned} \int_{\mathbf{C}} \Delta f \int_0^\infty U\, \mathrm{d}\eta\,\mathrm{d}^2 z &= \mathrm{i}2\sqrt{2}\int_{\mathbf{D}} \partial\overline{\partial} f(z)(1-|z|^2)\,\mathrm{d}^2 z = \mathrm{i}2\sqrt{2}\int_{\mathbf{D}} \overline{\partial} f(z)\overline{z}\,\mathrm{d}^2 z \\ &= -\mathrm{i}2\sqrt{2}\left(\int_{\mathbf{D}} f(z)\ \mathrm{d}^2 z + \frac{\mathrm{i}}{2}\int_{\partial(\mathbf{D})} f(z)\overline{z}\,\mathrm{d}z\right) \\ &= -\mathrm{i}2\sqrt{2}\left(\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z - \pi\widehat{f\restriction_{\partial\mathbf{D}}}(0)\right). \end{aligned}$$

This concludes the proof of this lemma. □

6.5 Local law for products of resolvents

The main technical result of this section is a local law for *products* of resolvents with different spectral parameters $z_1 \neq z_2$. Our goal is to find a deterministic approximation to $\langle AG^{z_1}BG^{z_2}\rangle$ for generic bounded deterministic matrices A, B. Due to the correlation between the two resolvents the deterministic approximation to $\langle AG^{z_1}BG^{z_2}\rangle$ is not simply $\langle AM^{z_1}BM^{z_2}\rangle$. In the context of linear statistics such local laws for products of resolvents have previously been obtained e.g. for Wigner matrices in [89] and for sample-covariance matrices in [56] albeit with weaker error bounds. In the current non-Hermitian setting we need such local law twice; for the resolvent CLT in Proposition 6.3.3, and for the asymptotic independence of resolvents in Proposition 6.3.5. The key point for the latter is to obtain an improvement in the error term for mesoscopic separation $|z_1 - z_2| \sim n^{-\epsilon}$, a fine effect that has not been captured before.

Our proof applies verbatim to both real and complex i.i.d. matrices, as well as to resolvents $G^z(w)$ evaluated at an arbitrary spectral parameter $w \in \mathbf{H}$. We therefore work with this more general setup in this section, even though for the application in the proofs of Propositions 6.3.3–6.3.5 this generality is not necessary.

We recall from [13] that with the shorthand notations

$$G_i := G^{z_i}(w_i), \quad M_i := M^{z_i}(w_i), \tag{6.68}$$

the deviation of G_i from M_i is computed from the identity

$$G_i = M_i - M_i \underline{W G_i} + M_i \mathcal{S}[G_i - M_i] G_i, \quad W := \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}. \tag{6.69}$$

The relation (6.69) requires some definitions. First, the linear *covariance* or *self-energy operator* $\mathcal{S}\colon \mathbf{C}^{2n\times 2n} \to \mathbf{C}^{2n\times 2n}$ is given by

$$\mathcal{S}\left[\begin{pmatrix} A & B \\ C & D \end{pmatrix}\right] := \widetilde{\mathbf{E}} \widetilde{W} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \widetilde{W} = \begin{pmatrix} \langle D \rangle & 0 \\ 0 & \langle A \rangle \end{pmatrix}, \quad \widetilde{W} = \begin{pmatrix} 0 & \widetilde{X} \\ \widetilde{X}^* & 0 \end{pmatrix}, \tag{6.70}$$

where $\widetilde{X} \sim \mathrm{Gin}_{\mathbf{C}}$, i.e. it averages the diagonal blocks and swaps them. Here $\mathrm{Gin}_{\mathbf{C}}$ stands for the standard complex Ginibre ensemble. The ultimate equality in (6.70) follows directly from $\mathbf{E}\, \widetilde{x}_{ab}^2 = 0$, $\mathbf{E}|\widetilde{x}_{ab}|^2 = n^{-1}$. Second, underlining denotes, for any given function $f\colon \mathbf{C}^{2n\times 2n} \to \mathbf{C}^{2n\times 2n}$, the *self-renormalisation* $\underline{Wf(W)}$ defined by

$$\underline{Wf(W)} := Wf(W) - \widetilde{\mathbf{E}} \widetilde{W} (\partial_{\widetilde{W}} f)(W), \tag{6.71}$$

where ∂ indicates a directional derivative in the direction $\widetilde{W}$ and $\widetilde{W}$ denotes an independent random matrix as in (6.70) with $\widetilde{X}$ a complex Ginibre matrix with expectation $\widetilde{\mathbf{E}}$. Note that we use complex Ginibre $\widetilde{X}$ irrespective of the symmetry class of X. Therefore, using the resolvent identity, it follows that

$$\underline{WG} = WG + \widetilde{\mathbf{E}} \widetilde{W} G \widetilde{W} G = WG + \mathcal{S}[G]G.$$

We now use (6.69) and (6.71) to compute

$$\begin{aligned} G_1 B G_2 &= M_1 B G_2 - M_1 \underline{W G_1} B G_2 + M_1 \mathcal{S}[G_1 - M_1] G_1 B G_2 \\ &= M_1 B M_2 + M_1 B (G_2 - M_2) - M_1 \underline{W G_1 B G_2} + M_1 \mathcal{S}[G_1 B G_2] M_2 \\ &\quad + M_1 \mathcal{S}[G_1 B G_2](G_2 - M_2) + M_1 \mathcal{S}[G_1 - M_1] G_1 B G_2, \end{aligned} \tag{6.72}$$

where, in the second equality, we used

$$\begin{aligned} \underline{W G_1 B G_2} &= W G_1 B G_2 + \mathcal{S}[G_1] G_1 B G_2 + \mathcal{S}[G_1 B G_2] G_2 \\ &= \underline{W G_1} B G_2 + \mathcal{S}[G_1 B G_2] G_2. \end{aligned}$$

Assuming that the self-renormalised terms and the ones involving $G_i - M_i$ in (6.72) are small, (6.72) implies

$$G_1 B G_2 \approx M_B^{z_1,z_2}, \tag{6.73}$$

where

$$M_B^{z_1,z_2}(w_1,w_2) := (1 - M^{z_1}(w_1) \mathcal{S}[\cdot] M^{z_2}(w_2))^{-1} [M^{z_1}(w_1) B M^{z_2}(w_2)]. \tag{6.74}$$

We define the corresponding 2-*body stability operator*

$$\widehat{\mathcal{B}} = \widehat{\mathcal{B}}_{12} = \widehat{\mathcal{B}}_{12}(z_1, z_2, w_1, w_2) := 1 - M_1 \mathcal{S}[\cdot] M_2, \tag{6.75}$$

acting on the space of $2n \times 2n$ matrices equipped with the usual Euclidean matrix norm which induces a natural norm for $\widehat{\mathcal{B}}$.

Our main technical result of this section is making (6.73) rigorous in the sense of Theorem 6.5.2 below. To keep notations compact, we first introduce a commonly used (see, e.g. [81]) notion of high-probability bound.

Definition 6.5.1 (Stochastic Domination)**.** *If*

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

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

$$\sup_{u \in U^{(n)}} \mathbf{P} \left[X^{(n)}(u) > n^{\epsilon} Y^{(n)}(u) \right] \le n^{-D}$$

for large enough $n \ge n_0(\epsilon, D)$. In this case we use the notation $X \prec Y$.

Theorem 6.5.2. *Fix $z_1, z_2 \in \mathbf{C}$ and $w_1, w_2 \in \mathbf{C}$ with $|\eta_i| := |\Im w_i| \ge n^{-1}$ such that*

$$\eta_* := \min\{|\eta_1|, |\eta_2|\} \ge n^{-1+\epsilon} \|\widehat{\mathcal{B}}_{12}^{-1}\|$$

for some $\epsilon > 0$. Assume that $G^{z_1}(w_1), G^{z_2}(w_2)$ satisfy the local laws in the form

$$|\langle A(G^{z_i} - M^{z_i}) \rangle| \prec \frac{\|A\|}{n|\eta_i|}, \quad |\langle \boldsymbol{x}, (G^{z_i} - M^{z_i}) \boldsymbol{y} \rangle| \prec \frac{\|\boldsymbol{x}\| \|\boldsymbol{y}\|}{\sqrt{n|\eta_i|}}$$

for any bounded deterministic matrix and vectors $A, \boldsymbol{x}, \boldsymbol{y}$. Then, for any bounded deterministic matrix B, with $\|B\| \lesssim 1$, the product of resolvents $G^{z_1} B G^{z_2} = G^{z_1}(w_1) B G^{z_2}(w_2)$ is approximated by $M_B^{z_1,z_2} = M_B^{z_1,z_2}(w_1, w_2)$ defined in (6.74) in the sense that

$$\begin{aligned} |\langle A(G^{z_1} B G^{z_2} - M_B^{z_1,z_2}) \rangle| &\prec \frac{\|A\| \|\widehat{\mathcal{B}}_{12}^{-1}\|}{n\eta_* |\eta_1 \eta_2|^{1/2}} \\ &\qquad \times \left(\eta_*^{1/12} + \eta_*^{1/4} \|\widehat{\mathcal{B}}_{12}^{-1}\| + \frac{1}{\sqrt{n\eta_*}} + \frac{\|\widehat{\mathcal{B}}_{12}^{-1}\|^{1/4}}{(n\eta_*)^{1/4}} \right), \\ |\langle \boldsymbol{x}, (G^{z_1} B G^{z_2} - M_B^{z_1,z_2}) \boldsymbol{y} \rangle| &\prec \frac{\|\boldsymbol{x}\| \|\boldsymbol{y}\| \|\widehat{\mathcal{B}}_{12}^{-1}\|}{(n\eta_*)^{1/2} |\eta_1 \eta_2|^{1/2}} \end{aligned} \tag{6.76}$$

for any deterministic $A, \boldsymbol{x}, \boldsymbol{y}$.

The estimates in (6.76) will be complemented by a upper bound on $\|\widehat{\mathcal{B}}^{-1}\|$ in Lemma 6.6.1, where we will prove in particular that $\|\widehat{\mathcal{B}}^{-1}\| \lesssim n^{2\delta}$ whenever $|z_1 - z_2| \gtrsim n^{-\delta}$, for some small fixed $\delta > 0$.

The proof of Theorem 6.5.2 will follow from a bootstrap argument once the main input, the following high-probability bound on $\underline{W G_1 B G_2}$ has been established.

Proposition 6.5.3. *Under the assumptions of Theorem 6.5.2, the following estimates hold uniformly in* $n^{-1} \lesssim |\eta_1|, |\eta_2| \lesssim 1$.

1. *We have the isotropic bound*

$$|\langle \boldsymbol{x}, \underline{WG_1BG_2\boldsymbol{y}}\rangle| \prec \frac{1}{(n\eta_*)^{1/2}|\eta_1\eta_2|^{1/2}} \tag{6.77a}$$

uniformly for deterministic vectors and matrix $\|\boldsymbol{x}\| + \|\boldsymbol{y}\| + \|B\| \le 1$.

2. *Assume that for some positive deterministic* $\theta = \theta(z_1, z_2, \eta_*)$ *an a priori bound*

$$|\langle AG_1BG_2\rangle| \prec \theta \tag{6.77b}$$

has already been established uniformly in deterministic matrices $\|A\| + \|B\| \le 1$. *Then we have the improved averaged bound*

$$|\langle \underline{WG_1BG_2A}\rangle| \prec \frac{1}{n\eta_*|\eta_1\eta_2|^{1/2}}\Big((\theta\eta_*)^{1/4} + \frac{1}{\sqrt{n\eta_*}} + \eta_*^{1/12}\Big), \tag{6.77c}$$

again uniformly in deterministic matrices $\|A\| + \|B\| \le 1$.

Proof of Theorem 6.5.2. We note that from (6.74) and (6.21) we have

$$\|M_B^{z_1,z_2}\| \lesssim \|\widehat{\mathcal{B}}^{-1}\| \tag{6.78}$$

and abbreviate $G_{12} := G_1BG_2$, $M_{12} := M_B^{z_1,z_2}$. We now assume an a priori bound $|\langle G_{12}A\rangle| \prec \theta_1$, i.e. that (6.77b) holds with $\theta = \theta_1$. In the first step we may take $\theta_1 = |\eta_1\eta_2|^{-1/2}$ due to the local law for G_i from which it follows that

$$\begin{aligned}|\langle AG_1BG_2\rangle| &\le \sqrt{\langle AG_1G_1^*A^*\rangle}\sqrt{\langle BG_2G_2^*B^*\rangle}\\ &= \frac{1}{\sqrt{|\eta_1||\eta_2|}}\sqrt{\langle A\Im G_1A^*\rangle}\sqrt{\langle B\Im G_2B^*\rangle}.\end{aligned}$$

By (6.72) and (6.74) we have

$$\begin{aligned}\widehat{\mathcal{B}}[G_{12} - M_{12}] = M_1B(G_2 - M_2) - M_1\underline{WG_{12}} + M_1\mathcal{S}[G_{12}](G_2 - M_2)\\ + M_1\mathcal{S}[G_1 - M_1]G_{12},\end{aligned} \tag{6.79}$$

and from (6.27) and (6.77c) we obtain

$$\begin{aligned}|\langle A(G_{12} - M_{12})\rangle| &= |\langle A^*, \widehat{\mathcal{B}}^{-1}\widehat{\mathcal{B}}[G_{12} - M_{12}]\rangle| = |\langle (\widehat{\mathcal{B}}^*)^{-1}[A^*]^*\widehat{\mathcal{B}}[G_{12} - M_{12}]\rangle|\\ &\prec \|\widehat{\mathcal{B}}^{-1}\|\Big[\frac{1}{n\eta_*} + \frac{(\theta_1\eta_*)^{1/4} + (\sqrt{n\eta_*})^{-1} + \eta_*^{1/12}}{n\eta_*|\eta_1\eta_2|^{1/2}} + \frac{\theta_1}{n\eta_*}\Big].\end{aligned}$$

For the terms involving $G_i - M_i$ we used that $\mathcal{S}[R] = \langle RE_2\rangle E_1 + \langle RE_1\rangle E_2$ with the $2n \times 2n$ block matrices

$$E_1 = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \qquad E_2 = \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}, \tag{6.80}$$

i.e. that $\mathcal{S}$ effectively acts as a trace, so that the averaged bounds are applicable. Therefore with (6.78) it follows that

$$|\langle G_{12}A\rangle| \prec \theta_2 := \|\widehat{\mathcal{B}}^{-1}\|\Big[1+\frac{1}{n\eta_*}+\frac{(\theta_1\eta_*)^{1/4}+(\sqrt{n\eta_*})^{-1}+\eta_*^{1/12}}{n\eta_*|\eta_1\eta_2|^{1/2}}+\frac{\theta_1}{n\eta_*}\Big]. \quad (6.81)$$

By iterating (6.81) we can use $|\langle G_{12}A\rangle| \prec \theta_2 \ll \theta_1$ as new input in (6.77b) to obtain $|\langle G_{12}A\rangle| \prec \theta_3 \ll \theta_2$ since $n\eta_* \gg \|\widehat{\mathcal{B}}^{-1}\|$. Here θ_j, for $j = 3, 4, \ldots$, is defined iteratively by replacing θ_1 with θ_{j-1} in the rhs. of the defining equation for θ_2 in (6.81). This improvement continues until the fixed point of this iteration, i.e. until $\theta_N^{3/4}$ approaches $\|\widehat{\mathcal{B}}^{-1}\|n^{-1}\eta_*^{-7/4}$. For any given $\xi > 0$, after finitely many steps $N = N(\xi)$ the iteration stabilizes to

$$\theta_* \lesssim n^\xi\Big[\|\widehat{\mathcal{B}}^{-1}\| + \frac{\|\widehat{\mathcal{B}}^{-1}\|}{n\eta_*}\frac{\eta_*^{1/12}}{|\eta_1\eta_2|^{1/2}}+\frac{1}{\eta_*}\Big(\frac{\|\widehat{\mathcal{B}}^{-1}\|}{n\eta_*}\Big)^{4/3}\Big],$$

from which

$$|\langle A(G_{12}-M_{12})\rangle| \prec \frac{\|\widehat{\mathcal{B}}^{-1}\|}{n\eta_*|\eta_1\eta_2|^{1/2}}\Big(\eta_*^{1/12}+\eta_*^{1/4}\|\widehat{\mathcal{B}}^{-1}\|+\frac{1}{\sqrt{n\eta_*}}+\Big(\frac{\|\widehat{\mathcal{B}}^{-1}\|}{n\eta_*}\Big)^{1/4}\Big),$$

and therefore the averaged bound in (6.76) follows.

For the isotropic bound in (6.76) note that

$$\langle \boldsymbol{x}, (G_{12}-M_{12})\boldsymbol{y}\rangle = \operatorname{Tr}\big[(\widehat{\mathcal{B}}^*)^{-1}[\boldsymbol{x}\boldsymbol{y}^*]\big]^*\widehat{\mathcal{B}}[G_{12}-M_{12}]$$

and that due to the block-structure of $\widehat{\mathcal{B}}$ we have

$$(\widehat{\mathcal{B}}^*)^{-1}[\boldsymbol{x}\boldsymbol{y}^*] = \sum_{i=1}^4 \boldsymbol{x}_i\boldsymbol{y}_i^*, \qquad \|\boldsymbol{x}_i\|\|\boldsymbol{y}_i\| \lesssim \|\widehat{\mathcal{B}}^{-1}\|,$$

for some vectors $\boldsymbol{x}_i, \boldsymbol{y}_i$. The isotropic bound in (6.76) thus follows in combination with the isotropic bound in (6.27), (6.79) and (6.77a) applied to the pairs of vectors $\boldsymbol{x}_i, \boldsymbol{y}_i$. This completes the proof of the theorem modulo the proof of Proposition 6.5.3. □

6.5.1 Probabilistic bound and the proof of Proposition 6.5.3

We follow the graphical expansion outlined in [83, 84] adapted to the current setting. We focus on the case when X has complex entries and additionally mention the few changes required when X is a real matrix. We abbreviate $G_{12} = G_1BG_2$ and use iterated cumulant expansions to expand $\mathbf{E}|\langle \boldsymbol{x}, \underline{WG_{12}\boldsymbol{y}}\rangle|^{2p}$ and $\mathbf{E}|\langle \underline{WG_{12}}A\rangle|^{2p}$ in terms of polynomials in entries of G. For the expansion of the first W we have in the complex case

$$\begin{aligned}
&\mathbf{E}\operatorname{Tr}(\underline{WG_{12}}A)\operatorname{Tr}(\underline{WG_{12}}A)^{p-1}\operatorname{Tr}(A^*\underline{G_{12}^*W})^p\\
&\quad= \frac{1}{n}\mathbf{E}\sum_{ab}R_{ab}\operatorname{Tr}(\Delta^{ab}G_{12}A)\partial_{ba}\Big[\operatorname{Tr}(\underline{WG_{12}}A)^{p-1}\operatorname{Tr}(A^*\underline{G_{12}^*W})^p\Big]\\
&\qquad+\sum_{k\ge 2}\sum_{ab}\sum_{\boldsymbol{\alpha}\in\{ab,ba\}^k}\frac{\kappa(ab,\boldsymbol{\alpha})}{k!}\\
&\qquad\qquad\times\mathbf{E}\,\partial_{\boldsymbol{\alpha}}\Big[\operatorname{Tr}(\Delta^{ab}G_{12}A)\operatorname{Tr}(\underline{WG_{12}}A)^{p-1}\operatorname{Tr}(A^*\underline{G_{12}^*W})^p\Big]
\end{aligned} \quad (6.82)$$

and similarly for $\langle \boldsymbol{x}, \underline{WG_{12}}\boldsymbol{y}\rangle$, where unspecified summations $\sum_a$ are understood to be over $\sum_{a\in[2n]}$, and $(\Delta^{ab})_{cd} := \delta_{ac}\delta_{bd}$. Here we introduced the matrix $R_{ab} := \mathbf{1}(a\le n, b > n) + \mathbf{1}(a > n, b\le n)$ which is the rescaled second order cumulant (variance), i.e. $R_{ab} = n\kappa(ab,ba)$. For $\boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_k)$ we denote the joint cumulant of $w_{ab}, w_{\alpha_1},\dots,w_{\alpha_k}$ by $\kappa(ab,\boldsymbol{\alpha})$ which is non-zero only for $\boldsymbol{\alpha}\in\{ab,ba\}^k$. The derivative $\partial_{\boldsymbol{\alpha}}$ denotes the derivative with respect to $w_{\alpha_1},\dots,w_{\alpha_k}$. Note that in (6.82) the $k=1$ term differs from the $k\ge 2$ terms in two aspects. First, we only consider the ∂_{ba} derivative since in the complex case we have $\kappa(ab,ab) = 0$. Second, the action of the derivative on the first trace is not present since it is cancelled by the *self-renormalisation* of $\underline{WG_{12}}$.

In the real case (6.82) differs slightly. First, for the $k=1$ terms both ∂_{ab} and ∂_{ba} have to be taken into account with the same weight R since $\kappa(ab,ab)=\kappa(ab,ba)$. Second, we chose only to renormalise the effect of the ∂_{ba}-derivative and hence the ∂_{ab}-derivative acts on all traces. Thus in the real case, compared to (6.82) there is an additional term given by

$$\frac{1}{n}\mathbf{E}\sum_{ab} R_{ab}\partial_{ab}\Big[\mathrm{Tr}(\Delta^{ab}G_{12}A)\,\mathrm{Tr}(\underline{WG_{12}}A)^{p-1}\,\mathrm{Tr}(A^*\underline{G_{12}^*W})^p\Big].$$

The main difference to [84, Section 4] and [83, Section 4] is that therein instead of $\underline{WG_{12}}$ the single-G renormalisation $\underline{WG}$ was considered. With respect to the action of the derivatives there is, however, little difference between the two since we have

$$\partial_{ab}G = -G\Delta^{ab}G,\quad \partial_{ab}G_{12} = -G_1\Delta^{ab}G_{12} - G_{12}\Delta^{ab}G_2.$$

Therefore after iterating the expansion (6.82) we structurally obtain the same polynomials as in [83, 84], except of the slightly different combinatorics and the fact that exactly $2p$ of the G's are G_{12}'s and the remaining G's are either G_1 or G_2. Thus, using the local law for G_i in the form

$$\begin{aligned}
&|\langle \boldsymbol{x}, G_i\boldsymbol{y}\rangle| \prec 1,\\
&|\langle \boldsymbol{x}, G_{12}\boldsymbol{y}\rangle| \le \sqrt{\langle \boldsymbol{x}, G_1G_1^*\boldsymbol{x}\rangle}\sqrt{\langle \boldsymbol{y}, BG_2G_2^*B^*\boldsymbol{y}\rangle}\\
&\qquad = \frac{1}{\sqrt{|\eta_1||\eta_2|}}\sqrt{\langle \boldsymbol{x}, (\Im G_1)\boldsymbol{x}\rangle}\sqrt{\langle \boldsymbol{y}, B(\Im G_2)B^*\boldsymbol{y}\rangle} \prec \frac{1}{\sqrt{|\eta_1||\eta_2|}}
\end{aligned}\tag{6.83}$$

for $\|\boldsymbol{x}\|+\|\boldsymbol{y}\|\lesssim 1$, we obtain exactly the same bound as in [84, Eq. (23a)] times a factor of $(|\eta_1||\eta_2|)^{-p}$ accounting for the $2p$ exceptional G_{12} edges, i.e.

$$\mathbf{E}|\langle \boldsymbol{x}, \underline{WG_{12}}\boldsymbol{y}\rangle|^{2p} \lesssim \frac{n^\epsilon}{(n\eta_*)^p|\eta_1|^p|\eta_2|^p},\quad \mathbf{E}|\langle \underline{WG_{12}}A\rangle|^{2p} \lesssim \frac{n^\epsilon}{(n\eta_*)^{2p}|\eta_1|^p|\eta_2|^p}.\tag{6.84}$$

The isotropic bound from (6.84) completes the proof of (6.77a).

It remains to improve the averaged bound in (6.84) in order to obtain (6.77c). We first have to identify where the bound (6.84) is suboptimal. By iterating the expansion (6.82) we obtain a complicated polynomial expression in terms of entries of G_{12}, G_1, G_2 which is most conveniently represented graphically as

$$\mathbf{E}|\langle \underline{WG_{12}}A\rangle|^{2p} = \sum_{\Gamma\in\mathrm{Graphs}(p)} c(\Gamma)\,\mathbf{E}\,\mathrm{Val}(\Gamma) + \mathcal{O}\Big(n^{-2p}\Big)\tag{6.85}$$

for some finite collection of Graphs(p). Before we precisely define the *value of* Γ, $\mathrm{Val}(\Gamma)$, we first give two examples. Continuing (6.82) in case $p=1$ we have

$$\begin{aligned} &\mathbf{E}\,\mathrm{Tr}(\underline{WG_{12}}A)\,\mathrm{Tr}(A^*\underline{G_{12}^*W}) \\ &= \sum_{ab}\frac{R_{ab}}{n}\,\mathbf{E}\,\mathrm{Tr}(\Delta^{ab}G_{12}A)\,\mathrm{Tr}(A^*G_{12}^*\Delta^{ba}) \\ &\quad -\sum_{ab}\frac{R_{ab}}{n}\,\mathbf{E}\,\mathrm{Tr}(\Delta^{ab}G_{12}A)\,\mathrm{Tr}(A^*\underline{G_2^*\Delta^{ba}G_{12}^*W}) \\ &\quad -\sum_{ab}\frac{R'_{ab}}{2n^{3/2}}\,\mathbf{E}\,\mathrm{Tr}(\Delta^{ab}G_1\Delta^{ba}G_{12}A)\,\mathrm{Tr}(A^*G_{12}^*\Delta^{ba})+\cdots \end{aligned} \tag{6.86a}$$

where, for illustration, we only kept two of the three Gaussian terms (the last being when W acts on G_1^*) and one non-Gaussian term. For the non-Guassian term we set $R'_{ab} := n^{3/2}\kappa(ab,ba,ba)$, $|R'_{ab}|\lesssim 1$. Note that in the case of i.i.d. matrices with $\sqrt{n}x_{ab}\stackrel{\mathrm{d}}{=}x$, we have $R'_{ab}=\kappa(x,\overline{x},\overline{x})$ for $a\le n, b>n$ and $R'_{ab}=\kappa(x,x,\overline{x})=\overline{\kappa(x,\overline{x},\overline{x})}$ for $a>n, b\le n$. For our argument it is of no importance whether matrices representing cumulants of degree at least three like R' are block-constant. It is important, however, that the variance $\kappa(ab,ba)$ represented by R is block-constant since later we will perform certain resummations. For the second term on the rhs. of (6.86a) we then obtain by another cumulant expansion that

$$\begin{aligned} &\sum_{ab}\frac{R_{ab}}{n}\,\mathbf{E}\,\mathrm{Tr}(\Delta^{ab}G_{12}A)\,\mathrm{Tr}(A^*\underline{G_2^*\Delta^{ba}G_{12}^*W}) \\ &= -\sum_{ab}\sum_{cd}\frac{R_{ab}R_{cd}}{n^2}\,\mathbf{E}(G_{12}\Delta^{dc}G_2A)_{ba}\,\mathrm{Tr}(A^*G_2^*\Delta^{ba}G_{12}^*\Delta^{cd})+\cdots \\ &\quad -\sum_{ab}\sum_{cd}\frac{R_{ab}R'_{cd}}{2!n^{5/2}}\,\mathbf{E}(G_{12}\Delta^{dc}G_2A)_{ba}\,\mathrm{Tr}(A^*G_2^*\Delta^{ba}G_{12}^*\Delta^{dc}G_1^*\Delta^{cd}), \end{aligned} \tag{6.86b}$$

where we kept one of the two Gaussian terms and one third order term. After writing out the traces, (6.86a)–(6.86b) become

$$\begin{aligned} &\sum_{ab}\frac{R_{ab}}{n}\,\mathbf{E}(G_{12}A)_{ba}(A^*G_{12}^*)_{ab}+\cdots \\ &-\sum_{ab}\frac{R'_{ab}}{n^{3/2}}\,\mathbf{E}(G_1)_{bb}(G_{12}A)_{aa}(A^*G_{12}^*)_{ab} \\ &+\sum_{ab}\sum_{cd}\frac{R_{ab}R_{cd}}{n^2}\,\mathbf{E}(G_{12})_{bd}(G_2A)_{ca}(A^*G_2^*)_{db}(G_{12}^*)_{ac} \\ &+\sum_{ab}\sum_{cd}\frac{R_{ab}R'_{cd}}{2!n^{5/2}}\,\mathbf{E}(G_{12})_{bd}(G_2A)_{ca}(G_1^*)_{cc}(A^*G_2^*)_{db}(G_{12}^*)_{ad}. \end{aligned} \tag{6.86c}$$

If X is real, then in (6.86) some additional terms appear since $\kappa(ab,ab)=\kappa(ab,ba)$ in the real case, while $\kappa(ab,ab)=0$ in the complex case. In the first equality of (6.86) this results in additional terms like

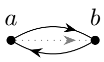
$$\begin{aligned}\sum_{ab}\frac{R_{ab}}{n}\mathbf{E}\Big(&-\operatorname{Tr}(\Delta^{ab}G_1\Delta^{ab}G_{12}A)\operatorname{Tr}(A^*\underline{G_{12}^*W})\\&+\operatorname{Tr}(\Delta^{ab}G_{12}A)\operatorname{Tr}(A^*G_{12}^*\Delta^{ab})\\&-\operatorname{Tr}(\Delta^{ab}G_{12}A)\operatorname{Tr}(A^*\underline{G_2^*\Delta^{ab}G_{12}^*W})+\dots\Big).\end{aligned}\tag{6.87}$$

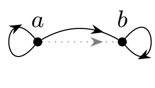
Out of the three terms in (6.87), however, only the first one is qualitatively different from the terms already considered in (6.86) since the other two are simply transpositions of already existing terms. After another expansion of the first term in (6.87) we obtain terms like

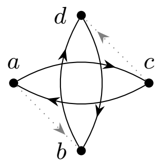
$$\begin{aligned}&\sum_{ab}\frac{R_{ab}}{n}(G_{12}A)_{ba}(A^*G_{12}^*)_{ba}+\cdots\\&+\sum_{ab}\sum_{cd}\frac{R_{ab}R_{cd}}{n^2}(G_1)_{ba}(G_{12}A)_{ba}(A^*G_2^*)_{dc}(G_{12}^*)_{dc}\\&+\sum_{ab}\sum_{cd}\frac{R_{ab}R'_{cd}}{2!n^{5/2}}(G_{12})_{bc}(G_2A)_{da}(A^*G_2^*)_{da}(G_{12}^*)_{bd}(G_2^*)_{cc}\end{aligned}\tag{6.88}$$

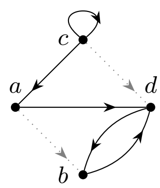
specific to the real case.

Now we explain how to encode (6.86) in the graphical formalism (6.85). The summation labels a_i, b_i correspond to vertices, while matrix entries correspond to edges between respective labelled vertices. We distinguish between the cumulant- or κ-edges E_κ, like R, R' and G-edges E_G, like $(A^*G_2^*)_{db}$ or $(G_{12}^*)_{ab}$, but do not graphically distinguish between G_1, G_{12}, $A^*G_2^*$, etc. The four terms from the rhs. of (6.86) would thus be represented as



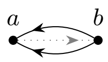
,


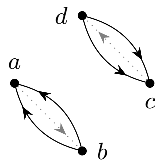
,


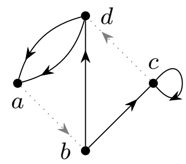
and


, (6.89)

where the edges from E_G are solid and those from E_κ dotted. Similarly, the three examples from (6.88) would be represented as



,


and


. (6.90)

It is not hard to see that after iteratively performing cumulant expansions up to order $4p$ for each remaining W we obtain a finite collection of polynomial expressions in R and G which correspond to graphs Γ from a certain set $\mathrm{Graphs}(p)$ with the following properties. We consider a directed graph $\Gamma=(V,E_\kappa\cup E_G)$ with an even number $|V|=2k$ of vertices, where k is the number of cumulant expansions along the iteration. The edge set is partitioned into two types of disjoint edges, the elements of E_κ are called *cumulant edges* and the elements of E_G are called G*-edges*. For $u \in V$ we define the G-degree of u as

$$\begin{aligned} d_G(u) &:= d_G^{\text{out}}(u) + d_G^{\text{in}}(u), \\ d_G^{\text{out}}(u) &:= |\{v \in V \mid (uv) \in E_G\}|, \quad d_G^{\text{in}}(u) := |\{v \in V \mid (vu) \in E_G\}|. \end{aligned}$$

We now record some structural attributes.

1. The graph (V, E_κ) is a perfect matching and in particular $|V| = 2|E_\kappa|$. We label the vertices by $u_1, \dots, u_k, v_1, \dots, v_k$ with cumulant edges $(u_1v_1), \dots, (u_kv_k)$. The ordering of the elements of E_κ indicated by $1, \dots, k$ is arbitrary and irrelevant.
2. The number of κ-edges is bounded by $|E_\kappa| \le 2p$ and therefore $|V| \le 4p$
3. For each $(u_iv_i) \in E_\kappa$, the G*-degree* of both vertices agrees, i.e. $d_G(u_i) = d_G(v_i) =: d_G(i)$. Furthermore the G-degree satisfies $2 \le d_G(i) \le 4p$. Note that loops (uu) contribute a value of 2 to the degree.
4. If $d_G(i) = 2$, then no loops are adjacent to either u_i or v_i.
5. We distinguish two types of G-edges $E_G = E_G^1 \cup E_G^2$ whose numbers are given by

$$|E_G^2| = 2p, \quad |E_G^1| = \sum_i d_G(i) - 2p, \quad |E_G| = |E_G^1| + |E_G^2|.$$

Note that in the examples (6.89) and (6.90) above we had $|E_\kappa| = 1$ in the first and $|E_\kappa| = 2$ in the other two cases. For the degrees we had $d_G(1) = 2$ in the first case, $d_G(1) = d_G(2) = 2$ in the second case, and $d_G(1) = 2, d_G(2) = 3$ in the third case. The number of G-edges involving G_{12} is 2 in all cases, while the number of remaining G-edges is 0, 2 and 3, respectively, in agreement with 5. We now explain how we relate the graphs to the polynomial expressions they represent.

(i) Each vertex $u \in V$ corresponds to a summation $\sum_{a \in [2n]}$ with a label a assigned to the vertex u.

(ii) Each G-edge $(uv) \in E_G^1$ represents a matrix $\mathcal{G}^{(uv)} = A_1 G_i A_2$ or $\mathcal{G}^{(uv)} = A_1 G_i^* A_2$ for some norm-bounded deterministic matrices A_1, A_2. Each G-edge $(uv) \in E_G^2$ represents a matrix $\mathcal{G}^{(uv)} = A_1 G_{12} A_2$ or $\mathcal{G}^{(uv)} = A_1 G_{12}^* A_2$ for norm bounded matrices A_1, A_2. We denote the matrices $\mathcal{G}^{(uv)}$ with a calligraphic "G" to avoid confusion with the ordinary resolvent matrix G.

(iii) Each κ-edge (uv) represents the matrix

$$R_{ab}^{(uv)} = \kappa(\underbrace{\sqrt{n}w_{ab}, \dots, \sqrt{n}w_{ab}}_{d_G^{\text{in}}(u)}, \underbrace{\sqrt{n}\overline{w_{ab}}, \dots, \sqrt{n}\overline{w_{ab}}}_{d_G^{\text{out}}(u)}),$$

where $d_G^{\text{in}}(u) = d_G^{\text{out}}(v)$ and $d_G^{\text{out}}(u) = d_G^{\text{in}}(v)$ are the in- and out degrees of u, v.

(iv) Given a graph Γ we define its value[4] as

$$\mathrm{Val}(\Gamma) := n^{-2p} \prod_{(u_i v_i)\in E_\kappa} \Big(\sum_{a_i,b_i\in[2n]} n^{-d_G(i)/2} R^{(u_i v_i)}_{a_i b_i} \Big) \prod_{(u_i v_i)\in E_G} \mathcal{G}^{(u_i v_i)}_{a_i b_i}, \tag{6.91}$$

where $R^{(u_i v_i)}$ is as in (iii) and a_i, b_i are the summation indices associated with u_i, v_i.

Proof of (6.85). In order to prove (6.85) we have to check that the graphs representing the polynomial expressions of the cumulant expansion up to order $4p$ have the attributes 1–5. Here 1–3 follow directly from the construction, with the lower bound $d_G(i) \geq 2$ being a consequence of $\mathbf{E}\, w_{ab} = 0$ and the upper bound $d_G(i) \leq 4p$ being a consequence of the fact that we trivially truncate the expansion after the $4p$-th cumulant. The error terms from the truncation are estimated trivially using (6.83). The fact 4 that no G-loops may be adjacent to degree two κ-edges follows since due to the self-renormalisation $\underline{WG_{12}}$ the the second cumulant of W can only act on some W or G in another trace, or if it acts on some G in its own trace then it generates a $\kappa(ab, ab)$ factor (only possible when X is real). In the latter case one of the two vertices has two outgoing, and the other one two incoming G-edges, and in particular no loops are adjacent to either of them. The counting of G_{12}-edges in E_G^2 in 5 is trivial since along the procedure no G_{12}-edges can be created or removed. For the counting of G_i edges in E_G^1 note that the action of the k-th order cumulant in the expansion of $\underline{WG_{12}}$ may remove k_1 W's and may create additional k_2 G_i's with $k = k_1 + k_2$, $k_1 \geq 1$. Therefore, since the number of G_i edges is 0 in the beginning, and the number of W's is reduced from $2p$ to 0 the second equality in 5 follows.

It now remains to check that with the interpretations (i)–(iv) the values of the constructed graphs are consistent in the sense of (6.85). The constant $c(\Gamma) \sim 1$ accounts for combinatorial factors in the iterated cumulant expansions and the multiplicity of identical graphs. The factor n^{-2p} in (iv) comes from the $2p$ normalised traces. The relation (iii) follows from the fact that the k-th order cumulant of k_1 copies of w_{ab} and k_2 copies of $\overline{w_{ab}} = w_{ba}$ comes together with k_1 copies of Δ^{ab} and k_2 copies of Δ^{ba}. Thus a is the first index of some G a total of k_2 times, while the remaining k_1 times the first index is b, and for the second indices the roles are reversed. □

Having established the properties of the graphs and the formula (6.85), we now estimate the value of any individual graph.

Naive estimate

We first introduce the so called *naive estimate*, N-Est(Γ), of a graph Γ as the bound on its value obtained by estimating the factors in (6.91) as $|\mathcal{G}^e_{ab}| \prec 1$ for $e \in E_G^1$ and $|\mathcal{G}^e_{ab}| \prec (|\eta_1||\eta_2|)^{-1/2}$ for $e \in E_G^2$, $|R^e_{ab}| \lesssim 1$ and estimating summations by their size. Thus, we obtain

$$\begin{aligned} \mathrm{Val}(\Gamma) \prec \text{N-Est}(\Gamma) :&= \frac{1}{n^{2p}|\eta_1|^p|\eta_2|^p} \prod_i \Big(n^{2-d_G(i)/2} \Big) \\ &\leq \frac{n^{|E_\kappa^2|} n^{|E_\kappa^3|/2}}{n^{2p}|\eta_1|^p|\eta_2|^p} \leq \frac{1}{|\eta_1|^p|\eta_2|^p}, \end{aligned} \tag{6.92}$$

[4]In [83] we defined the value with an expectation so that (6.85) holds without expectation. In the present paper we follow the convention of [84] and consider the value as a random variable.

where

$$E_\kappa^j := \{(u_i, v_i) \mid d_G(i) = j\}$$

is the set of degree j κ-edges, and in the last inequality we used $|E_\kappa^2| + |E_\kappa^3| \leq |E_\kappa| \leq 2p$.

Ward estimate

The first improvement over the naive estimate comes from the effect that sums of resolvent entries are typically smaller than the individual entries times the summation size. This effect can easily be seen from the *Ward* or resolvent identity $G^*G = \Im G/\eta = (G - G^*)/(2\mathrm{i}\eta)$. Indeed, the naive estimate of $\sum_a G_{ab}$ is n using $|G_{ab}| \prec 1$. However, using the the Ward identity we can improve this to

$$\Big|\sum_a G_{ab}\Big| \leq \sqrt{2n}\sqrt{\sum_a |G_{ab}|^2} = \sqrt{2n}\sqrt{(G^*G)_{bb}} = \sqrt{\frac{2n}{\eta}}\sqrt{(\Im G)_{bb}} \prec n\frac{1}{\sqrt{n\eta}},$$

i.e. by a factor of $(n\eta)^{-1/2}$. Similarly, we can gain two such factors if the summation index a appears in two G-factors off-diagonally, i.e.

$$\Big|\sum_a (G_1)_{ab}(G_2)_{ca}\Big| \leq \sqrt{(G_1^*G_1)_{bb}}\sqrt{(G_2G_2^*)_{cc}} \prec n\frac{1}{n\eta}.$$

However, it is impossible to gain more than two such factors per summation. We note that we have the same gain also for summations of G_{12}. For example, the naive estimate on $\sum_a (G_{12})_{ab}$ is $n|\eta_1\eta_2|^{-1/2}$ since $|(G_{12})_{ab}| \prec |\eta_1\eta_2|^{-1/2}$. Using the Ward identity, we obtain an improved bound of

$$\begin{aligned}\Big|\sum_a (G_{12})_{ab}\Big| &\leq \sqrt{2n}\sqrt{(G_{12}^*G_{12})_{bb}} = \sqrt{\frac{2n}{|\eta_1|}}\sqrt{(G_2^*B^*(\Im G_1)BG_2)_{bb}} \\ &\lesssim \sqrt{\frac{n}{|\eta_1|^2}}\sqrt{(G_2^*G_2)_{bb}} \prec \frac{\sqrt{n}}{|\eta_1||\eta_2|^{1/2}} \leq \frac{n}{|\eta_1\eta_2|^{1/2}}\frac{1}{\sqrt{n\eta_*}},\end{aligned}$$

where we recall $\eta_* = \min\{|\eta_1|, |\eta_2|\}$. Each of these improvements is associated with a specific G-edge with the restriction that one cannot gain simultaneously from more than two edges adjacent to any given vertex $u \in V$ while summing up the index a associated with u. Note, however, that globally it is nevertheless possible to gain from arbitrarily many G-edges adjacent to any given vertex, as long as the summation order is chosen correctly. In order to count the number edges giving rise to such improvements we recall a basic definition [140] from graph theory.

Definition 6.5.4. *For $k \geq 1$ a graph $\Gamma = (V, E)$ is called k-*degenerate *if any induced subgraph has minimal degree at most k.*

The relevance of this definition in the context of counting the number of gains of $(n\eta_*)^{-1/2}$ lies in the following equivalent characterisation [94].

Lemma 6.5.5. *A graph $\Gamma = (V, E)$ is k-degenerate if and only if there exists an ordering of vertices $\{v_1, \dots, v_n\} = V$ such that for each $m \in [n]$ it holds that*

$$\deg_{\Gamma[\{v_1,\dots,v_m\}]}(v_m) \leq k \tag{6.93}$$

where for $V' \subset V$, $\Gamma[V']$ denotes the induced subgraph on the vertex set V'.

We consider a subset of non-loop edges $E_{\text{Ward}} \subset E_G \setminus \{(vv) \mid v \in V\}$ for which Ward improvements will be obtained. We claim that if $\Gamma_{\text{Ward}} = (V, E_{\text{Ward}})$ is 2-degenerate, then we may gain a factor of $(n\eta_*)^{-1/2}$ from each edge in E_{Ward}. Indeed, take the ordering $\{v_1, \dots, v_{2|E_\kappa|}\}$ guaranteed to exist in Lemma 6.5.5 and first sum up the index a_1 associated with v_1. Since Γ_{Ward} is 2-degenerate there are at most two edges from E_{Ward} adjacent to v_1 and we can gain a factor of $(n\eta_*)^{-1/2}$ for each of them. Next, we can sum up the index associated with vertex v_2 and again gain the same factor for each edge in E_{Ward} adjacent to v_2. Continuing this way we see that in total we can gain a factor of $(n\eta_*)^{-|E_{\text{Ward}}|/2}$ over the naive bound (6.92).

Definition 6.5.6 (Ward estimate)**.** *For a graph Γ with fixed subset $E_{\text{Ward}} \subset E_G$ of edges we define*

$$\text{W-Est}(\Gamma) := \frac{\text{N-Est}(\Gamma)}{(n\eta_*)^{|E_{\text{Ward}}|/2}}.$$

By considering only G-edges adjacent to κ-edges of degrees 2 and 3 it is possible to find such a 2-degenerate set with

$$|E_{\text{Ward}}| = \sum_i (4 - d_G(i))_+$$

elements, cf. [83, Lemma 4.7]. As a consequence, as compared with the first inequality in (6.92), we obtain an improved bound

$$\begin{aligned}
\text{Val}(\Gamma) \prec \text{W-Est}(\Gamma) &= \frac{1}{n^{2p}|\eta_1\eta_2|^p}(n\eta_*)^{-|E_{\text{Ward}}|/2}\prod_i \Big(n^{2-d_G(i)/2}\Big) \\
&= \frac{1}{n^{2p}|\eta_1\eta_2|^p}\prod_{d_G(i)=2}\Big(\frac{n}{n\eta_*}\Big)\prod_{d_G(i)=3}\Big(\frac{\sqrt{n}}{\sqrt{n\eta_*}}\Big)\prod_{d_G(i)\geq 4}\Big(n^{2-d_G(i)/2}\Big) \\
&\lesssim \frac{1}{(n\eta_*)^{2p}|\eta_1\eta_2|^p}\eta_*^{2p+\sum_i(d_G(i)/2-2)} \lesssim \frac{1}{(n\eta_*)^{2p}|\eta_1\eta_2|^p},
\end{aligned} \tag{6.94}$$

where in the penultimate inequality we used $n^{-1} \leq \eta_*$, and in the ultimate inequality that $d_G(i) \geq 2$ and $|E_\kappa| \leq 2p$ which implies that the exponent of η_* is non-negative and $\eta_* \lesssim 1$. Thus we gained a factor of $(n\eta_*)^{-2p}$ over the naive estimate (6.92).

Resummation improvements

The bound (6.94) is optimal if $z_1 = z_2$ and if η_1, η_2 have opposite signs. In the general case $z_1 \neq z_2$ we have to use two additional improvements which both rely on the fact that the summations $\sum_{a_i,b_i}$ corresponding to $(u_i, v_i) \in E_\kappa^2$ can be written as matrix products since $d_G(u_i) = d_G(v_i) = 2$. Therefore we can sum up the G-edges adjacent to $(u_i v_i)$ as

$$\begin{aligned}
&\sum_{a_i b_i} G_{xa_i}G_{a_iy}G_{zb_i}G_{b_iw}R_{a_ib_i} \\
&\quad= \sum_{a_i b_i} G_{xa_i}G_{a_iy}G_{zb_i}G_{b_iw}\Big[\mathbf{1}(a_i > n, b_i \leq n) + \mathbf{1}(a_i \leq n, b_i > n)\Big] \\
&\quad= (GE_1G)_{xy}(GE_2G)_{zw} + (GE_2G)_{xy}(GE_1G)_{zw},
\end{aligned} \tag{6.95a}$$

where E_1, E_2 are defined in (6.80), in the case of four involved G's and $d_G^{\mathrm{in}} = d_G^{\mathrm{out}} = 1$. If one vertex has two incoming, and the other two outgoing edges (which is only possible if X is real), then we similarly can sum up

$$\sum_{ab} G_{xa}G_{ya}G_{bz}G_{bw}R_{ab} = (GE_1G^t)_{xy}(G^tE_2G)_{zw} + (GE_2G^t)_{xy}(G^tE_1G)_{zw}, \quad (6.95b)$$

so merely some G is replaced by its transpose G^t compared to (6.95a) which will not change any estimate. In the remaining cases with two and three involved G's we similarly have

$$\begin{aligned} \sum_{ab} G_{ba}G_{ab}R_{ab} &= \operatorname{Tr} GE_1GE_2 + \operatorname{Tr} GE_2GE_1 \\ \sum_{ab} G_{xa}G_{ab}G_{by}R_{ab} &= (GE_1GE_2G)_{xy} + (GE_2GE_1G)_{xy}. \end{aligned} \quad (6.95c)$$

By carrying out all available *partial summations* at degree-2 vertices as in (6.95) for the value $\mathrm{Val}(\Gamma)$ of some graph Γ we obtain a collection of *reduced graphs*, in which cycles of G's are contracted to the trace of their matrix product, and chains of G's are contracted to single edges, also representing the matrix products with two *external* indices. We denote generic cycle-subgraphs of k edges from E_G with vertices of degree two by Γ_k°, and generic chain-subgraphs of k edges from E_G with *internal* vertices of degree two and external vertices of degree at least three by Γ_k^-. With a slight abuse of notation we denote the *value* of Γ_k° by $\operatorname{Tr}\Gamma_k^\circ$, and the *value* of Γ_k^- with external indices (a, b) by $(\Gamma_k^-)_{ab}$, where for a fixed choice of E_1, E_2 in (6.95) the internal indices are summed up. The actual choice of E_1, E_2 is irrelevant for our analysis, hence we will omit it from the notation. The concept of the *naive* and *Ward* estimates of any graph Γ carry over naturally to these chain and cycle-subgraphs by setting

$$\begin{aligned} \text{N-Est}(\Gamma_k^\circ) &:= \frac{n^k}{|\eta_1\eta_2|^{|E_G^2(\Gamma_k^\circ)|/2}}, \quad \text{N-Est}(\Gamma_k^-) := \frac{n^{k-1}}{|\eta_1\eta_2|^{|E_G^2(\Gamma_k^-)|/2}}, \\ \text{W-Est}(\Gamma_k^{\circ/-}) &= \frac{\text{N-Est}(\Gamma_k^{\circ/-})}{(n\eta_*)^{|E_{\mathrm{Ward}}(\Gamma_k^{\circ/-})|/2}}, \quad E_{\mathrm{Ward}}(\Gamma_k^{\circ/-}) = E_G(\Gamma_k^{\circ/-}) \cap E_{\mathrm{Ward}}(\Gamma). \end{aligned} \quad (6.96)$$

After contracting the chain- and cycle-subgraphs we obtain $2^{|E_\kappa^2|}$ reduced graphs Γ_{red} on the vertex set

$$V(\Gamma_{\mathrm{red}}) := \{v \in V(\Gamma) \mid d_G(v) \geq 3\}$$

with κ-edges

$$E_\kappa(\Gamma_{\mathrm{red}}) := E_\kappa^{\geq 3}(\Gamma)$$

and G-edges

$$E_G(\Gamma_{\mathrm{red}}) := \{(uv) \in E_G(\Gamma) \mid \min\{d_G(u), d_G(v)\} \geq 3\} \cup E_G^{\mathrm{chain}}(\Gamma_{\mathrm{red}}),$$

with additional *chain-edges*

$$E_G^{\mathrm{chain}}(\Gamma_{\mathrm{red}}) := \left\{ (u_1u_{k+1}) \,\middle|\, \begin{array}{l} k \geq 2,\ u_1, u_{k+1} \in V(\Gamma_{\mathrm{red}}),\ \exists \Gamma_k^- \subset \Gamma, \\ V(\Gamma_k^-) = (u_1, \dots, u_{k+1}) \end{array} \right\}.$$

The additional chain edges $(u_1u_{k+1}) \in E_G^{\mathrm{chain}}$ naturally represent the matrices

$$\mathcal{G}^{(u_1u_{k+1})} := \big((\Gamma_k^-)_{ab}\big)_{a,b\in[2n]}$$

whose entries are the values of the chain-subgraphs. Note that due to the presence of E_1, E_2 in (6.95) the matrices associated with some G-edges can be multiplied by E_1, E_2. However, since in the definition (ii) of G-edges the multiplication with generic bounded deterministic matrices is implicitly allowed, this additional multiplication will not be visible in the notation. Note that the reduced graphs contain only vertices of at least degree three, and only κ-edges from $E_\kappa^{\geq 3}$. The definition of value, naive estimate and Ward estimate naturally extend to the reduced graphs and we have

$$\mathrm{Val}(\Gamma) = \sum \mathrm{Val}(\Gamma_{\mathrm{red}}) \prod_{\Gamma_k^\circ \subset \Gamma} \mathrm{Tr}\, \Gamma_k^\circ \tag{6.97}$$

and

$$\begin{aligned} \text{N-Est}(\Gamma) &= \text{N-Est}(\Gamma_{\mathrm{red}}) \prod_{\Gamma_k^\circ \subset \Gamma} \text{N-Est}(\Gamma_k^\circ), \\ \text{W-Est}(\Gamma) &= \text{W-Est}(\Gamma_{\mathrm{red}}) \prod_{\Gamma_k^\circ \subset \Gamma} \text{W-Est}(\Gamma_k^\circ). \end{aligned} \tag{6.98}$$

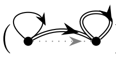
The irrelevant summation in (6.97) of size $2^{|E_\kappa^2|}$ is due to the sums in (6.95).

Let us revisit the examples (6.89) to illustrate the summation procedure. The first two graphs in (6.89) only have degree-2 vertices, so that the reduced graphs are empty with value $n^{-2p} = n^{-2}$, hence

$$\mathrm{Val}(\Gamma) = \frac{1}{n^2} \sum \mathrm{Tr}\, \Gamma_2^\circ \qquad \mathrm{Val}(\Gamma) = \frac{1}{n^2} \sum (\mathrm{Tr}\, \Gamma_2^\circ)(\mathrm{Tr}\, \Gamma_2^\circ),$$

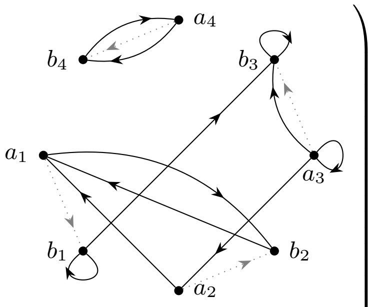
where the summation is over two and, respectively, four terms. The third graph in (6.89) results in no traces but in four reduced graphs

$$\mathrm{Val}(\Gamma) = \sum \mathrm{Val}(\quad),$$

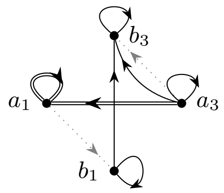


where for convenience we highlighted the chain-edges E_G^{chain} representing Γ_k^- by double lines (note that the two endpoints of a chain edge may coincide, but it is not interpreted as a cycle graph since this common vertex has degree more than two, so it is not summed up into a trace along the reduction process). Finally, to illustrate the reduction for a more complicated graph, we have

$$\mathrm{Val}\left(\right.$$



$$\left.\right) = \sum (\mathrm{Tr}\,\Gamma_2^-)\,\mathrm{Val}\left(\right.$$



$$\left.\right)$$

where we labelled the vertices for convenience, and the summation on the rhs. is over four assignments of E_1, E_2.

Since we have already established a bound on $\mathrm{Val}(\Gamma) \prec \text{W-Est}(\Gamma)$ we only have to identify the additional gain from the resummation compared to the *Ward-estimate* (6.94).

We will need to exploit two additional effects:

1. The Ward-estimate is sub-optimal whenever, after resummation, we have some contracted cycle $\mathrm{Tr}\,\Gamma_k^\circ$ or a reduced graph with a chain-edge Γ_k^- with $k \geq 3$.
2. When estimating $\mathrm{Tr}\,\Gamma_k^\circ$, $k \geq 2$ with Γ_k° containing some G_{12}, then also the improved bound from 1 is sub-optimal and there is an additional gain from using the a priori bound $|\langle G_{12}A\rangle| \prec \theta$.

We now make the additional gains 1–2 precise.

Lemma 6.5.7. *For $k \geq 2$ let Γ_k° and Γ_k^- be some cycle and chain subgraphs.*

1. We have

$$|\mathrm{Tr}\,\Gamma_k^\circ| \prec (n\eta_*)^{-(k-2)/2}\,\text{W-Est}(\Gamma_k^\circ) \tag{6.99a}$$

and for all a, b

$$|(\Gamma_k^-)_{ab}| \prec (n\eta_*)^{-(k-2)/2}\,\text{W-Est}(\Gamma_k^-). \tag{6.99b}$$

2. If Γ_k° contains at least one G_{12} then we have a further improvement of $(\eta_*\theta)^{1/2}$, i.e.

$$|\mathrm{Tr}\,\Gamma_k^\circ| \prec \sqrt{\eta_*\theta}(n\eta_*)^{-(k-2)/2}\,\text{W-Est}(\Gamma_k^\circ), \tag{6.99c}$$

where θ is as in (6.77b).

The proof of Lemma 6.5.7 follows from the following optimal bound on general products $G_{j_1\ldots j_k}$ of resolvents and generic deterministic matrices.

Lemma 6.5.8. *Let* $w_1, w_2, \dots, z_1, z_2, \dots$ *denote arbitrary spectral parameters with* $\eta_i = \Im w_i > 0$. *With* $G_j = G^{z_j}(w_j)$ *we then denote generic products of resolvents* $G_{j_1}, \dots G_{j_k}$ *or their adjoints/transpositions (in that order) with arbitrary bounded deterministic matrices in between by* $G_{j_1 \dots j_k}$, *e.g.* $G_{1i1} = A_1 G_1 A_2 G_i A_3 G_1 A_4$.

- *For* $j_1, \dots j_k$ *we have the isotropic bound*

$$|\langle \boldsymbol{x}, G_{j_1\dots j_k}\boldsymbol{y}\rangle| \prec \|\boldsymbol{x}\|\|\boldsymbol{y}\|\sqrt{\eta_{j_1}\eta_{j_k}}\Big(\prod_{n=1}^k \eta_{j_n}\Big)^{-1}. \tag{6.100a}$$

- *For* $j_1, \dots, j_k$ *and any* $1 \le s < t \le k$ *we have the averaged bound*

$$|\langle G_{j_1\dots j_k}\rangle| \prec \sqrt{\eta_{j_s}\eta_{j_t}}\Big(\prod_{n=1}^k \eta_{j_n}\Big)^{-1}. \tag{6.100b}$$

Lemma 6.5.8 for example implies $|(G_{1i})_{ab}| \prec (\eta_1\eta_i)^{-1/2}$ or $|(G_{i1i})_{ab}| \prec (\eta_1\eta_i)^{-1}$. Note that the averaged bound (6.100b) can be applied more flexibly by choosing s, t freely, e.g.

$$|\langle G_{1i1i}\rangle| \prec \min\{\eta_1^{-1}\eta_i^{-2}, \eta_1^{-2}\eta_i^{-1}\},$$

while $|\langle \boldsymbol{x}, G_{1i1i}\boldsymbol{y}\rangle| \prec \|\boldsymbol{x}\|\|\boldsymbol{y}\|(\eta_1\eta_i)^{-3/2}$.

Proof of Lemma 6.5.8. We begin with

$$\begin{aligned}
&|\langle \boldsymbol{x}, G_{j_1\dots j_k}\boldsymbol{y}\rangle| \\
&\le \sqrt{\langle \boldsymbol{x}, G_{j_1}G_{j_1}^*\boldsymbol{x}\rangle}\sqrt{\langle \boldsymbol{y}, G_{j_2\dots j_k}^* G_{j_2\dots j_k}\boldsymbol{y}\rangle} \prec \frac{\|\boldsymbol{x}\|}{\sqrt{\eta_{j_1}}}\sqrt{\langle \boldsymbol{y}, G_{j_2\dots j_k}^* G_{j_2\dots j_k}\boldsymbol{y}\rangle} \\
&\lesssim \frac{\|\boldsymbol{x}\|}{\sqrt{\eta_{j_1}}}\frac{1}{\eta_{j_2}}\sqrt{\langle \boldsymbol{y}, G_{j_3\dots j_k}^* G_{j_3\dots j_k}\boldsymbol{y}\rangle} \lesssim \dots \\
&\lesssim \frac{\|\boldsymbol{x}\|}{\sqrt{\eta_{j_1}}}\frac{1}{\eta_{j_2}\dots\eta_{j_{k-1}}}\sqrt{\langle \boldsymbol{y}, G_{j_k}^* G_{j_k}\boldsymbol{y}\rangle} \prec \frac{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}{\sqrt{\eta_{j_1}\eta_{j_k}}}\frac{1}{\eta_{j_2}\dots\eta_{j_{k-1}}},
\end{aligned}$$

where in each step we estimated the middle $G_{j_2}^*G_{j_2}, G_{j_3}^*G_{j_3}, \dots$ terms by $1/\eta_{j_2}^2, 1/\eta_{j_3}^2, \dots$, and in the last step we used Ward estimate. This proves (6.100a). We now turn to (6.100b) where by cyclicity without loss of generality we may assume $s = 1$. Thus

$$\begin{aligned}
|\langle G_{j_1\dots j_k}\rangle| &\le \sqrt{\langle G_{j_1\dots j_{t-1}}G_{j_1\dots j_{t-1}}^*\rangle}\sqrt{\langle G_{j_t\dots j_k}^* G_{j_t\dots j_k}\rangle} \\
&= \sqrt{\langle G_{j_1\dots j_{t-1}}G_{j_1\dots j_{t-1}}^*\rangle}\sqrt{\langle G_{j_t\dots j_k} G_{j_t\dots j_k}^*\rangle} \\
&\lesssim \Big(\prod_{n\ne 1,t}\frac{1}{\eta_{j_n}}\Big)\sqrt{\langle G_{j_1}G_{j_1}^*\rangle}\sqrt{\langle G_{j_t}G_{j_t}^*\rangle} \prec \frac{1}{\sqrt{\eta_{j_s}\eta_{j_t}}}\Big(\prod_{n\ne 1,t}\frac{1}{\eta_{j_n}}\Big),
\end{aligned}$$

where in the second step we used cyclicity of the trace, the norm-estimate in the third step und the Ward-estimate in the last step. □

Proof of Lemma 6.5.7. For the proof of (6.99a) we recall from the definition of the Ward-estimate in (6.96) that for a cycle Γ_k° we have

$$\text{W-Est}(\Gamma_k^\circ) \geq \frac{\text{N-Est}(\Gamma_k^\circ)}{(n\eta_*)^{k/2}} = \frac{n^{k/2}}{|\eta_1\eta_2|^{|E_G^2(\Gamma_k^\circ)|/2}} \frac{1}{\eta_*^{k/2}}$$

since $|E_{\text{Ward}}(\Gamma_k^\circ)| \leq |E_G(\Gamma_k^\circ)| \leq k$. Thus, together with (6.100b) and interpreting $\operatorname{Tr}\Gamma_k^\circ$ as a trace of a product of $k + |E_G^2(\Gamma_k^\circ)|$ factors of G's we conclude

$$|\operatorname{Tr}\Gamma_k^\circ| \prec \frac{n}{|\eta_1\eta_2|^{|E_G^2(\Gamma_k^\circ)|}\eta_*^{k-|E_G^2(\Gamma_k^\circ)|-1}} \leq \frac{n}{|\eta_1\eta_2|^{|E_G^2(\Gamma_k^\circ)|/2}\eta_*^{k-1}} \leq \frac{\text{W-Est}(\Gamma_k^\circ)}{(n\eta_*)^{k/2-1}}. \quad (6.101)$$

Note that Lemma 6.5.8 is applicable here even though therein (for convenience) it was assumed that all spectral parameters w_i have positive imaginary parts. However, the lemma also applies to spectral parameters with negative imaginary parts since it allows for adjoints and $G^z(\overline{w}) = (G^z(w))^*$. The first inequality in (6.101) elementarily follows from (6.100b) by distinguishing the cases $|E_G^2| = k, k-1$ or $\leq k-2$, and always choosing s and t such that the $\sqrt{\eta_{j_s}\eta_{j_t}}$ factor contains the highest possible η_* power. Similarly to (6.101), for (6.99b) we have, using (6.100a),

$$|(\Gamma_k^-)_{ab}| \prec \frac{n^{k-1}}{|\eta_1\eta_2|^{|E_G^2(\Gamma_k^-)|/2}} \frac{1}{(n\eta_*)^{k/2}} \leq \frac{\text{W-Est}(\Gamma_k^-)}{(n\eta_*)^{k/2-1}}. \quad (6.102)$$

For the proof of (6.99c) we use a Cauchy-Schwarz estimate to isolate a single G_{12} factor from the remaining G's in Γ_l°. We may represent the "square" of all the remaining factors by an appropriate cycle graph $\Gamma_{2(k-1)}^\circ$ of length $2(k-1)$ with $|E_G^2(\Gamma_{2(k-1)}^\circ)| = 2(|E_G^2(\Gamma_k^\circ)| - 1)$. We obtain

$$\begin{aligned}
|\operatorname{Tr}\Gamma_k^\circ| &\leq \sqrt{\operatorname{Tr}(G_{12}G_{12}^*)}\sqrt{|\operatorname{Tr}\Gamma_{2(k-1)}^\circ|} = \sqrt{\operatorname{Tr}G_1^*G_1BG_2G_2^*B^*}\sqrt{|\operatorname{Tr}\Gamma_{2(k-1)}^\circ|} \\
&= \frac{\sqrt{\operatorname{Tr}(\Im G_1)B(\Im G_2)B^*}\sqrt{|\operatorname{Tr}\Gamma_{2(k-1)}^\circ|}}{\sqrt{|\eta_1\eta_2|}} \\
&\prec \frac{\sqrt{\theta n}}{\sqrt{|\eta_1\eta_2|}} \frac{\sqrt{n}}{|\eta_1\eta_2|^{|E_2^G(\Gamma_k^\circ)|/2-1/2}\eta_*^{k-3/2}} \\
&\leq \sqrt{\eta_*\theta}(n\eta_*)^{-(k-2)/2}\,\text{W-Est}(\Gamma_k^\circ)
\end{aligned}$$

where in the penultimate step we wrote out $\Im G = (G - G^*)/(2\mathrm{i})$ in order to use (6.77b), and used (6.101) for $\Gamma_{2(k-1)}^\circ$. ☐

Now it remains to count the gains from applying Lemma 6.5.7 for each cycle- and chain subgraph of Γ. We claim that

$$\text{W-Est}(\Gamma) \leq (\eta_*^{1/6})^{d_{\geq 3}} \frac{1}{(n\eta_*)^{2p}|\eta_1\eta_2|^p}, \qquad d_{\geq 3} := \sum_{d_G(i)\geq 3} d_G(i). \quad (6.103a)$$

Furthermore, suppose that Γ has c degree-2 cycles Γ_k° which according to 3 has to satisfy $0 \leq c' := |E_\kappa^2| - c \leq |E_\kappa^2|$. Then we claim that

$$|\text{Val}(\Gamma)| \prec \Big(\frac{1}{n\eta_*}\Big)^{(c'-d_{\geq 3}/2)_+} \big(\sqrt{\eta_*\theta}\big)^{(p-c'-d_{\geq 3}/2)_+}\,\text{W-Est}(\Gamma). \quad (6.103b)$$

Assuming (6.103a)–(6.103b) it follows immediately that

$$|\mathrm{Val}(\Gamma)| \prec \frac{1}{(n\eta_*)^{2p}|\eta_1\eta_2|^p}\Big(\sqrt{\eta_*\theta} + \frac{1}{n\eta_*} + \eta_*^{1/6}\Big)^p,$$

implying (6.77c). In order to complete the proof of the Proposition 6.5.3 it remains to verify (6.103a) and (6.103b).

Proof of (6.103a). This follows immediately from the penultimate inequality in (6.94) and

$$\eta_*^{2p+\sum_i (d_G(i)/2-2)} \le \eta_*^{\sum_i (d_G(i)/2-1)} = \eta_*^{\frac{1}{2}\sum_{d_G(i)\ge 3}(d_G(i)-2)} \le \eta_*^{\frac{1}{6}\sum_{d_G(i)\ge 3} d_G(i)},$$

where we used 2 in the first inequality. □

Proof of (6.103b). For cycles Γ_k° or chain-edges Γ_k^- in the reduced graph we say that $\Gamma_k^{\circ/-}$ has $(k-2)_+$ *excess G-edges*. Note that for cycles Γ_k° every additional G beyond the minimal number $k \ge 2$ is counted as an excess G-edge, while for chain-edges Γ_k^- the first additional G beyond the minimal number $k \ge 1$ is not counted as an excess G-edge. We claim that:

a) The total number of excess G-edges is at least $2c' - d_{\ge 3}$.

b) There are at least $p - c' - d_{\ge 3}/2$ cycles in Γ containing G_{12}.

Since the vertices of the reduced graph are u_i, v_i for $d_G(i) \ge 3$, it follows that the reduced graph has $\sum_{d_G(i)\ge 3}(d_G(u_i)+d_G(v_i))/2 = d_{\ge 3}$ edges while the total number of G's beyond the minimally required G's (i.e. two for cycles and one for edges) is $2c'$. Thus in the worst case there are at least $2c' - d_{\ge 3}$ excess G-edges, confirming a).

The total number of G_{12}'s is $2p$, while the total number of G_i's is $2|E_\kappa^2| + d_{\ge 3} - 2p$, according to 5. For fixed c the number of cycles with G_{12}'s is minimised in the case when all G_i's are in cycles of length 2 which results in $|E_\kappa^2| - p + \lfloor d_{\ge 3}/2 \rfloor$ cycles without G_{12}'s. Thus, there are at least

$$c - \Big(|E_\kappa^2| - p + \lfloor d_{\ge 3}/2\rfloor\Big) = p - c' - \lfloor d_{\ge 3}/2 \rfloor \ge p - c' - d_{\ge 3}/2$$

cycles with some G_{12}, confirming also b).

The claim (6.103b) follows from a)-b) in combination with Lemma 6.5.7. □

6.6 Central limit theorem for resolvents

The goal of this section is to prove the CLT for resolvents, as stated in Proposition 6.3.3. We begin by analysing the 2-body stability operator $\widehat{\mathcal{B}}$ from (6.75), as well as its special case, the 1-body stability operator

$$\mathcal{B} := \widehat{\mathcal{B}}(z, z, w, w) = 1 - M\mathcal{S}[\cdot]M. \tag{6.104}$$

Note that other than in the previous Section 6.5, all spectral parameters $\eta, \eta_1, \dots, \eta_p$ considered in the present section are positive, or even, $\eta, \eta_i \ge 1/n$.

Lemma 6.6.1. *For $w_1 = \mathrm{i}\eta_1, w_2 = \mathrm{i}\eta_2 \in \mathrm{i}\mathbf{R} \setminus \{0\}$ and $z_1, z_2 \in \mathbf{C}$ we have*

$$\|\widehat{\mathcal{B}}^{-1}\|^{-1} \gtrsim (|\eta_1| + |\eta_2|) \min\{(\Im m_1)^2, (\Im m_2)^2\} + |z_1 - z_2|^2. \tag{6.105}$$

Moreover, for $z_1 = z_2 = z$ and $w_1 = w_2 = \mathrm{i}\eta$ the operator $\mathcal{B} = \widehat{\mathcal{B}}$ has two non-trivial eigenvalues β, β_* with β, β_* as in (6.22), (6.24), *and the remaining eigenvalues being* 1.

Proof. Throughout the proof we assume that $\eta_1, \eta_2 > 0$, all the other cases are completely analogous. With the shorthand notations $m_i := m^{z_i}(w_i), u_i := u^{z_i}(w_i)$ and the partial trace $\mathrm{Tr}_2 \colon \mathbf{C}^{2n \times 2n} \to \mathbf{C}^4$ rearranged into a 4-dimensional vector, the stability operator $\widehat{\mathcal{B}}$, written as a 4×4 matrix is given by

$$\widehat{\mathcal{B}} = 1 - \mathrm{Tr}_2^{-1} \circ \begin{pmatrix} T_1 & 0 \\ T_2 & 0 \end{pmatrix} \circ \mathrm{Tr}_2, \quad \mathrm{Tr}_2 \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} := \begin{pmatrix} \langle R_{11} \rangle \\ \langle R_{22} \rangle \\ \langle R_{12} \rangle \\ \langle R_{21} \rangle \end{pmatrix}. \tag{6.106}$$

Here we defined

$$T_1 := \begin{pmatrix} z_1 \overline{z_2} u_1 u_2 & m_1 m_2 \\ m_1 m_2 & \overline{z_1} z_2 u_1 u_2 \end{pmatrix}, \quad T_2 := \begin{pmatrix} -z_1 u_1 m_2 & -z_2 u_2 m_1 \\ -\overline{z_2} u_2 m_1 & -\overline{z_1} u_1 m_2 \end{pmatrix},$$

and Tr_2^{-1} is understood to map $\mathbf{C}^4$ into $\mathbf{C}^{2n \times 2n}$ in such a way that each $n \times n$ block is a constant multiple of the identity matrix. From (6.106) it follows that $\widehat{\mathcal{B}}$ has eigenvalue 1 in the $4(n^2 - 1)$-dimensional kernel of Tr_2, and that the remaining four eigenvalues are $1, 1$ and the eigenvalues $\widehat{\beta}, \widehat{\beta}_*$ of $B_1 := 1 - T_1$, i.e.

$$\widehat{\beta}, \widehat{\beta}_* := 1 - u_1 u_2 \Re z_1 \overline{z_2} \pm \sqrt{m_1^2 m_2^2 - u_1^2 u_2^2 (\Im z_1 \overline{z_2})^2}. \tag{6.107}$$

Thus the claim about the $w_1 = w_2$, $z_1 = z_2$ special case follows. The bound (6.105) follows directly from

$$\left| \widehat{\beta} \widehat{\beta}_* \right| \gtrsim (\eta_1 + \eta_2) \min\{(\Im m_1)^2, (\Im m_2)^2\} + |z_1 - z_2|^2, \tag{6.108}$$

since $|\widehat{\beta}|, |\widehat{\beta}_*| \lesssim 1$ and $\|\widehat{\mathcal{B}}^{-1}\| \lesssim \|B_1^{-1}\| = (\min\{|\widehat{\beta}|, |\widehat{\beta}_*|\})^{-1}$ due to B_1 being normal.

We now prove (6.108). By (6.107), using that $u_i = -m_i^2 + u_i^2 |z_i|^2$ repeatedly, it follows that

$$\begin{aligned} \widehat{\beta} \widehat{\beta}_* &= 1 - u_1 u_2 \Big[1 - |z_1 - z_2|^2 + (1 - u_1)|z_1|^2 + (1 - u_2)|z_2|^2 \Big] \\ &= u_1 u_2 |z_1 - z_2|^2 + (1 - u_1)(1 - u_2) - m_1^2 u_2 \left(\frac{1}{u_1} - 1 \right) \\ &\quad - m_2^2 u_1 \left(\frac{1}{u_2} - 1 \right). \end{aligned} \tag{6.109}$$

Then, using $1 - u_i = \eta_i / (\eta_i + \Im m_i) \gtrsim \eta_i / (\Im m_i)$, that $m_i = \mathrm{i} \Im m_i$, and assuming $u_1, u_2 \in [\delta, 1]$, for some small fixed $\delta > 0$, we get that

$$\begin{aligned} \left| \widehat{\beta} \widehat{\beta}_* \right| &\gtrsim |z_1 - z_2|^2 + (\Im m_1)^2 (1 - u_1) + (\Im m_2)^2 (1 - u_2) \\ &\gtrsim |z_1 - z_2|^2 + \min\{(\Im m_1)^2, (\Im m_2)^2\} (2 - u_1 - u_2) \\ &\gtrsim |z_1 - z_2|^2 + \min\{(\Im m_1)^2, (\Im m_2)^2\} \left(\frac{\eta_1}{\Im m_1} + \frac{\eta_2}{\Im m_2} \right). \end{aligned} \tag{6.110}$$

If instead at least one $u_i \in [0, \delta]$ then, by the second equality in the display above, the bound (6.108) is trivial. $\square$

We now turn to the computation of the expectation $\mathbf{E}\langle G^z(\mathrm{i}\eta)\rangle$ to higher precision beyond the approximation $\langle G\rangle \approx \langle M\rangle$. Recall the definition of the 1-body stability operator from (6.104) with non-trivial eigenvalues β, β_* as in (6.22), (6.24).

Lemma 6.6.2. *For $\kappa_4 \neq 0$ we have a correction of order n^{-1} to $\mathbf{E}\langle G\rangle$ of the form*

$$\mathbf{E}\langle G\rangle = \langle M\rangle + \mathcal{E} + \mathcal{O}\Big(\frac{1}{|\beta|}\Big(\frac{1}{n^{3/2}(1+\eta)} + \frac{1}{(n\eta)^2}\Big)\Big), \tag{6.111a}$$

where

$$\frac{1}{|\beta|} = \|(\mathcal{B}^*)^{-1}[1]\| \lesssim \frac{1}{|1-|z|^2| + \eta^{2/3}} \tag{6.111b}$$

and

$$\mathcal{E} := \frac{\kappa_4}{n} m^3\Big(\frac{1}{1-m^2-|z|^2} - 1\Big) = -\frac{\mathrm{i}\kappa_4}{4n}\partial_\eta(m^4). \tag{6.111c}$$

Proof. Using (6.69) we find

$$\begin{aligned}\langle G - M\rangle &= \langle 1, \mathcal{B}^{-1}\mathcal{B}[G-M]\rangle = \langle (\mathcal{B}^*)^{-1}[1], \mathcal{B}[G-M]\rangle \\ &= -\langle M^*(\mathcal{B}^*)^{-1}[1], \underline{WG}\rangle + \langle M^*(\mathcal{B}^*)^{-1}[1], \mathcal{S}[G-M](G-M)\rangle \\ &= -\langle M^*(\mathcal{B}^*)^{-1}[1], \underline{WG}\rangle + \mathcal{O}_\prec\Big(\frac{\|(\mathcal{B}^*)^{-1}[1]\|}{(n\eta)^2}\Big).\end{aligned} \tag{6.112}$$

With

$$A := (\mathcal{B}^*)^{-1}[1]^* M$$

we find from the explicit formula for $\mathcal{B}$ given in (6.106) and (6.24) that

$$\langle MA\rangle = \frac{1-\beta}{\beta} = \frac{1}{1-m^2-|z|^2u^2} - 1 = -\mathrm{i}\partial_\eta m, \tag{6.113}$$

and, using a cumulant expansion we find

$$\mathbf{E}\langle \underline{WGA}\rangle = \sum_{k\ge 2}\sum_{ab}\sum_{\boldsymbol{\alpha}\in\{ab,ba\}^k} \frac{\kappa(ba,\boldsymbol{\alpha})}{k!}\,\mathbf{E}\,\partial_{\boldsymbol{\alpha}}\langle \Delta^{ba} GA\rangle. \tag{6.114}$$

We first consider $k = 2$ where by parity at least one G factor is off-diagonal, e.g.

$$\frac{1}{n^{5/2}}\sum_{a\le n}\sum_{b>n}\mathbf{E}\, G_{ab}G_{aa}(GA)_{bb}$$

and similarly for $a > n$, $b \le n$. By writing $G = M + G - M$ and using the isotropic structure of the local law (6.27) we obtain

$$\begin{aligned}&\frac{1}{n^{5/2}}\sum_{a\le n}\sum_{b>n}\mathbf{E}\, G_{ab}G_{aa}(GA)_{bb} \\ &= \frac{1}{n^{5/2}}\mathbf{E}\, m(MA)_{n+1,n+1}\langle E_1\mathbf{1}, GE_2\mathbf{1}\rangle + \mathcal{O}_\prec\Big(n^2 n^{-5/2}(n\eta)^{-3/2}|\beta|^{-1}\Big) \\ &= \mathcal{O}_\prec\Big(\frac{1}{|\beta| n^{3/2}(1+\eta)} + \frac{1}{|\beta| n^2\eta^{3/2}}\Big),\end{aligned}$$

where $\mathbf{1} = (1, \dots, 1)$ denotes the constant vector of norm $\|\mathbf{1}\| = \sqrt{2n}$. Thus we can bound all $k = 2$ terms by $|\beta|^{-1}(n^{-3/2}(1+\eta)^{-1} + n^{-2}\eta^{-3/2})$.

For $k \geq 4$ we can afford bounding each G entrywise and obtain bounds of $|\beta|^{-1}n^{-3/2}$. Finally, for the $k = 3$ term there is an assignment $(\boldsymbol{\alpha}) = (ab, ba, ab)$ for which all G's are diagonal and which contributes a leading order term given by

$$-\frac{\kappa_4}{2n^3}\sum_{ab}{}' M_{aa}M_{bb}M_{aa}(MA)_{bb} = -\frac{\kappa_4}{n}\langle M\rangle^3\langle MA\rangle, \tag{6.115}$$

where

$$\sum_{ab}{}' := \sum_{a\leq n}\sum_{b>n} + \sum_{a>n}\sum_{b\leq n},$$

and thus

$$\begin{aligned}\sum_{k\geq 2}\sum_{ab}\sum_{\boldsymbol{\alpha}\in\{ab,ba\}^k}\frac{\kappa(ba,\boldsymbol{\alpha})}{k!}\partial_{\boldsymbol{\alpha}}\langle\Delta^{ba}GA\rangle &= -\frac{\kappa_4}{n}\langle M\rangle^3\langle MA\rangle \\ &\quad + \mathcal{O}\Big(\frac{1}{|\beta|n^{3/2}(1+\eta)} + \frac{1}{|\beta|n^2\eta^{3/2}}\Big),\end{aligned} \tag{6.116}$$

concluding the proof. □

We now turn to the computation of higher moments which to leading order due to Lemma 6.6.2 is equivalent to computing

$$\mathbf{E}\prod_{i\in[p]}\langle G_i - M_i - \mathcal{E}_i\rangle, \quad \mathcal{E}_i := \frac{\kappa_4}{n}\langle M_i\rangle^3\langle M_iA_i\rangle, \quad A_i := (\mathcal{B}_i^*)^{-1}[1]^*M_i,$$

with G_i, M_i as in (6.68) for $z_1, \dots, z_k \in \mathbf{C}$, $\eta_1, \dots, \eta_k > 1/n$. Using Lemma 6.6.2, Eq. (6.112), $|\mathcal{E}_i| \lesssim 1/n$ and the high-probability bound

$$|\langle \underline{WG_iA_i}\rangle| \prec \frac{1}{|\beta_i|n\eta_i} \tag{6.117}$$

we have

$$\prod_{i\in[p]}\langle G_i - \mathbf{E}\,G_i\rangle = \prod_{i\in[p]}\langle -\underline{WG_i}A_i - \mathcal{E}_i\rangle + \mathcal{O}_\prec\Big(\frac{\psi}{n\eta}\Big), \quad \psi := \prod_{i\in[p]}\frac{1}{|\beta_i|n|\eta_i|}. \tag{6.118}$$

In order to prove Proposition 6.3.3 we need to compute the leading order term in the local law bound

$$\Big|\prod_{i\in[p]}\langle -\underline{WG_i}A_i - \mathcal{E}_i\rangle\Big| \prec \psi. \tag{6.119}$$

Proof of Proposition 6.3.3. To simplify notations we will not carry the β_i-dependence within the proof because each A_i is of size $\|A_i\| \lesssim |\beta_i|^{-1}$ and the whole estimate is linear in each

$|\beta_i|^{-1}$. We first perform a cumulant expansion in $\underline{WG_1}$ to compute

$$\begin{aligned}
&\mathbf{E}\prod_{i\in[p]}\langle -\underline{WG_i}A_i-\mathcal{E}_i\rangle\\
&\quad=-\langle\mathcal{E}_1\rangle\,\mathbf{E}\prod_{i\neq 1}\langle -\underline{WG_i}A_i-\mathcal{E}_i\rangle\\
&\qquad+\sum_{i\neq 1}\mathbf{E}\,\widetilde{\mathbf{E}}\langle -\widetilde{W}G_1A_1\rangle\langle -\widetilde{W}G_iA_i+\underline{WG_i\widetilde{W}G_i}A_i\rangle\prod_{j\neq 1,i}\langle -\underline{WG_j}A_j-\mathcal{E}_j\rangle\\
&\qquad+\sum_{k\geq 2}\sum_{ab}\sum_{\boldsymbol{\alpha}\in\{ab,ba\}^k}\frac{\kappa(ba,\boldsymbol{\alpha})}{k!}\,\mathbf{E}\,\partial_{\boldsymbol{\alpha}}\Big[\langle -\Delta^{ba}G_1A_1\rangle\prod_{i\neq 1}\langle -\underline{WG_i}A_i-\mathcal{E}_i\rangle\Big],
\end{aligned}\tag{6.120}$$

where $\widetilde{W}$ denotes an independent copy of W with expectation $\widetilde{\mathbf{E}}$, and the underline is understood with respect to W and not $\widetilde{W}$. We now consider the terms of (6.120) one by one. For the second term on the rhs. we use the identity

$$\mathbf{E}\langle WA\rangle\langle WB\rangle=\frac{1}{2n^2}\langle AE_1BE_2+AE_2BE_1\rangle=\frac{\langle AEBE'\rangle}{2n^2},\tag{6.121}$$

where we recall the block matrix definition from (6.80) and follow the convention that E,E' are summed over both choices $(E,E')=(E_1,E_2),(E_2,E_1)$. Thus we obtain

$$\begin{aligned}
&\widetilde{\mathbf{E}}\langle -\widetilde{W}G_1A_1\rangle\langle -\widetilde{W}G_iA_i+\underline{WG_i\widetilde{W}G_i}A_i\rangle\\
&\quad=\frac{1}{2n^2}\langle G_1A_1EG_iA_iE'-G_1A_1E\underline{G_iA_iWG_iE'}\rangle\\
&\quad=\frac{1}{2n^2}\langle G_1A_1EG_iA_iE'+G_1\mathcal{S}[G_1A_1EG_iA_i]G_iE'-\underline{G_1A_1EG_iA_iWG_iE'}\rangle.
\end{aligned}\tag{6.122}$$

Here the self-renormalisation in the last term is defined analogously to (6.71), i.e.

$$\underline{f(W)Wg(W)}:=f(W)Wg(W)-\widetilde{\mathbf{E}}(\partial_{\widetilde{W}}f)(W)\widetilde{W}g(W)-\widetilde{\mathbf{E}}f(W)\widetilde{W}(\partial_{\widetilde{W}}g)(W),$$

which is only well-defined if it is clear to which W the action is associated, i.e. $\underline{WWf(W)}$ would be ambiguous. However, we only use the self-renormalisation notation for $f(W)$, $g(W)$ being (products of) resolvents and deterministic matrices, so no ambiguities should arise. For the first two terms in (6.122) we use $\|M^{z_1,z_i}_{AE_1}\|\lesssim\|\widehat{\mathcal{B}}_{1i}^{-1}\|\lesssim|z_1-z_i|^{-2}$ due to (6.105) and the first bound in (6.76) from Theorem 6.5.2 (estimating the big bracket by 1) to obtain

$$\begin{aligned}
&\langle G_1A_1EG_iA_iE'+G_1\mathcal{S}[G_1A_1EG_iA_i]G_iE'\rangle\\
&\qquad=\langle M^{z_1,z_i}_{A_1E}A_iE'+M^{z_i,z_1}_{E'}\mathcal{S}[M^{z_1,z_i}_{A_1E}A_i]\rangle\\
&\qquad\quad+\mathcal{O}_\prec\Big(\frac{1}{n|z_1-z_i|^4\eta_*^{1i}|\eta_1\eta_i|^{1/2}}+\frac{1}{n^2|z_1-z_i|^4(\eta_*^{1i})^2|\eta_1\eta_i|}\Big),
\end{aligned}\tag{6.123}$$

where $\eta_*^{1i}:=\min\{\eta_1,\eta_i\}$. For the last term in (6.122) we claim that

$$\mathbf{E}|\langle\underline{G_1A_1EG_iA_iWG_iE'}\rangle|^2\lesssim\Big(\frac{1}{n\eta_1\eta_i\eta_*^{1i}}\Big)^2,\tag{6.124}$$

the proof of which we present after concluding the proof of the proposition. Thus, using (6.124) together with (6.117),

$$\begin{aligned}
&\Big| n^{-2}\,\mathbf{E}\langle \underline{G_1 A_1 E G_i A_i W G_i E'}\rangle \prod_{j\neq 1,i}\langle -\underline{WG_j}A_j - \mathcal{E}_i\rangle \Big| \\
&\quad\lesssim \frac{n^\epsilon}{n^2}\Big[\prod_{j\neq 1,i}\frac{1}{n\eta_j}\Big]\Big(\mathbf{E}|\langle \underline{G_1A_1EG_iA_iWG_iE'}\rangle|^2\Big)^{1/2} \\
&\quad\lesssim \frac{n^\epsilon}{n\eta_*^{1i}}\prod_j \frac{1}{n\eta_j}\le \frac{n^\epsilon\psi}{n\eta_*}.
\end{aligned}$$

Together with (6.119) and (6.122)–(6.123) we obtain

$$\begin{aligned}
&\mathbf{E}\,\widetilde{\mathbf{E}}\langle -\widetilde{W}G_1A_1\rangle\langle -\widetilde{W}G_iA_i + \underline{WG_i\widetilde{W}G_i}A_i\rangle \prod_{j\neq 1,i}\langle -\underline{WG_j}A_j - \mathcal{E}_j\rangle \\
&\quad = \frac{V_{1,i}}{2n^2}\,\mathbf{E}\prod_{j\neq 1,i}\langle -\underline{WG_j}A_j - \mathcal{E}_j\rangle \\
&\qquad + \mathcal{O}\left(\psi n^\epsilon\Big(\frac{1}{n\eta_*} + \frac{|\eta_1\eta_i|^{1/2}}{n\eta_*^{1i}|z_1-z_i|^4} + \frac{1}{(n\eta_*^{1i})^2|z_1-z_i|^4}\Big)\right)
\end{aligned}\tag{6.125}$$

since, by an explicit computation the rhs. of (6.123) is given by $V_{1,i}$ as defined in (6.32). Indeed, from the explicit formula for $\mathcal{B}$ it follows that main term on the rhs. of (6.123) can be written as $\widetilde{V}_{1,i}$, where

$$\begin{aligned}
\widetilde{V}_{i,j} :=&\ \frac{2m_im_j\,[2u_iu_j\Re z_i\overline{z_j} + (u_iu_j|z_i||z_j|)^2[s_is_j-4]]}{t_it_j\,[1+(u_iu_j|z_i||z_j|)^2 - m_i^2m_j^2 - 2u_iu_j\Re z_i\overline{z_j}]^2} \\
&+ \frac{2m_im_j(m_i^2+u_i^2|z_i|^2)(m_j^2+u_j^2|z_j|^2)}{t_it_j\,[1+(u_iu_j|z_i||z_j|)^2 - m_i^2m_j^2 - 2u_iu_j\Re z_i\overline{z_j}]^2},
\end{aligned}\tag{6.126}$$

using the notations $t_i := 1 - m_i^2 - u_i^2|z_1|^2$, $s_i := m_i^2 - u_i^2|z_i|^2$. By an explicit computation using the equation (6.19) for m_i, m_j it can be checked that $\widetilde{V}_{i,j}$ can be written as a derivative and is given by $\widetilde{V}_{i,j} = V_{i,j}$ with $V_{i,j}$ from (6.32).

Next, we consider the third term on the rhs. of (6.120) for $k = 2$ and $k\geq 3$ separately. We first claim the auxiliary bound

$$|\langle \boldsymbol{x}, \underline{GBWG}\boldsymbol{y}\rangle| \prec \frac{\|\boldsymbol{x}\|\|\boldsymbol{y}\|\|B\|}{n^{1/2}\eta^{3/2}}.\tag{6.127}$$

Note that (6.127) is very similar to (6.77a) except that in (6.127) both G's have the same spectral parameters z,η and the order of W and G is interchanged. The proof of (6.127) is, however, very similar and we leave details to the reader.

After performing the $\boldsymbol{\alpha}$-derivative in (6.120) via the Leibniz rule, we obtain a product of $t\geq 1$ traces of the types $\langle(\Delta G_i)^{k_i}A_i\rangle$ and $\langle \underline{W(G_i\Delta)^{k_i}G_iA_i}\rangle$ with $k_i\geq 0$, $\sum k_i = k+1$, and $p-t$ traces of the type $\langle \underline{WG_iA_i} + \mathcal{E}_i\rangle$. For the term with multiple self-renormalised

G's, i.e. $\langle \underline{W(G_i\Delta)^{k_i}G_iA_i}\rangle$ with $k_i \geq 1$ we rewrite

$$\begin{aligned}\langle \underline{W(G\Delta)^k GA}\rangle &= \langle \underline{GAW(G\Delta)^k}\rangle \\ &= \langle \underline{GAWG}\Delta(G\Delta)^{k-1}\rangle + \sum_{j=1}^{k-1}\langle GA\mathcal{S}[(G\Delta)^j G](G\Delta)^{k-j}\rangle \\ &= \langle \underline{GAWG}\Delta(G\Delta)^{k-1}\rangle + \sum_{j=1}^{k-1}\langle GAE(G\Delta)^{k-j}\rangle\langle GE'(G\Delta)^j\rangle.\end{aligned} \tag{6.128}$$

Case $k = 2$, $t = 1$.

In this case the only possible term is given by $\langle \Delta G_1\Delta G_1\Delta G_1 A_1\rangle$ where by parity at least one $G = G_1$ is off-diagonal and in the worst case (only one off-diagonal factor) we estimate

$$\begin{aligned}n^{-1-3/2}\sum_{a\leq n}\sum_{b>n} G_{aa}G_{bb}(GA)_{ab} &= \frac{m^2}{n^{5/2}}\langle E_1\mathbf{1}, GAE_2\mathbf{1}\rangle + \mathcal{O}_\prec\Big(\frac{1}{n^{1/2}}\frac{1}{(n\eta_1)^{3/2}}\Big) \\ &= \mathcal{O}_\prec\Big(\frac{1}{n^{3/2}} + \frac{1}{n^2\eta_1^{3/2}}\Big),\end{aligned}$$

after replacing $G_{aa} = m + (G-M)_{aa}$ and using the isotropic structure of the local law in (6.27), and similarly for $\sum_{a>n}\sum_{b\leq n}$.

Case $k = 2$, $t = 2$.

In this case there are $2+2$ possible terms

$$\begin{aligned}&\langle \Delta G_1\Delta G_1 A_1\rangle\langle \Delta G_i A_i + \underline{WG_i\Delta G_i}A_i\rangle \\ &\qquad + \langle \Delta G_1 A_1\rangle\langle \Delta G_i\Delta G_i A_i + \underline{WG_i\Delta G_i\Delta G_i}A_i\rangle.\end{aligned}$$

For the first two, in the worst case, we have the estimate

$$\begin{aligned}&\frac{1}{n^{7/2}}\sum_{ab}{}'(G_1)_{aa}(G_1A_1)_{bb}\Big((G_iA_i)_{ab} + (\underline{G_iA_iWG_i})_{ab}\Big) \\ &\qquad = \mathcal{O}_\prec\Big(\frac{1}{n^{5/2}} + \frac{1}{n^3\eta_1\eta_i^{3/2}}\Big)\end{aligned}$$

using (6.127), where we recall the definition of $\sum'$ from (6.115). Similarly, using (6.128) and (6.127) for the ultimate two terms, we have the bound

$$\begin{aligned}&\frac{1}{n^{7/2}}\mathbf{E}\sum_{ab}{}'(G_1A_1)_{ab}\Big((\underline{G_iA_iWG_i})_{aa}(G_i)_{bb} + \frac{(G_iA_iEG_i)_{ab}(G_iE'G_i)_{ab}}{n}\Big) \\ &= \mathcal{O}_\prec\Big(\frac{1}{n^3\eta_1^{1/2}\eta_i^2}\Big).\end{aligned}$$

Case $k = 2, t = 3$.

In this final $k = 2$ case we have to consider four terms

$$\langle \Delta G_1 A_1 \rangle \langle \Delta G_i A_i + \underline{W G_i \Delta G_i A_i} \rangle \langle \Delta G_j A_j + \underline{W G_j \Delta G_j A_j} \rangle,$$

which, using (6.127), we estimate by

$$\begin{aligned} &\frac{1}{n^{9/2}} \sum_{ab}{}' (G_1 A_1)_{ab} \Big((G_i A_i)_{ab} + (\underline{G_i A_i W G_i})_{ab} \Big) \Big((G_j A_j)_{ab} + (\underline{G_j A_j W G_j})_{ab} \Big) \\ &\quad = \mathcal{O}_\prec \Big(\frac{1}{n^4 \eta_1^{1/2} \eta_i^{3/2} \eta_j^{3/2}} \Big). \end{aligned}$$

By inserting the above estimates back into (6.120), after estimating all untouched traces by $n^\epsilon/(n\eta_i)$ in high probability using (6.117), we obtain

$$\begin{aligned} &\sum_{k=2} \sum_{ab} \sum_{\boldsymbol{\alpha} \in \{ab, ba\}^k} \frac{\kappa(ba, \boldsymbol{\alpha})}{k!} \mathbf{E}\, \partial_{\boldsymbol{\alpha}} \Big[\langle -\Delta^{ba} G_1 A_1 \rangle \prod_{i \neq 1} \langle -\underline{W G_i} A_i - \mathcal{E}_i \rangle \Big] \\ &\quad = \mathcal{O}\Big(\frac{\psi n^\epsilon}{\sqrt{n \eta_*}} \Big). \end{aligned} \tag{6.129}$$

Case $k \geq 3$.

In case $k \geq 3$ after the action of the derivative in (6.120) there are $1 \leq t \leq k+1$ traces involving some Δ. By writing the normalised traces involving Δ as matrix entries we obtain a prefactor of $n^{-t-(k+1)/2}$ and a $\sum_{ab}$-summation over entries of $k+1$ matrices of the type G, GA, $\underline{GAWG}$ such that each summation index appears exactly $k+1$ times. There are some additional terms from the last sum in (6.128) which are smaller by a factor $(n\eta)^{-1}$ and which can be bounded exactly as in the $k = 2$ case. If there are only diagonal G or GA-terms, then we have a naive bound of $n^{-t-(k-3)/2}$ and therefore potentially some leading-order contribution in case $k = 3$. If, however, $k > 3$, or there are some off-diagonal G, GA or some $\underline{GAWG}$ terms, then, using (6.127) we obtain an improvement of at least $(n\eta)^{-1/2}$ over the naive bound (6.119). For $k = 3$, by parity, the only possibility of having four diagonal G, GA factors, is distributing the four Δ's either into a single trace or two traces with two Δ's each. Thus the relevant terms are

$$\langle \Delta G_1 \Delta G_1 \Delta G_1 \Delta G_1 A_1 \rangle, \quad \langle \Delta G_1 \Delta G_1 A_1 \rangle \langle \Delta G_i \Delta G_i A_i \rangle.$$

For the first one we recall from (6.116) for $k = 3$ that

$$\sum_{ab} \sum_{\boldsymbol{\alpha}} \kappa(ba, \boldsymbol{\alpha}) \langle \Delta^{ba} G_1 \Delta^{\alpha_1} G_1 \Delta^{\alpha_2} G_1 \Delta^{\alpha_3} G_1 A_1 \rangle = \mathcal{E}_1 + \mathcal{O}_\prec \Big(\frac{1}{n^{3/2}} + \frac{1}{n^2 \eta_1^{3/2}} \Big). \tag{6.130}$$

For the second one we note that only choosing $\boldsymbol{\alpha} = (ab, ab, ba), (ab, ba, ab)$ gives four diagonal factors, while any other choice gives at least two off-diagonal factors. Thus

$$\begin{aligned}
&\sum_{ab}\sum_{\boldsymbol{\alpha}} \kappa(ba,\boldsymbol{\alpha})\langle \Delta^{ba} G_1 \Delta^{\alpha_1} G_1\rangle\langle \Delta^{\alpha_2} G_i \Delta^{\alpha_3} G_i A_i\rangle \\
&= \frac{\kappa_4}{n^2}\sum_{ab}{}' \langle \Delta^{ba} G_1 \Delta^{ab} G_1 A_1\rangle \big[\langle \Delta^{ab} G_i \Delta^{ba} G_i A_i\rangle + \langle \Delta^{ba} G_i \Delta^{ab} G_i A_i\rangle\big] + \mathcal{O}_\prec(\mathcal{E}) \\
&= \frac{\kappa_4}{4n^4}\sum_{ab}{}' (G_1)_{aa}(G_1A_1)_{bb}\big[(G_i)_{bb}(G_iA_i)_{aa} + (G_i)_{aa}(G_iA_i)_{bb}\big] + \mathcal{O}_\prec(\mathcal{E}) \\
&= \frac{\kappa_4}{4n^4}\sum_{ab}{}' m_1 m_i (M_1A_1)_{bb}\big[(M_iA_i)_{aa} + (M_iA_i)_{bb}\big] + \mathcal{O}_\prec\left(\sqrt{n\eta_*}\mathcal{E}\right) \\
&= \frac{\kappa_4}{n^2}\langle M_1\rangle\langle M_i\rangle\langle M_1A_1\rangle\langle M_iA_i\rangle + \mathcal{O}_\prec\left(\frac{1}{n^{5/2}\eta_*^{1/2}}\right),
\end{aligned} \tag{6.131}$$

where $\mathcal{E} := (n^3\eta_*)^{-1}$. We recall from (6.113) that

$$\langle M_1\rangle\langle M_i\rangle\langle M_1A_1\rangle\langle M_iA_i\rangle = \frac{1}{2}U_1U_i$$

with U_i defined in (6.32). Thus, we can conclude for the $k \geq 3$ terms in (6.120) that

$$\begin{aligned}
&\sum_{k\geq 3}\sum_{ab}\sum_{\boldsymbol{\alpha}\in\{ab,ba\}^k} \frac{\kappa(ba,\boldsymbol{\alpha})}{k!}\,\mathbf{E}\,\partial_{\boldsymbol{\alpha}}\Big[\langle -\Delta^{ba}G_1A_1\rangle \prod_{i\neq 1}\langle -\underline{WG_i}A_i - \mathcal{E}_i\rangle\Big] \\
&\quad = \langle \mathcal{E}_1\rangle\,\mathbf{E}\prod_{i\neq 1}\langle -\underline{WG_i}A_i - \mathcal{E}_i\rangle + \sum_{i\neq 1}\frac{\kappa_4 U_1U_i}{2n^2}\,\mathbf{E}\prod_{j\neq 1,i}\langle -\underline{WG_j}A_j - \mathcal{E}_j\rangle \\
&\qquad + \mathcal{O}\Big(\frac{\psi n^\epsilon}{(n\eta_*)^{1/2}}\Big).
\end{aligned} \tag{6.132}$$

By combining (6.120) with (6.125), (6.129) and (6.132) we obtain

$$\begin{aligned}
\mathbf{E}\prod_i\langle -\underline{WG_i}A_i - \mathcal{E}_i\rangle &= \sum_{i\neq 1}\frac{V_{1,i} + \kappa_4 U_1U_i}{2n^2}\,\mathbf{E}\prod_{j\neq 1,i}\langle -\underline{WG_j}A_j - \mathcal{E}_j\rangle \\
&\quad + \mathcal{O}\left(\frac{\psi n^\epsilon}{\sqrt{n\eta_*}} + \frac{\psi n^\epsilon}{n\eta_*^{1/2}|z_1 - z_i|^4} + \frac{\psi n^\epsilon}{(n\eta_*)^2|z_1-z_i|^4}\right),
\end{aligned} \tag{6.133}$$

and thus by induction

$$\begin{aligned}
\mathbf{E}\prod_i\langle -\underline{WG_i}A_i - \mathcal{E}_i\rangle &= \frac{1}{n^p}\sum_{P\in\Pi_p}\prod_{\{i,j\}\in P}\frac{V_{i,j} + \kappa_4 U_iU_j}{2} \\
&\quad + \mathcal{O}\left(\frac{\psi n^\epsilon}{\sqrt{n\eta_*}} + \frac{\psi n^\epsilon}{n\eta_*^{1/2}|z_1 - z_i|^4} + \frac{\psi n^\epsilon}{(n\eta_*)^2|z_1-z_i|^4}\right),
\end{aligned} \tag{6.134}$$

from which the equality $\mathbf{E}\prod_i\langle G_i - \mathbf{E}\,G_i\rangle$ and the second line of (6.30) follows, modulo the proof of (6.124). The remaining equality then follows from applying the very same equality for each element of the pairing. Finally, (6.33) follows directly from Lemma 6.6.2. □

Proof of (6.124). Using the notation of Lemma 6.5.8, our goal is to prove that

$$\mathbf{E}|\langle \underline{WG_{i1i}}\rangle|^2 \lesssim \Big(\frac{1}{n\eta_1\eta_i\eta_*^{1i}}\Big)^2. \tag{6.135}$$

Since only η_1, η_i play a role within the proof of (6.124), we drop the indices from η_*^{1i} and simply write $\eta_* = \eta_*^{1i}$. Using a cumulant expansion we compute

$$\begin{aligned}
&\mathbf{E}|\langle \underline{WG_{i1i}}\rangle|^2\\
&= \mathbf{E}\,\widetilde{\mathbf{E}}\langle \widetilde{W}G_{i1i}\rangle\Big(\langle \widetilde{W}G_{i1i}\rangle + \langle \underline{WG_i\widetilde{W}G_{i1i}} + \underline{WG_{i1}\widetilde{W}G_{1i}} + \underline{WG_{i1i}\widetilde{W}G_i}\rangle\Big)\\
&\quad + \sum_{k\ge 2}\mathcal{O}\Big(\frac{1}{n^{(k+1)/2}}\Big)\sum_{ab}{}'\sum_{k_1+k_2=k-1}\sum_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2}\mathbf{E}\langle \Delta^{ab}\partial_{\boldsymbol{\alpha}_1}G_{i1i}\rangle\langle \Delta^{ab}\partial_{\boldsymbol{\alpha}_2}G_{i1i}\rangle\\
&\quad + \sum_{k\ge 2}\mathcal{O}\Big(\frac{1}{n^{(k+1)/2}}\Big)\sum_{ab}{}'\sum_{k_1+k_2=k}\sum_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2}\mathbf{E}\langle \Delta^{ab}\partial_{\boldsymbol{\alpha}_1}G_{i1i}\rangle\langle \underline{W\partial_{\boldsymbol{\alpha}_2}G_{i1i}}\rangle,
\end{aligned} \tag{6.136}$$

where $\boldsymbol{\alpha}_i$ is understood to be summed over $\boldsymbol{\alpha}_i \in \{ab, ba\}^{k_i}$. In (6.136) we only kept the scaling $|\kappa(ab,\boldsymbol{\alpha})| \lesssim n^{-(k+1)/2}$ of the cumulants, and also absorb combinatorial factors as $k!$ in $\mathcal{O}(\cdot)$. We first consider those terms in (6.136) which contain no self-renormalisations $\underline{Wf(W)}$ anymore since those do not have to be expanded further. For the very first term we obtain

$$\widetilde{\mathbf{E}}\langle \widetilde{W}G_{i1i}\rangle\langle \widetilde{W}G_{i1i}\rangle = \frac{\langle G_{i1ii1i}\rangle}{n^2} = \mathcal{O}_\prec\Big(\frac{1}{n^2\eta_1^2\eta_i^3}\Big). \tag{6.137}$$

To bound products of G_1 and G_i we use Lemma 6.5.8. For the second line on the rhs. of (6.136) we have to estimate

$$\mathcal{O}\Big(\frac{1}{n^{(k+1)/2+2}}\Big)\sum_{k\ge 2}\sum_{ab}{}'\sum_{k_1+k_2=k-1}\sum_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2}\mathbf{E}(\partial_{\boldsymbol{\alpha}_1}(G_{i1i})_{ba})(\partial_{\boldsymbol{\alpha}_2}(G_{i1i})_{ba})$$

and we note that without derivatives we have the estimate $|(G_{i1i})| \prec (\eta_1\eta_i)^{-1}$. Additional derivatives do not affect this bound since if e.g. G_i is derived we obtain one additional G_i but also one additional product of G's with G_i in the end, and one additional product with G_i in the beginning. Due to the structure of the estimate (6.100a) the bound thus remains invariant. For example $|(\partial_{ab}G_{i1i})_{ba}| = |(G_i)_{bb}(G_{i1i})_{aa} + \ldots| \prec (\eta_1\eta_i)^{-1}$. Thus, by estimating the sum trivially we obtain

$$\frac{1}{n^{(k+1)/2}}\sum_{\substack{k_1+k_2=k-1\\ k\ge 2}}\sum_{ab}{}'\sum_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2}\mathbf{E}\langle \Delta^{ab}\partial_{\boldsymbol{\alpha}_1}G_{i1i}\rangle\langle \Delta^{ab}\partial_{\boldsymbol{\alpha}_2}G_{i1i}\rangle = \mathcal{O}_\prec\Big(\frac{1}{n^{3/2}\eta_1^2\eta_i^2}\Big) \tag{6.138}$$

since $k \ge 2$.

It remains to consider the third line on the rhs. of (6.136) and the remaining terms from the first line. In both cases we perform a second cumulant expansion and again differentiate the Gaussian (i.e. the second order cumulant) term, and the terms from higher order cumulants. Since the two consecutive cumulant expansions commute it is clearly sufficient to consider the Gaussian term for the first line, and the full expansion for the third line. We begin with the latter and compute

$$\begin{aligned}
&\mathbf{E}\langle \Delta^{ab}\partial_{\boldsymbol{\alpha}_1} G_{i1i}\rangle \langle \underline{W\partial_{\boldsymbol{\alpha}_2} G_{i1i}}\rangle \\
&\quad = \widetilde{\mathbf{E}}\,\mathbf{E}\langle \Delta^{ab}\partial_{\boldsymbol{\alpha}_1}(G_i\widetilde{W}G_{i1i} + G_{i1}\widetilde{W}G_{1i} + G_{i1i}\widetilde{W}G_i)\rangle\langle \widetilde{W}\partial_{\boldsymbol{\alpha}_2} G_{i1i}\rangle \\
&\qquad + \sum_{l\ge 2}\sum_{cd}{}' \sum_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2} \mathbf{E}\langle \Delta^{ab}\partial_{\boldsymbol{\alpha}_1}\partial_{\boldsymbol{\beta}_1} G_{i1i}\rangle\langle \Delta^{cd}\partial_{\boldsymbol{\alpha}_2}\partial_{\boldsymbol{\beta}_2} G_{i1i}\rangle \\
&\quad = \frac{1}{n^2}\,\mathbf{E}\langle \partial_{\boldsymbol{\alpha}_1}(G_{i1i}\Delta^{ab}G_i + G_{1i}\Delta^{ab}G_{i1} + G_i\Delta^{ab}G_{i1i})\partial_{\boldsymbol{\alpha}_2}(G_{i1i})\rangle \\
&\qquad + \sum_{l\ge 2}\sum_{cd}{}' \sum_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2} \mathbf{E}\langle \Delta^{ab}\partial_{\boldsymbol{\alpha}_1}\partial_{\boldsymbol{\beta}_1} G_{i1i}\rangle\langle \Delta^{cd}\partial_{\boldsymbol{\alpha}_2}\partial_{\boldsymbol{\beta}_2} G_{i1i}\rangle,
\end{aligned} \tag{6.139}$$

where $\boldsymbol{\beta}_i$ are understood to be summed over $\boldsymbol{\beta}_i \in \{cd, dc\}^{l_i}$ with $l_1 + l_2 = l$. After inserting the first line of (6.139) back into (6.136) we obtain an overall factor of $n^{-3-(k+1)/2}$ as well as the $\sum_{ab}$-summation over some $\partial_{\boldsymbol{\alpha}}(\mathcal{G})_{ab}$, where $\mathcal{G}$ is a product of either $2+5$ or $3+4$ G_1's and G_i's respectively with G_i in beginning and end. We can bound $|\partial_{\boldsymbol{\alpha}}(\mathcal{G})_{ab}| \prec \eta_1^{-2}\eta_i^{-4} + \eta_1^{-3}\eta_i^{-3} \le \eta_1^{-2}\eta_i^{-2}\eta_*^{-2}$ and thus can estimate the sum by $n^{-5/2}\eta_1^{-2}\eta_i^{-2}\eta_*^{-2}$ since $k \ge 2$. Here we used (6.100a) to estimate all matrix elements of the form $\mathcal{G}'_{ab}, \mathcal{G}'_{aa}, \dots$ emerging after performing the derivative $\partial_{\boldsymbol{\alpha}}(\mathcal{G})_{ab}$.

Now we turn to the second line of (6.139) when inserted back into (6.136), where we obtain a total prefactor of $n^{-(k+l)/2-3}$, a summation $\sum_{abcd}$ over $(\partial_{\boldsymbol{\alpha}_1}\partial_{\boldsymbol{\beta}_1} G_{i1i})_{ab}(\partial_{\boldsymbol{\alpha}_2}\partial_{\boldsymbol{\beta}_2} G_{i1i})_{cd}$. In case $k = l = 2$, by parity, after performing the derivatives at least two factors are off-diagonal, while in case $k + l = 5$ at least one factor is off-diagonal. Thus we obtain a bound of $n^{1-(k+l)/2}\eta_1^{-2}\eta_i^{-2}$ multiplied by a Ward-improvement of $(n\eta_*)^{-1}$ in the first, and $(n\eta_*)^{-1/2}$ in the second case. Thus we conclude

$$\frac{1}{n^{(k+1)/2}} \sum_{\substack{k_1+k_2=k\\ k\ge 2}} \sum_{ab}{}' \sum_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2} \mathbf{E}\langle \Delta^{ab}\partial_{\boldsymbol{\alpha}_1} G_{i1i}\rangle \langle \underline{W\partial_{\boldsymbol{\alpha}_2} G_{i1i}}\rangle = \mathcal{O}\Big(\frac{1}{n^2\eta_1^2\eta_i^2\eta_*^2}\Big). \tag{6.140}$$

Finally, we consider the Gaussian part of the cumulant expansion of the remaining terms in the first line of (6.136), for which we obtain

$$\frac{1}{n^2}\widetilde{\mathbf{E}}\langle (G_{i1i}\widetilde{W}G_i + G_{1i}\widetilde{W}G_{i1} + G_i\widetilde{W}G_{i1i})^2\rangle = O_\prec\Big(\frac{1}{n^2\eta_1^2\eta_i^2\eta_*^2}\Big) \tag{6.141}$$

since

$$\begin{aligned}
&|\langle G_iG_i\rangle| \prec \frac{1}{\eta_i}, \quad |\langle G_iG_{i1}\rangle| \prec \frac{1}{\eta_1\eta_i}, \quad |\langle G_iG_{i1i}\rangle| \prec \frac{1}{\eta_1\eta_i^2}, \\
&|\langle G_{1i}G_{1i}\rangle| \prec \frac{1}{\eta_1^2\eta_i}, \quad |\langle G_{1i}G_{i1i}\rangle| \prec \frac{1}{\eta_1^2\eta_i^2}, \quad |\langle G_{i1i}G_{i1i}\rangle| \prec \frac{1}{\eta_1^2\eta_i^3}
\end{aligned}$$

due to (6.100b). By combining (6.137)–(6.141) we conclude the proof of (6.124) using (6.136). □

6.7 Independence of the small eigenvalues of H^{z_1} and H^{z_2}

Given an $n \times n$ i.i.d. complex matrix X, for any $z \in \mathbf{C}$ we recall that the Hermitisation of $X - z$ is given by

$$H^z := \begin{pmatrix} 0 & X - z \\ X^* - \overline{z} & 0 \end{pmatrix}. \tag{6.142}$$

The block structure of H^z induces a symmetric spectrum with respect to zero, i.e. denoting by $\{\lambda_{\pm i}^z\}_{i=1}^n$ the eigenvalues of H^z, we have that $\lambda_{-i}^z = -\lambda_i^z$ for any $i \in [n]$. Denote the resolvent of H^z by G^z, i.e. on the imaginary axis G^z is defined by $G^z(\mathrm{i}\eta) := (H^z - \mathrm{i}\eta)^{-1}$, with $\eta > 0$.

Convention 6.7.1. *We omitted the index $i = 0$ in the definition of the eigenvalues of H^z. In the remainder of this section we always assume that all the indices are not zero, e.g we use the notation*

$$\sum_{j=-n}^{n} := \sum_{j=-n}^{-1} + \sum_{j=1}^{n},$$

$|i| \leq A$, for some $A > 0$, to denote $0 < |i| \leq A$, etc.

The main result of this section is the proof of Proposition 6.3.5 which follows by Proposition 6.7.2 and rigidity estimates in Section 6.7.1.

Proposition 6.7.2. *Fix $p \in \mathbf{N}$. For any $\omega_d, \omega_f, \omega_h > 0$ sufficiently small constants such that $\omega_h \ll \omega_f$, there exits $\omega, \widehat{\omega}, \delta_0, \delta_1 > 0$ with $\omega_h \ll \delta_m \ll \widehat{\omega} \ll \omega \ll \omega_f$, for $m = 0, 1$, such that for any fixed $z_1, \dots, z_p \in \mathbf{C}$ such that $|z_l| \leq 1 - n^{-\omega_h}$, $|z_l - z_m| \geq n^{-\omega_d}$, with $l, m \in [p]$, $l \neq m$, it holds*

$$\begin{aligned} \mathbf{E} \prod_{l=1}^{p} \frac{1}{n} \sum_{|i_l| \leq n^{\widehat{\omega}}} \frac{\eta_l}{(\lambda_{i_l}^{z_l})^2 + \eta_l^2} &= \prod_{l=1}^{p} \mathbf{E} \frac{1}{n} \sum_{|i_l| \leq n^{\widehat{\omega}}} \frac{\eta_l}{(\lambda_{i_l}^{z_l})^2 + \eta_l^2} \\ &+ \mathcal{O}\left(\frac{n^{\widehat{\omega}}}{n^{1+\omega}} \sum_{l=1}^{p} \frac{1}{\eta_l} \times \prod_{m=1}^{p} \left(1 + \frac{n^{\xi}}{n \eta_m}\right) + \frac{n^{p\xi + 2\delta_0} n^{\omega_f}}{n^{3/2}} \sum_{l=1}^{p} \frac{1}{\eta_l} + \frac{n^{p\delta_0 + \delta_1}}{n^{\widehat{\omega}}} \right), \end{aligned} \tag{6.143}$$

for any $\xi > 0$, where $\eta_1, \dots, \eta_p \in [n^{-1-\delta_0}, n^{-1+\delta_1}]$ and the implicit constant in $\mathcal{O}(\cdot)$ may depend on p.

We recall that the eigenvalues of H^z are labelled by $\lambda_{-n} \leq \dots \leq \lambda_{-1} \leq \lambda_1 \leq \dots \lambda_n$, hence the summation over $|i_l| \leq n^{\widehat{\omega}}$ in (6.143) is over the smallest (in absolute value) eigenvalues of H^z.

The remainder of Section 6.7 is divided as follows: in Section 6.7.1 we state rigidity of the eigenvalues of the matrices H^{z_l} and a local law for $\operatorname{Tr} G^{z_l}$, then using these results and Proposition 6.7.2 we conclude the proof of Proposition 6.3.5. In Section 6.7.2 we state the main technical results needed to prove Proposition 6.7.2 and conclude its proof. In Section 6.7.3 we estimate the overlaps of eigenvectors, corresponding to small indices, of H^{z_l}, H^{z_m} for $l \neq m$, this is the main input to prove the asymptotic independence in Proposition 6.7.2. In Section 6.7.4 we present Proposition 6.7.13 which is a modification of the pathwise coupling of DBMs from [42, 129] (adapted to the 2×2 matrix model (6.142) in [54]) which is needed to deal with the (small) correlation of $\boldsymbol{\lambda}^{z_l}$, the eigenvalues of H^{z_l}, for different l's. In Section 6.7.5 we prove some technical lemmata used in Section 6.7.2. Finally, in Section 6.7.6 we prove Proposition 6.7.13.

6.7.1 Rigidity of eigenvalues and proof of Proposition 6.3.5

In this section, before proceeding with the actual proof of Proposition 6.7.2, we state the local law away from the imaginary axis, proven in [60], that will be used in the following sections. We remark that the averaged and entry-wise version of this local law for $|z| \leq 1-\epsilon$, for some small fixed $\epsilon > 0$, has already been established in [44, Theorem 3.4].

Proposition 6.7.3 (Theorem 3.1 of [60]). *Let $\omega_h > 0$ be sufficiently small, and define $\delta_l := 1 - |z_l|^2$. Then with very high probability it holds*

$$\left| \frac{1}{2n} \sum_{1 \leq |i| \leq n} \frac{1}{\lambda_i^{z_l} - w} - m^{z_l}(w) \right| \leq \frac{\delta_l^{-100} n^{\xi}}{n \Im w}, \tag{6.144}$$

uniformly in $|z_l|^2 \leq 1 - n^{-\omega_h}$ and $0 < \Im w \leq 10$. Here m^{z_l} denotes the solution of (6.19).

Note that $\delta_l := 1 - |z_l|^2$ introduced in Proposition 6.7.3 are not to be confused with the exponents δ_0, δ_1 introduced in Proposition 6.7.2.

Let $\{\lambda_{\pm i}^z\}_{i=1}^n$ denote the eigenvalues of H^z, and recall that $\rho^z(E) = \pi^{-1} \Im m^z(E + \mathrm{i}0)$ is the limiting (self-consistent) density of states. Then by Proposition 6.7.3 the rigidity of λ_i^z follows by a standard application of Helffer-Sjöstrand formula (see e.g. [81, Lemma 7.1, Theorem 7.6] or [93, Section 5] for a detailed derivation):

$$|\lambda_i^z - \gamma_i^z| \leq \frac{\delta^{-100} n^{\xi}}{n}, \qquad |i| \leq cn, \tag{6.145}$$

with $c > 0$ a small constant and $\delta := 1 - |z|^2$, with very high probability, uniformly in $|z| \leq 1 - n^{-\omega_h}$. The quantiles γ_i^z are defined by

$$\frac{i}{n} = \int_0^{\gamma_i^z} \rho^z(E) \, \mathrm{d}E, \qquad 1 \leq i \leq n, \tag{6.146}$$

and $\gamma_{-i}^z := -\gamma_i^z$ for $-n \leq i \leq -1$. Note that by (6.146) it follows that $\gamma_i^z \sim i/(n\rho^z(0))$ for $|i| \leq n^{1-10\omega_h}$, where $\rho^z(0) = \Im m^z(0) = (1 - |z|^2)^{1/2}$ for $|z| < 1$ by (6.25).

Using the rigidity bound in (6.145), by Proposition 6.7.2 we conclude the proof of Proposition 6.3.5.

Proof of Proposition 6.3.5. Let $z_1, \dots, z_p$ such that $|z_l| \leq 1 - n^{-\omega_h}$ and $|z_l - z_m| \geq n^{-\omega_d}$, for any $l, m \in [p]$, with ω_d, ω_h defined in Proposition 6.3.5. Let $\omega, \widehat{\omega}, \delta_0, \delta_1$ be as in Proposition 6.7.2, i.e.

$$\omega_h \ll \delta_m \ll \widehat{\omega} \ll \omega \ll \omega_f,$$

for $m = 0, 1$. For a detailed summary about all the different scales in the proof of Proposition 6.7.2 and so of Proposition 6.3.5 see Section 6.7.2.3 later. Write

$$\langle G^{z_l}(\mathrm{i}\eta_l) \rangle = \frac{\mathrm{i}}{2n} \left[\sum_{|i| \leq \widehat{\omega}} + \sum_{\widehat{\omega} < |i| \leq n} \right] \frac{\eta_l}{(\lambda_i^{z_l})^2 + \eta_l^2}, \tag{6.147}$$

for $\eta_l \in [n^{-1-\delta_0}, n^{-1+\delta_1}]$. As a consequence of Proposition 6.7.2, the summations over $|i| \leq n^{\widehat{\omega}}$ are asymptotically independent for different l's. We now prove that the sum over $n^{\widehat{\omega}} < |i| \leq n$ in (6.147) is much smaller n^{-c} for some small constant $c > 0$.

Since $\omega_h \ll \widehat{\omega}$ the rigidity of the eigenvalues in (6.145) holds for $n^{\widehat{\omega}} \leq |i| \leq n^{1-10\omega_h}$, hence we conclude the following bound with very high probability:

$$\frac{1}{n} \sum_{n^{\widehat{\omega}} \leq |i| \leq n} \frac{\eta_l}{(\lambda_i^{z_l})^2 + \eta_l^2} \lesssim n^{40\omega_h} \sum_{n^{\widehat{\omega}} \leq |i| \leq n} \frac{n\eta_l}{i^2 (\rho^{z_l}(0))^2} \lesssim \frac{n^{\delta_1 + 40\omega_h}}{n^{\widehat{\omega}}}, \tag{6.148}$$

where we used that $(\lambda_i^z)^2+\eta^2 \gtrsim n^{-40\omega_h}$ for $n^{1-10\omega_h} \le |i| \le n$, and that $\eta_l \in [n^{-1-\delta_0}, n^{-1+\delta_1}]$. In particular, in (6.148) we used that by (6.146) it follows $\gamma_i^{z_l} \sim i/(n\rho^{z_l}(0))$ for $|i| \le n^{1-10\omega_h}$, where $\rho^{z_l}(0) = \Im m^{z_l}(0) = (1-|z_l|^2)^{1/2}$ for $|z_l|^2 \le 1$ by (6.25).

Combining (6.147)–(6.148) with Proposition 6.7.2 we immediately conclude that

$$\begin{aligned}\mathbf{E}\prod_{l=1}^{p}\langle G^{z_l}(\mathrm{i}\eta_l)\rangle &= \mathbf{E}\prod_{l=1}^{p}\frac{\mathrm{i}}{2n}\sum_{|i|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\lambda_i^{z_l})^2+\eta_l^2}+\mathcal{O}\left(\frac{n^{\delta_1+40\omega_h}}{n^{\widehat{\omega}}}\right)\\ &= \prod_{l=1}^{p}\mathbf{E}\,\frac{\mathrm{i}}{2n}\sum_{|i|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\lambda_i^{z_l})^2+\eta_l^2}+\mathcal{O}\left(\frac{n^{p\delta_0+\widehat{\omega}}}{n^{\omega}}+\frac{n^{\delta_1+40\omega_h}}{n^{\widehat{\omega}}}\right)\\ &= \prod_{l=1}^{p}\mathbf{E}\langle G^{z_l}(\mathrm{i}\eta_l)\rangle+\mathcal{O}\left(\frac{n^{\delta_1+40\omega_h}}{n^{\widehat{\omega}}}+\frac{n^{p\delta_0+\widehat{\omega}}}{n^{\omega}}\right).\end{aligned}$$

This concludes the proof of Proposition 6.3.5 since $\omega_h \ll \delta_m \ll \widehat{\omega} \ll \omega$, with $m=0,1$. □

We conclude Section 6.7.1 with some properties of m^z, the unique solution of (6.19). Fix $z \in \mathbf{C}$, and consider the $2n\times 2n$ matrix $A+F$, with F a Wigner matrix, whose entries are centred random variables of variance $(2n)^{-1}$, and A is a deterministic diagonal matrix $A := \mathrm{diag}(|z|,\dots,|z|,-|z|,\dots,-|z|)$. Then by [57, Eq. (2.1)], [83, Eq. (2.2)] it follows that the corresponding *Dyson equation* is given by

$$\begin{cases}-\frac{1}{m_1} = w-|z|+\frac{m_1+m_2}{2}\\ -\frac{1}{m_2} = w+|z|+\frac{m_1+m_2}{2},\end{cases} \tag{6.149}$$

which has a unique solution under the assumption $\Im m_1, \Im m_2 > 0$. By (6.149) it readily follows that m^z, the solution of (6.19), satisfies

$$m^z(w) = \frac{m_1(w)+m_2(w)}{2}. \tag{6.150}$$

In particular, this implies that all the regularity properties of m_1+m_2 (see e.g. [7, Theorem 2.4, Lemma A.7], [14, Proposition 2.3, Lemma A.1]) hold for m^z as well, e.g. m^z is $1/3$-Hölder continuous for any $z\in\mathbf{C}$.

6.7.2 Overview of the proof of Proposition 6.7.2

The main result of this section is the proof of Proposition 6.7.2, which is divided into two further sub-sections. In Lemma 6.7.5, we prove that we can add a common small Ginibre component to the matrices H^{z_l}, with $l\in[p]$, $p\in\mathbf{N}$, without changing their joint eigenvalue distribution much. In Section 6.7.2.1, we introduce comparison processes for the process defined in (6.156) below, with initial data $\boldsymbol{\lambda}^{z_l} = \{\lambda_{\pm i}^{z_l}\}_{i=1}^n$, where we recall that $\{\lambda_i^{z_l}\}_{i=1}^n$ are the singular values of $\check{X}_{t_f}-z_l$, and $\lambda_{-i}^{z_l} = -\lambda_i^{z_l}$ (the matrix $\check{X}_{t_f}$ is defined in (6.153) below). Finally, in Section 6.7.2.2 we conclude the proof of Proposition 6.7.2. Additionally, in Section 6.7.2.3 we summarize the different scales used in the proof of Proposition 6.7.2.

Let X be an i.i.d. complex $n\times n$ matrix, and run the Ornstein-Uhlenbeck (OU) flow

$$d\widehat{X}_t = -\frac{1}{2}\widehat{X}_t\,\mathrm{d}t+\frac{\mathrm{d}\widehat{B}_t}{\sqrt{n}}, \qquad \widehat{X}_0 = X, \tag{6.151}$$

for a time

$$t_f := \frac{n^{\omega_f}}{n}, \tag{6.152}$$

with some small exponent $\omega_f > 0$ given in Proposition 6.7.2, in order to add a small Gaussian component to X. $\widehat{B}_t$ in (6.151) is a standard matrix valued complex Brownian motion independent of $\widehat{X}_0$, i.e. $\sqrt{2}\Re \widehat{B}_{ab}$, $\sqrt{2}\Im \widehat{B}_{ab}$ are independent standard real Brownian motions for any $a, b \in [n]$. Then we construct an i.i.d. matrix $\check{X}_{t_f}$ such that

$$\widehat{X}_{t_f} \stackrel{d}{=} \check{X}_{t_f} + \sqrt{ct_f}U, \tag{6.153}$$

for some constant $c > 0$ very close to 1, and U is a complex Ginibre matrix independent of $\check{X}_{t_f}$.

Next, we define the matrix flow

$$\mathrm{d}X_t = \frac{\mathrm{d}B_t}{\sqrt{n}}, \quad X_0 = \check{X}_{t_f}, \tag{6.154}$$

where B_t is a standard matrix valued complex Brownian motion independent of X_0 and $\widehat{B}_t$. Note that by construction X_{ct_f} is such that

$$X_{ct_f} \stackrel{d}{=} \widehat{X}_{t_f}. \tag{6.155}$$

Define the matrix $H_t^{z_l}$ as in (6.142) replacing $X - z$ by $X_t - z_l$, for any $l \in [p]$, then the flow in (6.154) induces the following DBM flow on the eigenvalues of $H_t^{z_l}$ (cf. [87, Eq. (5.8)]):

$$\mathrm{d}\lambda_i^{z_l}(t) = \sqrt{\frac{1}{2n}}\,\mathrm{d}b_i^{z_l} + \frac{1}{2n}\sum_{j\neq i}\frac{1}{\lambda_i^{z_l}(t) - \lambda_j^{z_l}(t)}\,\mathrm{d}t, \qquad 1 \le |i| \le n, \tag{6.156}$$

with initial data $\{\lambda_{\pm i}^{z_l}(0)\}_{i=1}^n$, where $\lambda_i^{z_l}(0)$, with $i \in [n]$ and $l \in [p]$, are the singular values of $\check{X}_{t_f} - z_l$, and $\lambda_{-i}^{z_l} = -\lambda_i^{z_l}$. The well-posedness of (6.156) follows by [54, Appendix A]. It follows from this derivation that the Brownian motions $\{b_i^{z_l}\}_{i=1}^n$, omitting the t-dependence, are defined as

$$\mathrm{d}b_i^{z_l} := \sqrt{2}\left(\mathrm{d}B_{ii}^{z_l} + \mathrm{d}\overline{B_{ii}^{z_l}}\right), \qquad \mathrm{d}B_{ij}^{z_l} := \sum_{a,b=1}^n \overline{u_i^{z_l}(a)}\,\mathrm{d}B_{ab} v_j^{z_l}(b), \tag{6.157}$$

where $(\boldsymbol{u}_i^{z_l}, \pm\boldsymbol{v}_i^{z_l})$ are the orthonormal eigenvectors of $H_t^{z_l}$ with corresponding eigenvalues $\lambda_{\pm i}^{z_l}$, and B_{ab} are the entries of the Brownian motion defined in (6.154). For negative indices we define $b_{-i}^{z_l} := -b_i^{z_l}$. It follows from (6.157) that for each fixed l the collection of Brownian motions $\boldsymbol{b}^{z_l} = \{b_i^{z_l}\}_{i=1}^n$ consists of i.i.d. Brownian motions, however the families $\boldsymbol{b}^{z_l}$ are not independent for different l's, in fact their joint distribution is not necessarily Gaussian. The derivation of (6.156) follows standard steps, see e.g. [90, Section 12.2]. For the convenience of the reader we included this derivation in Appendix 6.B.

Remark 6.7.4. *We point out that in the formula [54, Eq. (3.9)] analogous to* (6.156) *the term* $j = -i$ *in* (6.156) *is apparently missing. This additional term does not influence the results in [54, Section 3] (that are proven for the real DBM for which the term* $j = -i$ *is actually not present).*

As a consequence of (6.155) we conclude the following lemma.

Lemma 6.7.5. *Let* $\boldsymbol{\lambda}^{z_l} = \{\lambda^{z_l}_{\pm i}\}_{i=1}^n$ *be the eigenvalues of* H^{z_l} *and let* $\boldsymbol{\lambda}^{z_l}(t)$ *be the solution of* (6.156) *with initial data* $\boldsymbol{\lambda}^{z_l}$*, then*

$$\begin{aligned}\mathbf{E}\prod_{l=1}^p \frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}} \frac{\eta_l}{(\lambda_{i_l}^{z_l})^2+\eta_l^2} &= \mathbf{E}\prod_{l=1}^p \frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}} \frac{\eta_l}{(\lambda_{i_l}^{z_l}(ct_f))^2+\eta_l^2} \\ &\quad + \mathcal{O}\left(\frac{n^{p\xi+2\delta_0}t_f}{n^{1/2}}\sum_{l=1}^p\frac{1}{\eta_l}+\frac{n^{k\delta_0+\delta_1}}{n^{\widehat{\omega}}}\right),\end{aligned} \tag{6.158}$$

for any sufficiently small $\widehat{\omega}, \delta_0, \delta_1 > 0$ *such that* $\delta_m \ll \widehat{\omega}$*, where* $\eta_l \in [n^{-1-\delta_0}, n^{-1+\delta_1}]$ *and* t_f *defined in* (6.152).

Proof. The equality in (6.158) follows by a standard Green's function comparison (GFT) argument (e.g. see [59, Proposition 3.1]) for the $\langle G^{z_l}(\mathrm{i}\eta_l)\rangle$, combined with the same argument as in the proof of Proposition 6.3.5, using the local law [11, Theorem 5.1] and (6.155), to show that the summation over $n^{\widehat{\omega}} < |i| \le n$ is negligible. We remark that the GFT used in this lemma is much easier than the one in [59, Proposition 3.1] since here we used GFT only for a very short time $t_f \sim n^{-1+\omega_f}$, for a very small $\omega_f > 0$, whilst in [59, Proposition 3.1] the GFT is considered up to a time $t = +\infty$. The scaling in the error term in [59, Proposition 3.1] is different compared to the error term in (6.158) since the scaling therein refers to the cusp-scaling. □

6.7.2.1 Definition of the comparison processes for $\boldsymbol{\lambda}^{z_l}(t)$

The philosophy behind the proof of Proposition 6.7.2 is to compare the distribution of $\boldsymbol{\lambda}^{z_l}(t) = \{\lambda^{z_l}_{\pm i}(t)\}$, the strong solutions of (6.156) for $l \in [p]$, which are correlated for different l's and realized on a probability space Ω_b, with carefully constructed independent processes $\boldsymbol{\mu}^{(l)}(t) = \{\mu^{(l)}_{\pm i}(t)\}_{i=1}^n$ on a different probability space Ω_β. We choose $\boldsymbol{\mu}^{(l)}(t)$ to be the solution of

$$\mathrm{d}\mu_i^{(l)}(t) = \frac{\mathrm{d}\beta_i^{(l)}}{\sqrt{2n}} + \frac{1}{2n}\sum_{j\ne i}\frac{1}{\mu_i^{(l)}(t)-\mu_j^{(l)}(t)}\,\mathrm{d}t, \quad \mu_i^{(l)}(0) = \mu_i^{(l)}, \tag{6.159}$$

for $|i| \le n$, with $\mu_i^{(l)}$ the eigenvalues of the matrix

$$H^{(l)} := \begin{pmatrix} 0 & X^{(l)} \\ (X^{(l)})^* & 0 \end{pmatrix}$$

where $X^{(l)}$ are independent Ginibre matrices, $\boldsymbol{\beta}^{(l)} = \{\beta_i^{(l)}\}_{i=1}^n$ are independent vectors of i.i.d. standard real Brownian motions, and $\beta_{-i}^{(l)} = -\beta_i^{(l)}$. We let $\mathcal{F}_{\beta,t}$ denote the common filtration of the Brownian motions $\boldsymbol{\beta}^{(l)}$ on Ω_β.

In the remainder of this section we define two processes $\widetilde{\boldsymbol{\lambda}}^{(l)}$, $\widetilde{\boldsymbol{\mu}}^{(l)}$ so that for a time $t \ge 0$ large enough $\widetilde{\lambda}_i^{(l)}(t)$, $\widetilde{\mu}_i^{(l)}(t)$ for small indices i will be close to $\lambda_i^{z_l}(t)$ and $\mu_i^{(l)}(t)$, respectively, with very high probability. Additionally, the processes $\widetilde{\boldsymbol{\lambda}}^{(l)}$, $\widetilde{\boldsymbol{\mu}}^{(l)}$ will be such that they have the same joint distribution:

$$\left(\widetilde{\boldsymbol{\lambda}}^{(1)}(t),\dots,\widetilde{\boldsymbol{\lambda}}^{(p)}(t)\right)_{t\ge 0} \stackrel{d}{=} \left(\widetilde{\boldsymbol{\mu}}^{(1)}(t),\dots,\widetilde{\boldsymbol{\mu}}^{(p)}(t)\right)_{t\ge 0}. \tag{6.160}$$

Fix $\omega_A > 0$ and define the process $\widetilde{\boldsymbol{\lambda}}(t)$ to be the solution of

$$\mathrm{d}\widetilde{\lambda}_i^{(l)}(t) = \frac{1}{2n}\sum_{j\neq i}\frac{1}{\widetilde{\lambda}_i^{(l)}(t)-\widetilde{\lambda}_j^{(l)}(t)}\,\mathrm{d}t + \begin{cases}\sqrt{\frac{1}{2n}}\,\mathrm{d}b_i^{z_l} & \text{if} \quad |i|\leq n^{\omega_A} \\ \sqrt{\frac{1}{2n}}\,\mathrm{d}\widetilde{b}_i^{(l)} & \text{if} \quad n^{\omega_A}<|i|\leq n,\end{cases} \tag{6.161}$$

with initial data $\widetilde{\boldsymbol{\lambda}}^{(l)}(0)$ being the singular values, taken with positive and negative sign, of independent Ginibre matrices $\widetilde{Y}^{(l)}$ independent of $\boldsymbol{\lambda}^{z_l}(0)$. Here $\mathrm{d}b_i^{z_l}$ is from (6.156); this is used for small indices. For large indices we define the driving Brownian motions to be an independent collection $\{\{\widetilde{b}_i^{(l)}\}_{i=n^{\omega_A}+1}^n \mid l\in[p]\}$ of p vector-valued i.i.d. standard real Brownian motions which are also independent of $\{\{b_{\pm i}^{z_l}\}_{i=1}^n \,|\, l\in[p]\}$, and that $\widetilde{b}_{-i}^{(l)}=-\widetilde{b}_i^{(l)}$. The Brownian motions $\boldsymbol{b}^{z_l}$, with $l\in[p]$, and $\{\{\widetilde{b}_i^{(l)}\}_{i=n^{\omega_A}+1}^n \mid l\in[p]\}$ are defined on a common probability space that we continue to denote by Ω_b with the common filtration $\mathcal{F}_{b,t}$.

We conclude this section by defining $\widetilde{\boldsymbol{\mu}}^{(l)}(t)$, the comparison process of $\boldsymbol{\mu}^{(l)}(t)$. It is given as the solution of the following DBM:

$$\mathrm{d}\widetilde{\mu}_i^{(l)}(t) = \frac{1}{2n}\sum_{j\neq i}\frac{1}{\widetilde{\mu}_i^{(l)}(t)-\widetilde{\mu}_j^{(l)}(t)}\,\mathrm{d}t + \begin{cases}\sqrt{\frac{1}{2n}}\,\mathrm{d}\zeta_i^{z_l} & \text{if} \quad |i|\leq n^{\omega_A} \\ \sqrt{\frac{1}{2n}}\,\mathrm{d}\widetilde{\zeta}_i^{(l)} & \text{if} \quad n^{\omega_A}<|i|\leq n,\end{cases} \tag{6.162}$$

with initial data $\widetilde{\boldsymbol{\mu}}^{(l)}(0)$ so that they are the singular values of independent Ginibre matrices $Y^{(l)}$, which are also independent of $\widetilde{Y}^{(l)}$. We now explain how to construct the driving Brownian motions in (6.162) so that (6.160) is satisfied. We only consider positive indices, since the negative indices are defined by symmetry. For indices $n^{\omega_A}<i\leq n$ we choose $\{\widetilde{\zeta}_{\pm i}^{(l)}\}_{n^{\omega_A}+1}^n$ to be independent families (for different l's) of i.i.d. Brownian motions, defined on the same probability space of $\{\boldsymbol{\beta}^{(l)} : l\in[p]\}$, that are independent of the Brownian motions $\{\beta_{\pm i}^{(l)}\}_{i=1}^n$ used in (6.159). For indices $1\leq i\leq n^{\omega_A}$ the families $\{\{\zeta_i^{z_l}\}_{i=1}^{n^{\omega_A}} \mid l\in[p]\}$ will be constructed from the independent families $\{\{\beta_i^{(l)}\}_{i=1}^{n^{\omega_A}} \mid l\in[p]\}$ as follows.

Arranging $\{\{\beta_i^{(l)}\}_{i=1}^{n^{\omega_A}} \mid l\in[p]\}$ into a single vector, we define the pn^{ω_A}-dimensional vector

$$\underline{\beta} := (\beta_1^{(1)},\dots,\beta_{n^{\omega_A}}^{(1)},\dots,\beta_1^{(p)},\dots,\beta_{n^{\omega_A}}^{(p)}). \tag{6.163}$$

Similarly we define the pn^{ω_A}-dimensional vector

$$\underline{b} := (b_1^{z_1},\dots,b_{n^{\omega_A}}^{z_1},\dots,b_1^{z_p},\dots,b_{n^{\omega_A}}^{z_p}) \tag{6.164}$$

which is a continuous martingale. To make our notation easier, in the following we assume that $n^{\omega_A}\in\mathbf{N}$. For any $i,j\in[pn^{\omega_A}]$, we use the notation

$$i=(l-1)n^{\omega_A}+\mathfrak{i}, \qquad j=(m-1)n^{\omega_A}+\mathfrak{j}, \tag{6.165}$$

with $l,m\in[p]$ and $\mathfrak{i},\mathfrak{j}\in[n^{\omega_A}]$. Note that in the definitions in (6.165) we used $(l-1)$, $(m-1)$ instead of l,m so that l and m exactly indicate in which block of the matrix $C(t)$ in (6.166) the indices i,j are. With this notation, the covariance matrix of the increments of $\underline{b}$ is the matrix $C(t)$ consisting of p^2 blocks of size n^{ω_A} is defined as

$$C_{ij}(t)\,\mathrm{d}t := \mathbf{E}\big[\mathrm{d}b_{\mathfrak{i}}^{z_l}\,\mathrm{d}b_{\mathfrak{j}}^{z_m} \mid \mathcal{F}_{b,t}\big] = \begin{cases}\Theta_{\mathfrak{i}\mathfrak{j}}^{z_l,z_m}(t)\,\mathrm{d}t & \text{if} \quad l\neq m, \\ \delta_{\mathfrak{i}\mathfrak{j}}\,\mathrm{d}t & \text{if} \quad l=m.\end{cases} \tag{6.166}$$

Here

$$\Theta_{\mathrm{ij}}^{z_l,z_m}(t) := 4\Re\big[\langle \boldsymbol{u}_{\mathrm{i}}^{z_l}(t), \boldsymbol{u}_{\mathrm{j}}^{z_m}(t)\rangle\langle \boldsymbol{v}_{\mathrm{i}}^{z_m}(t), \boldsymbol{v}_{\mathrm{j}}^{z_l}(t)\rangle\big], \tag{6.167}$$

with $\{\boldsymbol{w}_{\pm i}\}_{i\in[n]} = \{(\boldsymbol{u}_i^{z_l}(t), \pm\boldsymbol{v}_i^{z_l}(t))\}_{i\in[n]}$ the orthonormal eigenvectors of $H_t^{z_l}$. Note that $\{\boldsymbol{w}_i\}_{|i|\leq n}$ are not well-defined if $H_t^{z_l}$ has multiple eigenvalues. However, without loss of generality, we can assume that almost surely $H_t^{z_l}$ does not have multiple eigenvalues for any $l \in [p]$, as a consequence of [55, Lemma 6.2] (which is the adaptation of [53, Proposition 2.3] to the 2×2 block structure of $H_t^{z_l}$).

By Doob's martingale representation theorem [120, Theorem 18.12] there exists a standard Brownian motion $\boldsymbol{\theta}_t \in \mathbf{R}^{pN^{\omega_A}}$ realized on an extension $(\widetilde{\Omega}_b, \widetilde{\mathcal{F}}_{b,t})$ of the original filtrated probability space $(\Omega_b, \mathcal{F}_{b,t})$ such that $\mathrm{d}\underline{\boldsymbol{b}} = \sqrt{C}\,\mathrm{d}\boldsymbol{\theta}$. Here $\boldsymbol{\theta}_t$ and $C(t)$ are adapted to the filtration $\widetilde{\mathcal{F}}_{b,t}$ and note that $C = C(t)$ is a positive semi-definite matrix and $\sqrt{C}$ denotes its positive semi-definite matrix square root.

For the clarity of the presentation the original processes $\boldsymbol{\lambda}^{z_l}$ and the comparison processes $\boldsymbol{\mu}^{(l)}$ will be realized on completely different probability spaces. We thus construct another copy $(\Omega_\beta, \mathcal{F}_{\beta,t})$ of the filtrated probability space $(\widetilde{\Omega}_b, \widetilde{\mathcal{F}}_{b,t})$ and we construct a matrix valued process $C^{\#}(t)$ and a Brownian motion $\underline{\beta}$ on $(\Omega_\beta, \mathcal{F}_{\beta,t})$ such that $(C^{\#}(t), \underline{\beta}(t))$ are adapted to the filtration $\mathcal{F}_{\beta,t}$ and they have the same joint distribution as $(C(t), \boldsymbol{\theta}(t))$. The Brownian motion $\underline{\beta}$ is used in (6.159) for small indices.

Define the process

$$\underline{\zeta}(t) := \int_0^t \sqrt{C^{\#}(s)}\,\mathrm{d}\underline{\beta}(s), \quad \underline{\zeta} = (\zeta_1^{z_1}, \dots, \zeta_{n^{\omega_A}}^{z_1}, \dots, \zeta_1^{z_p}, \dots, \zeta_{n^{\omega_A}}^{z_p}), \tag{6.168}$$

on the probability space Ω_β and define $\zeta_{-i}^{z_l} := -\zeta_i^{z_l}$ for any $1 \leq i \leq n^{\omega_A}$, $l \in [p]$. Since $\underline{\beta}$ are i.i.d. Brownian motions, we clearly have

$$\mathbf{E}\big[\mathrm{d}\zeta_{\mathrm{i}}^{z_l}(t)\,\mathrm{d}\zeta_{\mathrm{j}}^{z_m}(t) \mid \mathcal{F}_{\beta,t}\big] = C^{\#}(t)_{ij}\,\mathrm{d}t, \qquad |\mathrm{i}|, |\mathrm{j}| \leq n^{\omega_A}. \tag{6.169}$$

By construction we see that the processes $(\{b_{\pm i}^{z_l}\}_{i=1}^{n^{\omega_A}})_{l=1}^k$ and $(\{\zeta_{\pm i}^{z_l}\}_{i=1}^{n^{\omega_A}})_{l=1}^k$ have the same distribution. Furthermore, since by definition the two collections

$$\Big\{\{\widetilde{b}_{\pm i}^{(l)}\}_{i=n^{\omega_A}+1}^n, \{\widetilde{\zeta}_{\pm i}^{(l)}\}_{i=n^{\omega_A}+1}^n \,\Big|\, l \in [k]\Big\}$$

are independent of

$$\Big\{\{b_{\pm i}^{z_l}\}_{i=1}^{n^{\omega_A}}, \{\beta_{\pm i}^{(l)}\}_{i=1}^{n^{\omega_A}} \,\Big|\, l \in [k]\Big\}$$

and among each other, we have

$$\Big(\{b_{\pm i}^{z_l}\}_{i=1}^{n^{\omega_A}}, \{\widetilde{b}_{\pm i}^{(l)}\}_{i=n^{\omega_A}+1}^n\Big)_{l=1}^p \overset{d}{=} \Big(\{\zeta_{\pm i}^{z_l}\}_{i=1}^{n^{\omega_A}}, \{\widetilde{\zeta}_{\pm i}^{(l)}\}_{i=n^{\omega_A}+1}^n\Big)_{l=1}^p. \tag{6.170}$$

Finally, by the definitions in (6.161), (6.162), and (6.170), it follows that the Dyson Brownian motions $\widetilde{\boldsymbol{\lambda}}^{(l)}$ and $\widetilde{\boldsymbol{\mu}}^{(l)}$ have the same distribution, i.e.

$$\Big(\widetilde{\boldsymbol{\lambda}}^{(1)}(t), \dots, \widetilde{\boldsymbol{\lambda}}^{(p)}(t)\Big) \overset{d}{=} \Big(\widetilde{\boldsymbol{\mu}}^{(1)}(t), \dots, \widetilde{\boldsymbol{\mu}}^{(p)}(t)\Big) \tag{6.171}$$

since their initial conditions, as well as their driving processes (6.170), agree in distribution. Note that these processes are Brownian motions for each fixed l since $C_{ij}(t) = \delta_{\mathrm{ij}}$ if $l = m$, but jointly they are not necessarily Gaussian due to the non-trivial correlation $\Theta_{\mathrm{ij}}^{z_l,z_m}$ in (6.166).

6.7.2.2 Proof of Proposition 6.7.2

In this section we conclude the proof of Proposition 6.7.2 using the comparison processes defined in Section 6.7.2.1. More precisely, we use that the processes $\boldsymbol{\lambda}^{z_l}(t)$, $\widetilde{\boldsymbol{\lambda}}^{(l)}(t)$ and $\boldsymbol{\mu}^{(l)}(t)$, $\widetilde{\boldsymbol{\mu}}^{(l)}(t)$ are close pathwise at time t_f, as stated below in Lemma 6.7.6 and Lemma 6.7.7, respectively. The proofs of these lemmas are postponed to Section 6.7.5. They will be a consequence of Proposition 6.7.13, which is an adaptation to our case of the main technical estimate of [129]. The main input is the bound on the eigenvector overlap in Lemma 6.7.9, since it gives an upper bound on the correlation structure in (6.169). Let $\rho_{sc}(E) = \frac{1}{2\pi}\sqrt{4-E^2}$ denote the semicircle density.

Lemma 6.7.6. *Fix $p \in \mathbf{N}$, and let $\boldsymbol{\lambda}^{z_l}(t)$, $\widetilde{\boldsymbol{\lambda}}^{(l)}(t)$, with $l \in [p]$, be the processes defined in* (6.156) *and* (6.161)*, respectively. For any small $\omega_h, \omega_f > 0$ such that $\omega_h \ll \omega_f$ there exist $\omega, \widehat{\omega} > 0$ with $\omega_h \ll \widehat{\omega} \ll \omega \ll \omega_f$, such that for any $|z_l| \le 1 - n^{-\omega_h}$ it holds*

$$\left|\rho^{z_l}(0)\lambda_i^{z_l}(ct_f) - \rho_{sc}(0)\widetilde{\lambda}_i^{(l)}(ct_f)\right| \le n^{-1-\omega}, \qquad |i| \le n^{\widehat{\omega}}, \tag{6.172}$$

with very high probability, where $t_f := n^{-1+\omega_f}$ and $c > 0$ is defined in (6.155)*.*

Lemma 6.7.7. *Fix $p \in \mathbf{N}$, and let $\boldsymbol{\mu}^{(l)}(t)$, $\widetilde{\boldsymbol{\mu}}^{(l)}(t)$, with $l \in [p]$, be the processes defined in* (6.159) *and* (6.162)*, respectively. For any small $\omega_h, \omega_f, \omega_d > 0$ such that $\omega_h \ll \omega_f$ there exist $\omega, \widehat{\omega} > 0$ with $\omega_h \ll \widehat{\omega} \ll \omega \ll \omega_f$, such that for any $|z_l| \le 1 - n^{-\omega_h}$, $|z_l - z_m| \ge n^{-\omega_d}$, with $l \ne m$, it holds*

$$\left|\mu_i^{(l)}(ct_f) - \widetilde{\mu}_i^{(l)}(ct_f)\right| \le n^{-1-\omega}, \qquad |i| \le n^{\widehat{\omega}}, \tag{6.173}$$

with very high probability, where $t_f := n^{-1+\omega_f}$ and $c > 0$ is defined in (6.155)*.*

Proof of Proposition 6.7.2. In the following we omit the trivial scaling factors $\rho^{z_l}(0)$, $\rho_{sc}(0)$ in the second term in the lhs. of (6.172) to make our notation easier. We recall that by Lemma 6.7.5 we have

$$\begin{aligned}\mathbf{E}\prod_{l=1}^p \frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}} \frac{\eta_l}{(\lambda_{i_l}^{z_l})^2+\eta_l^2} &= \mathbf{E}\prod_{l=1}^p \frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}} \frac{\eta_l}{(\lambda_{i_l}^{z_l}(ct_f))^2+\eta_l^2} \\ &\quad + \mathcal{O}\left(\frac{n^{p\xi+2\delta_0}t_f}{n^{1/2}}\sum_{l=1}^p\frac{1}{\eta_l} + \frac{n^{p\delta_0+\delta_1}}{n^{\widehat{\omega}}}\right),\end{aligned} \tag{6.174}$$

where $\lambda_i^{z_l}(t)$ is the solution of (6.156) with initial data $\lambda_i^{z_l}$. Next we replace $\lambda_i^{z_l}(t)$ with $\widetilde{\lambda}_i^{z_l}(t)$ for small indices by using Lemma 6.7.6; this is formulated in the following lemma whose detailed proof is postponed to the end of this section.

Lemma 6.7.8. *Fix $p \in \mathbf{N}$, and let $\lambda_i^{z_l}(t)$, $\widetilde{\lambda}_i^{(l)}(t)$, with $l \in [p]$, be the solution of* (6.156) *and* (6.161)*, respectively. Then*

$$\mathbf{E}\prod_{l=1}^p \frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}} \frac{\eta_l}{(\lambda_{i_l}^{z_l})^2+\eta_l^2} = \mathbf{E}\prod_{l=1}^p \frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}} \frac{\eta_l}{(\widetilde{\lambda}_{i_l}^{(l)}(ct_f))^2+\eta_l^2} + \mathcal{O}(\Psi), \tag{6.175}$$

where $\lambda_{i_l}^{z_l} = \lambda_{i_l}^{z_l}(0)$, $t_f = n^{-1+\omega_f}$, and the error term is given by

$$\Psi := \frac{n^{\widehat{\omega}}}{n^{1+\omega}}\left(\sum_{l=1}^p \frac{1}{\eta_l}\right)\cdot\prod_{l=1}^p\left(1+\frac{n^\xi}{n\eta_l}\right) + \frac{n^{p\xi+2\delta_0}t_f}{n^{1/2}}\sum_{l=1}^p\frac{1}{\eta_l} + \frac{n^{p\delta_0+\delta_1}}{n^{\widehat{\omega}}}.$$

By (6.171) it readily follows that

$$\mathbf{E}\prod_{l=1}^{p}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\widetilde{\lambda}_{i_l}^{(l)}(ct_f))^2+\eta_l^2}=\mathbf{E}\prod_{l=1}^{p}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\widetilde{\mu}_{i_l}^{(l)}(ct_f))^2+\eta_l^2}. \tag{6.176}$$

Moreover, by (6.173), similarly to Lemma 6.7.8, we conclude

$$\mathbf{E}\prod_{l=1}^{p}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\widetilde{\mu}_{i_l}^{(l)}(ct_f))^2+\eta_l^2}=\mathbf{E}\prod_{l=1}^{p}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\mu_{i_l}^{(l)}(ct_f))^2+\eta_l^2}+\mathcal{O}(\Psi). \tag{6.177}$$

Additionally, by the definition of the processes $\boldsymbol{\mu}^{(l)}(t)$ in (6.159) it follows that $\boldsymbol{\mu}^{(l)}(t)$, $\boldsymbol{\mu}^{(m)}(t)$ are independent for $l\neq m$ and so that

$$\mathbf{E}\prod_{l=1}^{p}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\mu_{i_l}^{(l)}(ct_f))^2+\eta_l^2}=\prod_{l=1}^{p}\mathbf{E}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\mu_{i_l}^{(l)}(ct_f))^2+\eta_l^2}. \tag{6.178}$$

Combining (6.175)–(6.178), we get

$$\mathbf{E}\prod_{l=1}^{p}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\lambda_{i_l}^{z_l})^2+\eta_l^2}=\prod_{l=1}^{p}\mathbf{E}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\mu_{i_l}^{(l)}(ct_f))^2+\eta_l^2}+\mathcal{O}(\Psi). \tag{6.179}$$

Then, by similar computation to the ones in (6.174)–(6.179) we conclude that

$$\prod_{l=1}^{p}\mathbf{E}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\lambda_{i_l}^{z_l})^2+\eta_l^2}=\prod_{l=1}^{p}\mathbf{E}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\mu_{i_l}^{(l)}(ct_f))^2+\eta_l^2}+\mathcal{O}(\Psi). \tag{6.180}$$

We remark that in order to prove (6.180) it would not be necessary to introduce the additional comparison processes $\widetilde{\boldsymbol{\lambda}}^{(l)}$ and $\widetilde{\boldsymbol{\mu}}^{(l)}$ of Section 6.7.2.1, since in (6.180) the product is outside the expectation, so one can compare the expectations one by one; the correlation between these processes for different l's plays no role. Hence, already the usual coupling (see e.g. [42, 54, 129]) between the processes $\boldsymbol{\lambda}^{z_l}(t)$, $\boldsymbol{\mu}^{(l)}(t)$ defined in (6.156) and (6.159), respectively, would be sufficient to prove (6.180).

Finally, combining (6.179)–(6.180) we conclude the proof of Proposition 6.7.2. □

Proof of Lemma 6.7.8. We show the proof for $p=2$ in order to make our presentation easier. The case $p\ge 3$ proceeds exactly in the same way. In order to make our notation shorter, for $l\in\{1,2\}$, we define

$$T_{i_l}^{(l)}:=\frac{\eta_l}{(\lambda_{i_l}^{z_l}(ct_f))^2+\eta_l^2}.$$

Similarly, replacing $\lambda_{i_l}^{z_l}(ct_f)$ with $\widetilde{\lambda}_{i_l}^{(l)}(ct_f)$, we define $\widetilde{T}_l$. Then, by telescopic sum, we have

$$\begin{aligned}
&\left| \mathbf{E} \prod_{l=1}^{2} \frac{1}{n} \sum_{|i_l| \le n^{\widehat{\omega}}} T_{i_l}^{(l)} - \mathbf{E} \prod_{l=1}^{2} \frac{1}{n} \sum_{|i_l| \le n^{\widehat{\omega}}} \widetilde{T}_{i_l}^{(l)} \right| \\
&= \frac{1}{n^2} \left| \mathbf{E} \sum_{|i_1|,|i_2| \le n^{\widehat{\omega}}} \left[T_{i_1}^{(1)} - \widetilde{T}_{i_1}^{(1)} \right] T_{i_2}^{(2)} - \mathbf{E} \sum_{|i_1|,|i_2| \le n^{\widehat{\omega}}} \left[T_{i_2}^{(2)} - \widetilde{T}_{i_2}^{(2)} \right] \widetilde{T}_{i_1}^{(1)} \right| \\
&\lesssim \sum_{\substack{l,m=1 \\ l \neq m}}^{2} \left(1 + \frac{n^\xi}{n\eta_l} \right) \mathbf{E} \frac{1}{n} \sum_{|i_m| \le n^{\widehat{\omega}}} \frac{T_{i_m}^{(m)} \widetilde{T}_{i_m}^{(m)}}{\eta_m} \left| (\widetilde{\lambda}_{i_m}^{(m)}(ct_f))^2 - (\lambda_{i_m}^{z_m}(ct_f))^2 \right| \\
&\lesssim \frac{n^{\widehat{\omega}}}{n^{1+\omega}} \left(\frac{1}{\eta_1} + \frac{1}{\eta_2} \right) \cdot \prod_{l=1}^{2} \left(1 + \frac{n^\xi}{n\eta_l} \right),
\end{aligned} \tag{6.181}$$

where we used the local law (6.3.1) in the first inequality and (6.172) in the last step. Combining (6.181) with (6.174) we conclude the proof of Lemma 6.7.8. □

Before we continue, we summarize the scales used in the entire Section 6.7.

6.7.2.3 Relations among the scales in the proof of Proposition 6.7.2

Scales in the proof of Proposition 6.7.2 are characterized by various exponents ω's of n that we will also refer to scales, for simplicity. The basic input scales in the proof of Proposition 6.7.2 are $0 < \omega_d, \omega_h, \omega_f \ll 1$, the others will depend on them. The exponents ω_h, ω_d are chosen within the assumptions of Lemma 6.7.9 to control the location of z's as $|z_l| \le 1 - n^{-\omega_h}$, $|z_l - z_m| \ge n^{-\omega_p}$, with $l \neq m$. The exponent ω_f defines the time $t_f = n^{-1+\omega_f}$ so that the local equilibrium of the DBM is reached after t_f. This will provide the asymptotic independence of $\lambda_i^{z_l}$, $\lambda_j^{z_m}$ for small indices and for $l \neq m$.

The primary scales created along the proof of Proposition 6.7.2 are ω, $\widehat{\omega}$, δ_0, δ_1, ω_E, ω_B. The scales ω_E, ω_B are given in Lemma 6.7.9: $n^{-\omega_E}$ measures the size of the eigenvector overlaps from (6.167) while the exponent ω_B describes the range of indices for which these overlap estimates hold. Recall that the overlaps determine the correlations among the driving Brownian motions. The scale ω quantifies the $n^{-1-\omega}$ precision of the coupling between various processes. These couplings are effective only for small indices i, their range is given by $\widehat{\omega}$ as $|i| \le n^{\widehat{\omega}}$. Both these scales are much bigger than ω_h but much smaller than ω_f. They are determined in Lemma 6.7.6, Lemma 6.7.7, in fact both lemmas give only a necessary upper bound on the scales $\omega, \widehat{\omega}$, so we can pick the smaller of them. The exponents δ_0, δ_1 determine the range of $\eta \in [n^{-1-\delta_0}, n^{-1+\delta_1}]$ for which Proposition 6.7.2 holds; these are determined in Lemma 6.7.5 after $\omega, \widehat{\omega}$ have already been fixed. These steps yield the scales $\omega, \widehat{\omega}, \delta_0, \delta_1$ claimed in Proposition 6.7.2 and hence also in Proposition 6.3.5. We summarize order relation among all these scales as

$$\omega_h \ll \delta_m \ll \widehat{\omega} \ll \omega \ll \omega_B \ll \omega_f \ll \omega_E \ll 1, \qquad m = 0, 1. \tag{6.182}$$

We mention that three further auxiliary scales emerge along the proof but they play only a local, secondary role. For completeness we also list them here; they are $\omega_1, \omega_A, \omega_l$. Their meanings are the following: $t_1 := n^{-1+\omega_1}$, with $\omega_1 \ll \omega_f$, is the time needed for the DBM process $x_i(t,\alpha)$, defined in (6.196), to reach local equilibrium, hence to prove its universality; $t_0 := t_f - t_1$ is the initial time we run the DBM before starting with the actual proof of universality so that the solution $\boldsymbol{\lambda}^{z_l}(t_0)$ of (6.156) at time t_0 and the density $\mathrm{d}\rho(E,t,\alpha)$ (which we will define in Section 6.7.6.2) satisfy certain technical regularity conditions [54, Lemma 3.3-3.5], [129, Lemma 3.3-3.5]. Note that $t_0 \sim t_f$, in fact they are almost the same. The other two scales are technical: ω_l is the scale of the short range interaction, and ω_A is a cut-off scale such that $x_i(t,\alpha)$ is basically independent of α for $|i| \le n^{\omega_A}$. These scales are inserted in the above chain of inequalities (6.182) between ω, ω_B as follows

$$\omega_h \ll \delta_m \ll \widehat{\omega} \ll \omega \ll \omega_1 \ll \omega_l \ll \omega_A \le \omega_B \ll \omega_f \ll \omega_E \ll 1, \quad m = 0,1.$$

In particular, the relation $\omega_A \ll \omega_E$ ensures that the effect of the correlation is small, see the bound in (6.195) later.

We remark that introducing the additional initial time layer t_0 is not really necessary for our proof of Proposition 6.7.2 since the initial data $\boldsymbol{\lambda}^z(0)$ of the DBM in (6.156) and their deterministic density ρ^z already satisfy [54, Lemma 3.3-3.5], [129, Lemma 3.3-3.5] as a consequence of (6.144) (see Remark 6.7.10 and Remark 6.7.15 for more details). We keep it only to facilitate the comparison with [54, 129].

6.7.3 Bound on the eigenvector overlap for large $|z_1 - z_2|$

For any $z \in \mathbf{C}$, let $\{\boldsymbol{w}^z_{\pm i}\}_{i=1}^n$ be the eigenvectors of the matrix H^z. They are of the form $\boldsymbol{w}^z_{\pm i} = (\boldsymbol{u}^z_i, \pm \boldsymbol{v}^z_i)$, with $\boldsymbol{u}^z_i, \boldsymbol{v}^z_i \in \mathbf{C}^n$, as a consequence of the symmetry of the spectrum of H^z induced by its block structure. The main input to prove Lemma 6.7.6–6.7.7 is the following high probability bound on the almost orthogonality of the eigenvectors belonging to distant z_l, z_m parameters and eigenvalues close to zero. With the help of the Dyson Brownian motion (DBM), this information will then be used to establish almost independence of these eigenvalues.

Lemma 6.7.9. *Let* $\{\boldsymbol{w}^{z_l}_{\pm i}\}_{i=1}^n = \{(\boldsymbol{u}^{z_l}_i, \pm \boldsymbol{v}^{z_l}_i)\}_{i=1}^n$*, for* $l = 1,2$*, be the eigenvectors of matrices* H^{z_l} *of the form* (6.142) *with i.i.d. entries. Then for any sufficiently small* $\omega_d, \omega_h > 0$ *there exist* $\omega_B, \omega_E > 0$ *such that if* $|z_1 - z_2| \ge n^{-\omega_d}$, $|z_l| \le 1 - n^{-\omega_h}$ *then*

$$\left|\langle \boldsymbol{u}^{z_1}_i, \boldsymbol{u}^{z_2}_j\rangle\right| + \left|\langle \boldsymbol{v}^{z_1}_i, \boldsymbol{v}^{z_2}_j\rangle\right| \le n^{-\omega_E}, \quad 1 \le i,j \le n^{\omega_B}, \tag{6.183}$$

with very high probability.

Proof. Using the spectral symmetry of H^z, for any $z \in \mathbf{C}$ we write G^z in spectral decomposition as

$$G^z(\mathrm{i}\eta) = \sum_{j>0} \frac{2}{(\lambda^z_j)^2 + \eta^2} \begin{pmatrix} \mathrm{i}\eta \boldsymbol{u}^z_j(\boldsymbol{u}^z_j)^* & \lambda^z_j \boldsymbol{u}^z_j(\boldsymbol{v}^z_j)^* \\ \lambda^z_j \boldsymbol{v}^z_j(\boldsymbol{u}^z_j)^* & \mathrm{i}\eta \boldsymbol{v}^z_j(\boldsymbol{v}^z_j)^* \end{pmatrix}.$$

Let $\eta \ge n^{-1}$, then by rigidity of the eigenvalues in (6.145), for any $i_0, j_0 \ge 1$ such that $\lambda^{z_l}_{i_0}, \lambda^{z_l}_{j_0} \lesssim \eta$, with $l = 1,2$, and any z_1, z_2 such that $n^{-\omega_d} \lesssim |z_1 - z_2| \lesssim 1$, for some

$\omega_d > 0$ we will choose shortly, it follows that

$$\begin{aligned}
&\left|\langle \boldsymbol{u}_{i_0}^{z_1}, \boldsymbol{u}_{j_0}^{z_2}\rangle\right|^2 + \left|\langle \boldsymbol{v}_{i_0}^{z_1}, \boldsymbol{v}_{j_0}^{z_2}\rangle\right|^2 \\
&\quad \lesssim \sum_{i,j=1}^{n} \frac{4\eta^4}{((\lambda_i^{z_1})^2+\eta^2)((\lambda_j^{z_2})^2+\eta^2)} \left(\left|\langle \boldsymbol{u}_{i}^{z_1}, \boldsymbol{u}_{j}^{z_2}\rangle\right|^2 + \left|\langle \boldsymbol{v}_{i}^{z_1}, \boldsymbol{v}_{j}^{z_2}\rangle\right|^2\right) \\
&\quad = \eta^2 \operatorname{Tr}(\Im G^{z_1})(\Im G^{z_2}) \lesssim \frac{n^{8\omega_d/3}}{(n\eta)^{1/4}} + (\eta^{1/12} + n\eta^2) n^{2\omega_d} \\
&\quad \lesssim \frac{n^{2\omega_d + 100\omega_h}}{n^{1/23}}.
\end{aligned} \tag{6.184}$$

The first inequality in the second line of (6.184) is from Theorem 6.5.2 and the lower bound on $|\widehat{\beta}_*|$ from (6.105). In the last inequality we choose $\eta = n^{-12/23}$, under the assumption that $\omega_d \leq 1/100$ and that $i_0, j_0 \leq n^{1/5}$ (in order to make sure that the first inequality in (6.184) hold). We also used that the first term in the lhs. of the last inequality is always smaller than the other two for $\eta \geq n^{-4/3}$, and in the second line of (6.184) we used that M_{12}, the deterministic approximation of $\operatorname{Tr} \Im G^{z_1} \Im G^{z_2}$ in Theorem 6.5.2, is bounded by $\|M_{12}\| \lesssim |z_1 - z_2|^{-2}$.

This concludes the proof by choosing $\omega_B \leq 1/5$ and $\omega_d = 1/100$, which implies a choice of $\omega_E = -(2\omega_d + 100\omega_h - 1/23)$. □

6.7.4 Pathwise coupling of DBM close to zero

This section is the main technical result used in the proof of Lemma 6.7.6 and Lemma 6.7.7. We compare the evolution of two DBMs whose driving Brownian motions are nearly the same for small indices and are independent for large indices. In Proposition 6.7.13 we will show that the points with small indices in the two processes become very close to each other on a certain time scale t_f. This time scale is chosen to be larger than the local equilibration time, but not too large so that the independence of the driving Brownian motions for large indices do not yet have an effect on particles with small indices.

Remark 6.7.10. *The main result of this section (Proposition 6.7.13) is stated for general deterministic initial data $\boldsymbol{s}(0)$ satisfying Definition 6.7.11 even if for its applications in the proof of Proposition 6.7.2 we only consider initial data which are eigenvalues of i.i.d. random matrices.*

The proof of Proposition 6.7.13 follows the proof of fixed energy universality in [42, 54, 129], adapted to the block structure (6.142) in [54] (see also [53, 55] for further adaptations of [42, 129] to different matrix models). The main novelty in our DBM analysis compared to [42, 54, 129] is that we analyse a process for which we allow not (fully) coupled driving Brownian motions (see Assumption (6.B)).

Define the processes $s_i(t)$, $r_i(t)$ to be the solution of

$$\mathrm{d}s_i(t) = \sqrt{\frac{1}{2n}}\,\mathrm{d}\mathfrak{b}_i^s(t) + \frac{1}{2n}\sum_{j\neq i} \frac{1}{s_i(t) - s_j(t)}\,\mathrm{d}t, \qquad 1 \leq |i| \leq n, \tag{6.185}$$

and

$$\mathrm{d}r_i(t) = \sqrt{\frac{1}{2n}}\,\mathrm{d}\mathfrak{b}_i^r(t) + \frac{1}{2n}\sum_{j\neq i} \frac{1}{r_i(t) - r_j(t)}\,\mathrm{d}t, \qquad 1 \leq |i| \leq n, \tag{6.186}$$

with initial data $s_i(0) = s_i$, $r_i(0) = r_i$, where $\boldsymbol{s} = \{s_{\pm i}\}_{i=1}^n$ and $\boldsymbol{r} = \{r_{\pm i}\}_{i=1}^n$ are two independent sets of particles such that $s_{-i} = -s_i$ and $r_{-i} = -r_i$ for $i \in [n]$. The driving standard real Brownian motions $\{\mathfrak{b}_i^s\}_{i=1}^n$, $\{\mathfrak{b}_i^r\}_{i=1}^n$ in (6.185)–(6.186) are two i.i.d. families and they are such that $\mathfrak{b}_{-i}^s = -\mathfrak{b}_i^s$, $\mathfrak{b}_{-i}^r = -\mathfrak{b}_i^r$ for $i \in [n]$. For convenience we also assume that $\{r_{\pm i}\}_{i=1}^n$ are the singular values of $\widetilde{X}$, with $\widetilde{X}$ a Ginibre matrix. This is not a restriction; indeed, once a process with general initial data $\boldsymbol{s}$ is shown to be close to the reference process with Ginibre initial data, then processes with any two initial data will be close.

Fix an n-dependent parameter $K = K_n = n^{\omega_K}$, for some $\omega_K > 0$. On the correlation structure between the two families of i.i.d. Brownian motions $\{\mathfrak{b}_i^s\}_{i=1}^n$, $\{\mathfrak{b}_i^r\}_{i=1}^n$ we make the following assumptions:

Assumption (6.B). *Suppose that the families $\{\mathfrak{b}_{\pm i}^s\}_{i=1}^n$, $\{\mathfrak{b}_{\pm i}^r\}_{i=1}^n$ in* (6.185) *and*(6.186) *are realised on a common probability space with a common filtration $\mathcal{F}_t$. Let*

$$L_{ij}(t)\,\mathrm{d}t := \mathbf{E}\Big[\big(\mathrm{d}\mathfrak{b}_i^s(t) - \mathrm{d}\mathfrak{b}_i^r(t)\big)\big(\mathrm{d}\mathfrak{b}_j^s(t) - \mathrm{d}\mathfrak{b}_j^r(t)\big)\;\Big|\;\mathcal{F}_t\Big] \tag{6.187}$$

denote the covariance of the increments conditioned on $\mathcal{F}_t$. The processes satisfy the following assumptions:

1. $\{\mathfrak{b}_i^s\}_{i=1}^n$, $\{\mathfrak{b}_i^r\}_{i=1}^n$ *are two families of i.i.d. standard real Brownian motions.*
2. $\{\mathfrak{b}_{\pm i}^r\}_{i=K+1}^n$ *is independent of* $\{\mathfrak{b}_{\pm i}^s\}_{i=1}^n$, *and* $\{\mathfrak{b}_{\pm i}^s\}_{i=K+1}^n$ *is independent of* $\{\mathfrak{b}_{\pm i}^r\}_{i=1}^n$.
3. *Fix $\omega_Q > 0$ so that $\omega_K \ll \omega_Q$. We assume that the subfamilies $\{\mathfrak{b}_{\pm i}^s\}_{i=1}^K$, $\{\mathfrak{b}_{\pm i}^r\}_{i=1}^K$ are very strongly dependent in the sense that for any $|i|, |j| \le K$ it holds*

$$|L_{ij}(t)| \le n^{-\omega_Q} \tag{6.188}$$

   with very high probability for any fixed $t \ge 0$.

Furthermore we assume that the initial data $\{s_{\pm i}\}_{i=1}^n$ is regular in the following sense (cf. [54, Definition 3.1], [129, Definition 2.1], motivated by [130, Definition 2.1]).

Definition 6.7.11 ((g, G)-regular points)**.** *Fix a very small $\nu > 0$, and choose g and G such that*

$$n^{-1+\nu} \le g \le n^{-2\nu}, \qquad G \le n^{-\nu}.$$

A set of $2n$-points $\boldsymbol{s} = \{s_i\}_{i=1}^{2n}$ on $\mathbf{R}$ is called (g, G)-regular *if there exist constants $c_\nu, C_\nu > 0$ such that*

$$c_\nu \le \frac{1}{2n}\Im \sum_{i=-n}^{n} \frac{1}{s_i - (E + \mathrm{i}\eta)} \le C_\nu, \tag{6.189}$$

for any $|E| \le G$, $\eta \in [g, 10]$, and if there is a constant C_s large enough such that $\|\boldsymbol{s}\|_\infty \le n^{C_s}$. Moreover, $c_\nu, C_\nu \sim 1$ if $\eta \in [g, n^{-2\nu}]$ and $c_\nu \ge n^{-100\nu}$, $C_\nu \le n^{100\nu}$ if $\eta \in [n^{-2\nu}, 10]$.

Remark 6.7.12. *We point out that in [54, Definition 3.1] and [129, Definition 2.1] the constants c_ν, C_ν do not depend on $\nu > 0$, but this change does not play any role since ν will always be the smallest exponent of scale involved in the analysis of the DBMs* (6.185)–(6.186)*, hence negligible.*

Let $\rho_{\mathrm{fc},t}(E)$ be the deterministic approximation of the density of the particles $\{s_{\pm i}(t)\}_{i=1}^n$ that is obtained from the semicircular flow acting on the empirical density of the initial data $\{s_{\pm i}(0)\}_{i=1}^n$, see [129, Eq. (2.5)–(2.6)]. Recall that $\rho_{sc}(E)$ denotes the semicircular density.

Proposition 6.7.13. *Let the processes* $\boldsymbol{s}(t) = \{s_{\pm i}(t)\}_{i=1}^n$, $\boldsymbol{r}(t) = \{r_{\pm i}(t)\}_{i=1}^n$ *be the solutions of* (6.185) *and* (6.186)*, respectively, and assume that the driving Brownian motions in* (6.185)–(6.186) *satisfy Assumption (6.B). Additionally, assume that* $\boldsymbol{s}(0)$ *is* (g, G)*-regular in the sense of Definition 6.7.11 and that* $\boldsymbol{r}(0)$ *are the singular values of a Ginibre matrix. Then for any small* $\nu, \omega_f > 0$ *such that* $\nu \ll \omega_K \ll \omega_f \ll \omega_Q$ *and that* $gn^\nu \le t_f \le n^{-\nu}G^2$*, there exist* $\omega, \widehat{\omega} > 0$ *with* $\nu \ll \widehat{\omega} \ll \omega \ll \omega_f$*, and such that it holds*

$$\left|\rho_{\mathrm{fc},t_f}(0)s_i(t_f) - \rho_{\mathrm{sc}}(0)r_i(t_f)\right| \le n^{-1-\omega}, \qquad |i| \le n^{\widehat{\omega}}, \tag{6.190}$$

with very high probability, where $t_f := n^{-1+\omega_f}$.

The proof of Proposition 6.7.13 is postponed to Section 6.7.6.

Remark 6.7.14. *Note that, without loss of generality, it is enough to prove Proposition 6.7.13 only for the case* $\rho_{\mathrm{fc},t_f}(0) = \rho_{sc}(0)$*, since we can always rescale the time: we may define* $\widetilde{s}_i := (\rho_{\mathrm{fc},t_f}(0)s_i/\rho_{sc}(0))$ *and notice that* $\widetilde{s}_i(t)$ *is a solution of the DBM* (6.185) *after rescaling as* $t' = (\rho_{\mathrm{fc},t_f}(0)/\rho_{sc}(0))^2 t$.

6.7.5 Proof of Lemma 6.7.6 and Lemma 6.7.7

In this section we prove that by Lemma 6.7.9 and Proposition 6.7.13 Lemmas 6.7.6–6.7.7 follow.

6.7.5.1 Application of Proposition 6.7.13 to $\boldsymbol{\lambda}^{z_l}(t)$ and $\tilde{\boldsymbol{\lambda}}^{(l)}(t)$

In this section we prove that for any fixed l the processes $\boldsymbol{\lambda}^{z_l}(t)$ and $\widetilde{\boldsymbol{\lambda}}^{(l)}(t)$ satisfy Assumption (6.B), Definition 6.7.11 and so that by Proposition 6.7.13 we conclude the lemma.

Proof of Lemma 6.7.6. For any fix $l \in [p]$, by the definition of the driving Brownian motions of the processes (6.156) and (6.161) it is clear that they satisfy Assumption (6.B) choosing $\boldsymbol{s}(t) = \boldsymbol{\lambda}^{z_l}(t)$, $\boldsymbol{r}(t) = \widetilde{\boldsymbol{\lambda}}^{(l)}(t)$, and $K = n^{\omega_A}$, since $L_{ij}(t) \equiv 0$ for $|i|, |j| \le K$.

We now show that the set of points $\{\lambda_{\pm i}^{z_l}\}_{i=1}^n$, rescaled by $\rho^{z_l}(0)/\rho_{sc}(0)$, is (g, G)-*regular* for

$$g = n^{-1+\omega_h}\delta_l^{-100}, \qquad G = n^{-\omega_h}\delta_l^{10}, \qquad \nu = \omega_h. \tag{6.191}$$

with $\delta_l := 1 - |z_l|^2$, for any $l \in [p]$. By the local law (6.144), together with the regularity properties of m^{z_l} which follow by (6.150), namely that m^{z_l} is $1/3$-Hölder continuous, we conclude that there exist constants $c_{\omega_h}, C_{\omega_h} > 0$ such that

$$c_{\omega_h} \le \Im \frac{1}{2n} \sum_{i=-n}^{n} \frac{1}{[\rho^{z_l}(0)\lambda_i^{z_l}/\rho_{sc}(0)] - (E + \mathrm{i}\eta)} \le C_{\omega_h}, \tag{6.192}$$

for any $|E| \le n^{-\omega_h}\delta_l^{10}$, $n^{-1}\delta_l^{-100} \le \eta \le 10$. In particular, $c_{\omega_h}, C_{\omega_h} \sim 1$ for $\eta \in [g, n^{-2\omega_h}]$, and $c_{\omega_h} \gtrsim n^{-100\omega_h}$, $C_{\omega_h} \lesssim n^{100\omega_h}$ for $\eta \in [n^{-2\omega_h}, 10]$. This implies that the set $\boldsymbol{\lambda}^{z_l} = \{\lambda_{\pm i}^{z_l}\}_{i=1}^n$ satisfies Definition 6.7.11 and it concludes the proof of this lemma. □

6.7.5.2 Application of Proposition 6.7.13 to $\boldsymbol{\mu}^{(l)}(t)$ and $\tilde{\boldsymbol{\mu}}^{(l)}(t)$

We now prove that for any fixed l the processes $\boldsymbol{\mu}^{(l)}(t)$ and $\widetilde{\boldsymbol{\mu}}^{(l)}(t)$ satisfy Assumption (6.B), Definition 6.7.11 and so that by Proposition 6.7.13 we conclude the lemma.

Proof of Lemma 6.7.7. For any fixed $l \in [p]$, we will apply Proposition 6.7.13 with the choice $\boldsymbol{s}(t) = \boldsymbol{\mu}^{(l)}(t)$, $\boldsymbol{r}(t) = \widetilde{\boldsymbol{\mu}}^{(l)}(t)$ and $K = n^{\omega_A}$. Since the initial data $s_i(0) = \mu_i^{(l)}(0)$ are the singular values of a Ginibre matrix $X^{(l)}$, it is clear that the assumption in Definition 6.7.11 holds choosing $g = n^{-1+\delta}$ and $G = n^{-\delta}$, and $\nu = 0$, for any small $\delta > 0$ (see e.g. the local law in (6.144)).

We now check Assumption (6.B). By the definition of the families of i.i.d. Brownian motions

$$\Big(\{\zeta_{\pm i}^{z_l}\}_{i=1}^{n^{\omega_A}}, \{\widetilde{\zeta}_{\pm i}^{(l)}\}_{i=n^{\omega_A}+1}^{n}\Big)_{l=1}^{p}, \qquad \Big(\{\beta_{\pm i}^{(l)}\}_{i=1}^{n}\Big)_{l=1}^{p}, \tag{6.193}$$

defined in (6.162) and (6.159), respectively, it immediately follows that they satisfy 1 and 2 of Assumption (6.B), since $\{\widetilde{\zeta}_{\pm i}^{(l)}\}_{i=n^{\omega_A}+1}^{n}$ are independent of $\{\beta_{\pm i}^{(l)}\}_{i=1}^{n}$ as well as $\{\beta_{\pm i}^{(l)}\}_{i=n^{\omega_A}+1}^{n}$ are independent of $\{\widetilde{\zeta}_{\pm i}^{(l)}\}_{i=1}^{n}$ by construction. Recall that $\mathcal{F}_{\beta,t}$ denotes the common filtration of all the Brownian motions $\boldsymbol{\beta}^{(m)} = \{\beta_i^{(m)}\}_{i=1}^{n}$, $m \in [p]$.

Finally, we prove that also 3 of Assumption (6.B) is satisfied. We recall the relations $i = \mathfrak{i} + (l-1)n^{\omega_A}$ and $j = \mathfrak{j} + (l-1)n^{\omega_A}$ from (6.165) which, for any fixed l, establish a one to one relation between a pair $\mathfrak{i}, \mathfrak{j} \in [n^{\omega_B}]$ and a pair i, j with $(l-1)n^{\omega_A} + 1 \le i, j \le l n^{\omega_A}$. By the definition of $\{\zeta_{\pm i}^{z_l}\}_{i=1}^{n^{\omega_A}}$ it follows that

$$\mathrm{d}\zeta_{\mathfrak{i}}^{z_l} - \mathrm{d}\beta_{\mathfrak{i}}^{(l)} = \sum_{m=1}^{pn^{\omega_A}} \Big(\sqrt{C\#(t)} - I\Big)_{im} \mathrm{d}(\underline{\beta})_m, \qquad 1 \le \mathfrak{i} \le n^{\omega_A}, \tag{6.194}$$

with $\underline{\beta}$ defined in (6.163), and so that for any $1 \le \mathfrak{i}, \mathfrak{j} \le n^{\omega_A}$ and fixed l we have

$$\begin{aligned} &\mathbf{E}\Big[(\mathrm{d}\zeta_{\mathfrak{i}}^{z_l} - \mathrm{d}\beta_{\mathfrak{i}}^{(l)})(\mathrm{d}\zeta_{\mathfrak{j}}^{z_l} - \mathrm{d}\beta_{\mathfrak{j}}^{(l)}) \,\Big|\, \mathcal{F}_{\beta,t}\Big] \\ &= \sum_{m_1,m_2=1}^{pn^{\omega_A}} \Big(\sqrt{C\#(t)} - I\Big)_{im_1} \Big(\sqrt{C\#(t)} - I\Big)_{jm_2} \mathbf{E}\Big[\mathrm{d}(\underline{\beta})_{m_1}\, \mathrm{d}(\underline{\beta})_{m_2} \,\Big|\, \mathcal{F}_{\beta,t}\Big] \\ &= \Big[\Big(\sqrt{C\#(t)} - I\Big)^2\Big]_{ij} \mathrm{d}t, \end{aligned}$$

since $\sqrt{C\#(t)}$ is real symmetric. Hence, $L_{ij}(t)$ defined in (6.187) in this case is given by

$$L_{ij}(t) = \Big[\Big(\sqrt{C\#(t)} - I\Big)^2\Big]_{ij}.$$

Then, by Cauchy-Schwarz inequality, we have that

$$\begin{aligned} |L_{ij}(t)| &\le \Big[\Big(\sqrt{C\#(t)} - I\Big)^2\Big]_{ii}^{1/2} \Big[\Big(\sqrt{C\#(t)} - I\Big)^2\Big]_{jj}^{1/2} \\ &\le \mathrm{Tr}\Big[(\sqrt{C\#(t)} - I)^2\Big] \le \mathrm{Tr}\Big[(C\#(t) - I)^2\Big] \lesssim \frac{p^2 n^{2\omega_A}}{n^{4\omega_E}}, \end{aligned} \tag{6.195}$$

with very high probability, where in the last inequality we used that $C^{\#}(t)$ and $C(t)$ have the same distribution and the bound (6.183) of Lemma 6.7.9 holds for $C(t)$ hence for $C^{\#}(t)$ as well. This implies that for any fixed $l \in [p]$ the two families of Brownian motions $\{\beta_{\pm i}^{(l)}\}_{i=1}^{n}$ and $(\{\zeta_{\pm i}^{z_l}\}_{i=1}^{n^{\omega_A}}, \{\widetilde{\zeta}_{\pm i}^{(l)}\}_{i=n^{\omega_A}+1}^{n})$ satisfy Assumption (6.B) with $K = n^{\omega_A}$ and $\omega_Q = 4\omega_E - 2\omega_A$. Applying Proposition 6.7.13 this concludes the proof of Lemma 6.7.7. □

6.7.6 Proof of Proposition 6.7.13

We divide the proof of Proposition 6.7.13 into four sub-sections. In Section 6.7.6.1 we introduce an interpolating process $\boldsymbol{x}(t, \alpha)$ between the processes $\boldsymbol{s}(t)$ and $\boldsymbol{r}(t)$ defined in (6.185)–(6.186), and in Section 6.7.6.2 we introduce a measure which approximates the particles $\boldsymbol{x}(t, \alpha)$ and prove their rigidity. In Section 6.7.6.3 we introduce a cut-off near zero (this scale will be denoted by ω_A later) such that we only couple the dynamics of the particles $|i| \leq n^{\omega_A}$, as defined in 3 of Assumption (6.B), i.e. we will choose $\omega_A = \omega_K$. Additionally, we also localise the dynamics on a scale ω_l (see Section 6.7.2.3) since the main contribution to the dynamics comes from the nearby particles. We will refer to the new process $\widehat{\boldsymbol{x}}(t, \alpha)$ (see (6.209) later) as the *short range approximation* of the process $\boldsymbol{x}(t, \alpha)$. Finally, in Section 6.7.6.4 we conclude the proof of Proposition 6.7.13.

Large parts of our proof closely follow [54, 129] and for brevity we will focus on the differences. We use [54, 129] as our main references since the 2×2 block matrix setup of [54] is very close to the current one and [54] itself closely follows [129]. However, we point out that many key ideas of this technique have been introduced in earlier papers on universality; e.g. short range cut-off and finite speed of propagation in [39, 91], coupling and homogenisation in [42]; for more historical references, see [129]. The main novelty of [129] itself is a mesoscopic analysis of the fundamental solution $p_t(x, y)$ of (6.220) which enables the authors to prove short time universality for general deterministic initial data. They also proved the result with very high probability unlike [42] that relied on level repulsion estimates. We also mention a related but different more recent technique to prove universality [40], which has been recently adapted to the singular values setup, or equivalently to the 2×2 block matrix structure, in [208].

6.7.6.1 Definition of the interpolated process

For $\alpha \in [0, 1]$ we introduce the continuous interpolation process $\boldsymbol{x}(t, \alpha)$, between the processes $\boldsymbol{s}(t)$ and $\boldsymbol{r}(t)$ in (6.185)–(6.186), defined as the solution of the flow

$$\mathrm{d}x_i(t, \alpha) = \alpha \frac{\mathrm{d}\mathfrak{b}_i^s}{\sqrt{2n}} + (1 - \alpha) \frac{\mathrm{d}\mathfrak{b}_i^r}{\sqrt{2n}} + \frac{1}{2n} \sum_{j \neq i} \frac{1}{x_i(t, \alpha) - x_j(t, \alpha)} \, \mathrm{d}t, \tag{6.196}$$

with initial data

$$\boldsymbol{x}(0, \alpha) = \alpha \boldsymbol{s}(t_0) + (1 - \alpha) \boldsymbol{r}(t_0), \tag{6.197}$$

with some t_0 that is a slightly smaller than t_f. In fact we will write $t_0 + t_1 = t_f$ with $t_1 \ll t_f$, where t_1 is the time scale for the equilibration of the DBM with initial condition (6.197) (see (6.205)). To make our notation consistent with [54, 129] in the remainder of this section we assume that $t_0 = n^{-1+\omega_0}$, for some small $\omega_0 > 0$, such that $\omega_K \ll \omega_0 \ll \omega_Q$. The reader can think of $\omega_0 = \omega_f$. Note that the strong solution of (6.196) is well defined since the variance of its driving Brownian motion is smaller than $\frac{1}{2n}(1 - 2\alpha(1 - \alpha)n^{-\omega_Q})$

by (6.188), which is below the critical variance for well-posedness of the DBM since we are in the complex symmetry class (see e.g. [17, Lemma 4.3.3]).

By (6.196) it clearly follows that $\boldsymbol{x}(t,0) = \boldsymbol{r}(t+t_0)$ and $\boldsymbol{x}(t,1) = \boldsymbol{s}(t+t_0)$, for any $t \geq 0$. Note that the process (6.196) is almost the same as [129, Eq. (3.13)], [54, Eq. (3.13)], except for the stochastic term, which in our case depends on α. Also, to make the notation clearer, we remark that in [54, 129] the interpolating process is denoted by $\boldsymbol{z}(t,\alpha)$. We changed this notation to $\boldsymbol{x}(t,\alpha)$ to avoid confusions with the z_l-parameters introduced in the previous sections where we apply Proposition 6.7.13 to the processes defined in Section 6.7.2.1.

Remark 6.7.15. *Even if all processes $\boldsymbol{\lambda}(t)$, $\widetilde{\boldsymbol{\lambda}}(t)$, $\widetilde{\boldsymbol{\mu}}(t)$, $\boldsymbol{\mu}(t)$ introduced in Section 6.7.2.1 already satisfy [54, Lemma 3.3-3.5], [129, Lemma 3.3-3.5] as a consequence of the local law* (6.144) *and the rigidity estimates* (6.145)*, we decided to present the proof of Proposition 6.7.13 for general deterministic initial data $\boldsymbol{s}(0)$ satisfying Definition 6.7.11 (see Remark 6.7.10). Hence, an additional time t_0 is needed to ensure the validity of [54, Lemma 3.3-3.5], [129, Lemma 3.3-3.5]. More precisely, we first let the DBMs* (6.185)–(6.186) *evolve for a time $t_0 := n^{-1+\omega_0}$, and then we consider the process* (6.196) *whose initial data in* (6.197) *is given by a linear interpolation of the solutions of* (6.185)–(6.186) *at time t_0.*

Before proceeding with the analysis of (6.196) we give some definitions and state some preliminary results necessary for its analysis.

6.7.6.2 Interpolating measures and particle rigidity

Using the convention of [54, Eq. (3.10)–(3.11)], given a probability measure $\mathrm{d}\rho(E)$, we define the $2n$-quantiles γ_i by

$$
\begin{aligned}
\gamma_i &:= \inf\left\{x \,\middle|\, \int_{-\infty}^{x} \mathrm{d}\rho(E) \geq \frac{n+i-1}{2n}\right\}, \quad 1 \leq i \leq n, \\
\gamma_i &:= \inf\left\{x \,\middle|\, \int_{-\infty}^{x} \mathrm{d}\rho(E) \geq \frac{n+i}{2n}\right\}, \qquad -n \leq i \leq -1,
\end{aligned} \tag{6.198}
$$

Note that $\gamma_1 = 0$ if $\mathrm{d}\rho(E)$ is symmetric with respect to 0.

Let $\rho_{fc,t}(E)$ be defined above Proposition 6.7.13 (see e.g. [129, Eq. (2.5)–(2.6)] for more details), and let $\rho_{sc}(E)$ denote the semicircular density, then by $\gamma_i(t)$, γ_i^{sc} we denote the $2n$-quantiles, defined as in (6.198), of $\rho_{fc,t}$ and ρ_{sc}, respectively.

Following the construction of [129, Lemma 3.3-3.4, Appendix A], [54, Section 3.2.1], we define the interpolating (random) measure $\mathrm{d}\rho(E,t,\alpha)$ for any $\alpha \in [0,1]$. More precisely, the measure $\mathrm{d}\rho(E,t,\alpha)$ is deterministic close to zero, and it consists of delta functions of the position of the particles $x_i(t,\alpha)$ away from zero.

Denote by $\gamma_i(t,\alpha)$ the quantiles of $\mathrm{d}\rho(E,\alpha,t)$, and by $m(w,t,\alpha)$, with $w \in \mathbf{H}$, its Stieltjes transform. Fix $q_* \in (0,1)$ throughout this section, and let $k_0 = k_0(q_*) \in \mathbf{N}$ be the largest index such that

$$
|\gamma_{\pm k_0}(t_0)|, |\gamma_{\pm k_0}^{sc}| \leq q_* G, \tag{6.199}
$$

with G defined in (6.191), then the measure $\mathrm{d}\rho(E,t,\alpha)$ has a deterministic density (denoted by $\rho(E,\alpha,t)$ with a slight abuse of notation) on the interval

$$
\mathcal{G}_\alpha := [\alpha\gamma_{-k_0}(t_0) + (1-\alpha)\gamma_{-k_0}^{sc}, \alpha\gamma_{k_0}(t_0) + (1-\alpha)\gamma_{k_0}^{sc}]. \tag{6.200}
$$

Outside $\mathcal{G}_\alpha$ the measure $\mathrm{d}\rho(E,t,\alpha)$ consists of $1/(2n)$ times delta functions of the particle locations $\delta_{x_i(t,\alpha)}$.

Remark 6.7.16. *By the construction* $\mathrm{d}\rho(E,t,\alpha)$ *as in [129, Lemma 3.3–3.4, Appendix A], [54, Section 3.2.1] all the regularity properties of* $\mathrm{d}\rho(E,\alpha,t)$*, its quantiles* $\gamma_i(t,\alpha)$*, and its Stieltjes transform* $m(E+\mathrm{i}\eta,t,\alpha)$ *in [129, Lemma 3.3–3.4], [54, Lemma 3.3–3.4] hold without any change. In particular, it follows that*

$$|\gamma_i(t,\alpha)-\gamma_j(t,\alpha)|\sim\frac{|i-j|}{n},\qquad |i|,|j|\le q_*G, \tag{6.201}$$

with q_* *defined above* (6.199)*, and* G *in* (6.191).

Define the Stieltjes transform of the empirical measure of the particles $\{x_{\pm i}(t,\alpha)\}_{i=1}^n$ by

$$m_n(w,t,\alpha):=\frac{1}{2n}\sum_{i=-n}^{n}\frac{1}{x_i(t,\alpha)-w},\quad w\in\mathbf{H}. \tag{6.202}$$

We recall that the summation does not include the term $i=0$ (see Remark 6.7.1). Then by the local law and optimal rigidity for short time for singular values in [54, Lemma 3.5], which has been proven adapting the local laws for short time of [129, Appendix A-B] and [130, Section 3], we conclude the following local law and optimal eigenvalue rigidity.

Lemma 6.7.17. *Fix* $q\in(0,1)$ *and* $\tilde{\epsilon}>0$*. Define* $\widehat{C}_q:=\{j:|j|\le qk_0\}$*, with* k_0 *defined in* (6.199)*. Then for any* $\xi>0$*, with very high probability we have the optimal rigidity*

$$\sup_{0\le t\le t_0 n^{-\tilde{\epsilon}}}\sup_{i\in\widehat{C}_q}\sup_{0\le\alpha\le1}|x_i(t,\alpha)-\gamma_i(t,\alpha)|\le\frac{n^{\xi+100\nu}}{n}, \tag{6.203}$$

and the local law

$$\sup_{n^{-1+\tilde{\epsilon}}\le\eta\le10}\sup_{0\le t\le t_0 n^{-\tilde{\epsilon}}}\sup_{0\le\alpha\le1}\sup_{E\in q\mathcal{G}\alpha}|m_n(E+\mathrm{i}\eta,t,\alpha)-m(E+\mathrm{i}\eta,t,\alpha)|\le\frac{n^{\xi+100\nu}}{n\eta}, \tag{6.204}$$

for sufficiently large n*, with* $\nu>0$ *from in Definition 6.7.11.*

Without loss of generality in Lemma 6.7.17 we assumed $k_1=k_0$ in [54, Eq. (3.25)–(3.26)].

6.7.6.3 Short range analysis

In the following of this section we perform a local analysis of (6.196) adapting the analysis of [54, 129] and explaining the minor changes needed for the analysis of the flow (6.196), for which the driving Brownian motions $\mathfrak{b}^s$, $\mathfrak{b}^r$ satisfy Assumption (6.B), compared to the analysis of [54, Eq. (3.13)], [129, Eq. (3.13)]. More precisely, we run the DBM (6.196) for a time

$$t_1:=\frac{n^{\omega_1}}{n}, \tag{6.205}$$

for any $\omega_1>0$ such that $\nu\ll\omega_1\ll\omega_K$, with ν,ω_K defined in Definition 6.7.11 and above Assumption (6.B), respectively, so that (6.196) reaches its local equilibrium (see Section 6.7.2.3 for a summary on the different scales). Moreover, since the dynamics of $x_i(t,\alpha)$ is mostly influenced by the particles close to it, in the following we define a short range approximation of the process $\boldsymbol{x}(t,\alpha)$ (see (6.209) later), denoted by $\widehat{\boldsymbol{x}}(t,\alpha)$, and use the homogenisation theory developed in [129], adapted in [54] for the singular values flow, for the short range kernel.

Remark 6.7.18. *We do not need to define the shifted process $\widetilde{\boldsymbol{x}}(t,\alpha)$ as in [54, Eq. (3.29)–(3.32)] and [129, Eq. (3.36)–(3.40)], since in our case the measure $\mathrm{d}\rho(E,t,\alpha)$ is symmetric with respect to 0 by assumption, hence, using the notation in [54, Eq. (3.29)–(3.32)], we have $\widetilde{\boldsymbol{x}}(t,\alpha) = \boldsymbol{x}(t,\alpha) - \gamma_1(t,\alpha) = \boldsymbol{x}(t,\alpha)$. Hence, from now on we only use $\boldsymbol{x}(t,\alpha)$ and the reader can think $\widetilde{\boldsymbol{x}}(t,\alpha) \equiv \boldsymbol{x}(t,\alpha)$ for a direct analogy with [54, 129].*

Our analysis will be completely local, hence we introduce a short range cut-off. Fix $\omega_l, \omega_A > 0$ so that

$$0 < \omega_1 \ll \omega_l \ll \omega_A \ll \omega_0 \ll \omega_Q, \tag{6.206}$$

with ω_1 defined in (6.205), ω_0 defined below (6.197), and ω_Q in 3 of Assumption (6.B). Moreover, we assume that ω_A is such that

$$K_n = n^{\omega_A}, \tag{6.207}$$

with $K_n = n^{\omega_K}$ in Assumption (6.B), i.e. $\omega_A = \omega_K$. We remark that it is enough to choose $\omega_A \ll \omega_K$, but to avoid further splitting in (6.209) we assumed $\omega_K = \omega_A$.

For any $q \in (0,1)$, define the set

$$A_q := \Big\{ (i,j) \,\Big|\, |i-j| \le n^{\omega_l} \text{ or } ij > 0, i \notin \widehat{C}_q, j \notin \widehat{C}_q \Big\}, \tag{6.208}$$

and denote $A_{q,(i)} := \{ j \mid (i,j) \in A_q \}$. In the remainder of this section we will often use the notations

$$\sum_j^{A_{q,(i)}} := \sum_{j \in A_{q,(i)}}, \qquad \sum_j^{A_{q,(i)}^c} := \sum_{j \notin A_{q,(i)}}.$$

Let $q_* \in (0,1)$ be defined above (6.199), then we define the short range process $\widehat{\boldsymbol{x}}(t,\alpha)$ (cf. [54, Eq. (3.35)–(3.36)], [129, Eq. (3.45)–(3.46)]) as follows

$$\begin{aligned} \mathrm{d}\widehat{x}_i(t,\alpha) &= \frac{1}{2n} \sum_j^{A_{q_*,(i)}} \frac{1}{\widehat{x}_i(t,\alpha) - \widehat{x}_j(t,\alpha)} \mathrm{d}t \\ &+ \begin{cases} \alpha \frac{\mathrm{d}\mathfrak{b}^s}{\sqrt{2n}} + (1-\alpha) \frac{\mathrm{d}\mathfrak{b}^r}{\sqrt{2n}} & \text{if } \quad |i| \le n^{\omega_A}, \\ \alpha \frac{\mathrm{d}\mathfrak{b}^s}{\sqrt{2n}} + (1-\alpha) \frac{\mathrm{d}\mathfrak{b}^r}{\sqrt{2n}} + J_i(\alpha,t)\,\mathrm{d}t & \text{if } \quad n^{\omega_A} < |i| \le n, \end{cases} \end{aligned} \tag{6.209}$$

where

$$J_i(\alpha,t) := \frac{1}{2n} \sum_j^{A_{q_*,(i)}^c} \frac{1}{x_i(t,\alpha) - x_j(t,\alpha)}, \tag{6.210}$$

and initial data $\widehat{\boldsymbol{x}}(0,\alpha) = \boldsymbol{x}(0,\alpha)$. Note that

$$\sup_{0 \le t \le t_1} \sup_{0 \le \alpha \le 1} |J_1(\alpha,t)| \le \log n, \tag{6.211}$$

with very high probability.

Remark 6.7.19. *Note that the SDE defined in* (6.209) *has the same form as in [129, Eq. (3.70)], with $F_i = 0$ in our case, except for the stochastic term in* (6.209) *that looks slightly different, in particular it depends on α. Nevertheless, by Assumption (6.B), the quadratic variation of the driving Brownian motions in* (6.209) *is also bounded by one uniformly in $\alpha \in [0,1]$. Moreover, the process defined in* (6.209) *and the measure $\mathrm{d}\rho(E,t,\alpha)$ satisfy [129, Eq. (3.71)–(3.77)].*

Since when we consider the difference process $\widehat{\boldsymbol{x}}(t,\alpha)-\boldsymbol{x}(t,\alpha)$ the stochastic differential disappears, by [129, Lemma 3.8], without any modification, it follows that

$$\sup_{0\le t\le t_1}\sup_{0\le\alpha\le 1}\sup_{|i|\le n}|\widehat{x}_i(t,\alpha)-x_i(t,\alpha)|\le n^{\xi+100\nu}t_1\left(\frac{1}{n^{\omega_l}}+\frac{n^{\omega_A}}{n^{\omega_0}}+\frac{1}{\sqrt{nG}}\right), \qquad (6.212)$$

for any $\xi>0$ with very high probability, with G defined in (6.191). In particular, (6.212) implies that the short range process $\widehat{\boldsymbol{x}}(t,\alpha)$, defined in (6.209), approximates very well (i.e. they are closer than the fluctuation scale) the process $\boldsymbol{x}(t,\alpha)$ defined in (6.196).

Next, in order to use the smallness of (6.187)–(6.188) in Assumption (6.B) for $|i|\le n^{\omega_A}$, we define $\boldsymbol{u}(t,\alpha):=\partial_\alpha\widehat{\boldsymbol{x}}(t,\alpha)$, which is the solution of the following discrete SPDE (cf. [54, Eq. (3.38)], [129, Eq. (3.63)]):

$$\mathrm{d}\boldsymbol{u}=\sum_j^{A_{q*,(i)}}B_{ij}(u_j-u_i)\,\mathrm{d}t+\mathrm{d}\boldsymbol{\xi}_1+\boldsymbol{\xi}_2\,\mathrm{d}t=-\mathcal{B}\boldsymbol{u}\,\mathrm{d}t+\mathrm{d}\boldsymbol{\xi}_1+\boldsymbol{\xi}_2\,\mathrm{d}t, \qquad (6.213)$$

where

$$\begin{aligned} B_{ij}&:=\frac{\mathbf{1}_{j\neq\pm i}}{2n(\widehat{x}_i-\widehat{x}_j)^2},\quad \mathrm{d}\xi_{1,i}:=\frac{\mathrm{d}\mathfrak{b}_i^s}{\sqrt{2n}}-\frac{\mathrm{d}\mathfrak{b}_i^r}{\sqrt{2n}}\\ \xi_{2,i}&:=\begin{cases}0 & \text{if}\quad |i|\le n^{\omega_A},\\ \partial_\alpha J_i(\alpha,t) & \text{if}\quad n^{\omega_A}<|i|\le n,\end{cases}\end{aligned} \qquad (6.214)$$

with $J_i(\alpha,t)$ defined in (6.210). We remark that the operator[5] $\mathcal{B}$ defined via the kernel in (6.214) depends on α and t. It is not hard to see (e.g. see [129, Eq. (3.65), Eq. (3.68)–(3.69)]) that the forcing term $\boldsymbol{\xi}_2$ is bounded with very high probability by n^C, for some $C>0$, for $n^{\omega_A}<|i|\le n$. Note that the only difference in (6.213) compared to [54, Eq. (3.38)], [129, Eq. (3.63)] is the additional term $\mathrm{d}\boldsymbol{\xi}_1$ which will be negligible for our analysis.

Let $\mathcal{U}$ be the semigroup associated to $\mathcal{B}$, i.e. if $\partial_t\boldsymbol{v}=-\mathcal{B}\boldsymbol{v}$, then for any $0\le s\le t$ we have that

$$v_i(t)=\sum_{j=-n}^n\mathcal{U}_{ij}(s,t,\alpha)v_j(s),\qquad |i|\le n.$$

The first step to analyse the equation in (6.213) is the following finite speed of propagation estimate (cf. [54, Lemma 3.9], [129, Lemma 3.7]).

Lemma 6.7.20. *Let $0\le s\le t\le t_1$. Fix $0<q_1<q_2<q_*$, with $q_*\in(0,1)$ defined in* (6.199)*, and $\epsilon_1>0$ such that $\epsilon_1\ll\omega_A$. Then for any $\alpha\in[0,1]$ we have*

$$|U_{ji}(s,t,\alpha)|+|U_{ij}(s,t,\alpha)|\le n^{-D}, \qquad (6.215)$$

for any $D>0$ with very high probability, if either $i\in\widehat{C}_{q_2}$ and $|i-j|>n^{\omega_l+\epsilon_1}$, or if $i\notin\widehat{C}_{q_2}$ and $j\in\widehat{C}_{q_1}$.

[5]The operator $\mathcal{B}$ defined here is not to be confused with the completely unrelated one in (6.104).

Proof. The proof of this lemma follows the same lines as [129, Lemma 3.7]. There are only two differences that we point out. The first one is that [129, Eq. (4.15)], using the notation therein, has to be replaced by

$$\sum_k v_k^2(\nu^2(\psi_k')^2 + \nu\psi_k'')\,\mathbf{E}[\mathrm{d}C_k(\alpha,t)\,\mathrm{d}C_k(\alpha,t)\mid\mathcal{F}_t], \tag{6.216}$$

where $\mathcal{F}_t$ is the filtration defined in Assumption (6.B), and $C_k(\alpha,t)$ is defined as

$$C_k(\alpha,t) := \alpha\frac{\mathfrak{b}_k^s(t)}{\sqrt{2n}} + (1-\alpha)\frac{\mathfrak{b}_k^r(t)}{\sqrt{2n}}. \tag{6.217}$$

We remark that ν in (6.216) should not to be confused with ν in Definition 6.7.11. Then, by Kunita-Watanabe inequality, it is clear that

$$\mathbf{E}[\mathrm{d}C_k(\alpha,t)\,\mathrm{d}C_k(\alpha,t)\mid\mathcal{F}_t] \lesssim \frac{\mathrm{d}t}{n}, \tag{6.218}$$

uniformly in $|k|\le n$, $t\ge 0$, and $\alpha\in[0,1]$. The fact that (6.218) holds is the only input needed to bound [129, Eq. (4.21)].

The second difference is that the stochastic differential $(\sqrt{2}\,\mathrm{d}B_k)/\sqrt{n}$ in [129, Eq. (4.21)] has to be replaced by $\mathrm{d}C_k(\alpha,t)$ defined in (6.217). This change is inconsequential in the bound [129, Eq. (4.26)], since $\mathbf{E}\,\mathrm{d}C_k(\alpha,t)=0$. □

Moreover, the result in [54, Lemma 3.8], [129, Lemma 3.10] hold without any change, since its proof is completely deterministic and the stochastic differential in the definition of the process $\widehat{\boldsymbol{x}}(t,\alpha)$ does not play any role.

In the remainder of this section, before completing the proof of Proposition 6.7.13, we describe the homogenisation argument to approximate the t-dependent kernel of $\mathcal{B}$ with a continuous kernel (denoted by $p_t(x,y)$ below). We follow verbatim [129, Section 3-4] and its adaptation to the singular value flow of [54, Section 3.4], except for the bound of the rhs. of (6.233), where we handle the additional term $\mathrm{d}\boldsymbol{\xi}_1$ in (6.214).

Fix a constant $\epsilon_B>0$ such that $\omega_A-\epsilon_B>\omega_l$, and let $a\in\mathbf{Z}$ be such that $0<|a|\le n^{\omega_A-\epsilon_B}$. Define also the equidistant points $\gamma_j^f := j(2n\rho_{sc}(0))^{-1}$, which approximate the quantiles $\gamma_j(t,\alpha)$ very well for small j, i.e. $|\gamma_j^f-\gamma_j(t,\alpha)|\lesssim n^{-1}$ for $|j|\le n^{\omega_0/2}$ (see [129, Eq. (3.91)]). Consider the solution of

$$\partial_t w_i = -(\mathcal{B}w)_i, \quad w_i(0) = 2n\delta_{ia}, \tag{6.219}$$

and define the cut-off $\eta_l := n^{\omega_l}(2n\rho_{sc}(0))^{-1}$. Let $p_t(x,y)$ be the fundamental solution of the equation

$$\partial_t f(x) = \int_{|x-y|\le\eta_l}\frac{f(y)-f(x)}{(x-y)^2}\rho_{sc}(0)\,\mathrm{d}y. \tag{6.220}$$

The idea of the homogenisation argument is that the deterministic solution f of (6.220) approximates very well the random solution of (6.219). This is formulated in terms of the solution kernels of the two equations in Proposition 6.7.21. Following [54, Lemma 3.9-3.13, Corollary 3.14, Theorem 3.15-3.17], which are obtained adapting the proof of [129, Section 3.6], we will conclude the following proposition.

Proposition 6.7.21. *Let $a, i \in \mathbf{Z}$ such that $|a| \le n^{\omega_A - \epsilon_B}$ and $|i-a| \le n^{\omega_l}/10$. Fix $\epsilon_c > 0$ such that $\omega_1 - \epsilon_c > 0$, let $t_1 := n^{-1+\omega_1}$ and $t_2 := n^{-\epsilon_c} t_1$, then for any $\alpha \in [0,1]$ and for any $|u| \le t_2$ we have*

$$\left| \mathcal{U}_{ia}(0, t_1 + u, \alpha) - \frac{p_{t_1}(\gamma_i^f, \gamma_a^f)}{n} \right| \le \frac{n^{100\nu + \epsilon_c}}{n t_1} \left(\frac{(nt_1)^2}{n^{\omega_l}} + \frac{1}{(nt_1)^{1/10}} + \frac{1}{n^{3\epsilon_c/2}} \right), \quad (6.221)$$

with very high probability.

Proof. The proof of this proposition relies on [129, Section 3.6], which has been adapted to the 2×2 block structure in [54, Lemma 3.9–3.13, Corollary 3.14, Theorem 3.15–3.17]. We thus present only the differences compared to[54, 129]; for a complete proof we defer the reader to these works.

The only difference in the proof of this proposition compared to the proof of [54, Theorem 3.17], [129, Theorem 3.11] is in [129, Eq. (3.121) of Lemma 3.14] and [129, Eq. (3.148) of Lemma 3.14]. The main goal of [129, Lemma 3.14] and [129, Lemma 3.14] is to prove that

$$\mathrm{d}\frac{1}{2n} \sum_{1 \le |i| \le n} (w_i - f_i)^2 = -\langle \boldsymbol{w}(t) - \boldsymbol{f}(t), \mathcal{B}(\boldsymbol{w}(t) - \boldsymbol{f}(t)) \rangle + \text{Lower order}, \quad (6.222)$$

where $f_i := f(\widehat{x}_i(t,\alpha), t)$, with $\widehat{x}_i(t,\alpha)$ being the solution of (6.209), and $\boldsymbol{w}(t)$, $\boldsymbol{f}(t)$ being the solutions of (6.219) and (6.220) with $x = \widehat{x}_i(t,\alpha)$, respectively. In order to prove (6.222), following [129, Eq. (3.121)] and using the notation therein (with $N = 2n$ and replacing $\widehat{z}_i$ by $\widehat{x}_i$), we compute

$$\begin{aligned} &\mathrm{d}\frac{1}{2n} \sum_{1 \le |i| \le n} (w_i - f_i)^2 \\ &= \frac{1}{n} \sum_{1 \le |i| \le n} (w_i - f_i) \left[\partial_t w_i \,\mathrm{d}t - (\partial_t f)(t, \widehat{x}_i) \,\mathrm{d}t - f'(t, \widehat{x}_i) \,\mathrm{d}\widehat{x}_i \right] \\ &\quad + \frac{1}{n} \sum_{1 \le |i| \le n} \left(-(w_i - f_i) f''(t, \widehat{x}_i) + (f'(t, \widehat{x}_i))^2 \right) \mathbf{E}[\mathrm{d}C_i(\alpha,t) \,\mathrm{d}C_i(\alpha,t) \mid \mathcal{F}_t], \end{aligned} \quad (6.223)$$

where

$$C_i(\alpha, t) := \alpha \frac{\mathfrak{b}_i^s(t)}{\sqrt{2n}} + (1-\alpha) \frac{\mathfrak{b}_i^r(t)}{\sqrt{2n}}.$$

As a consequence of the slight difference in definition of $\mathrm{d}\widehat{x}_i$ in (6.209), compared to the definition of $\mathrm{d}\widehat{z}_i$ in [129, Eq. (3.70)], the martingale term in (6.223) is given by (cf. [129, Eq. (3.148)])

$$\mathrm{d}M_t = \frac{1}{2n} \sum_{1 \le |i| \le n} (w_i - f_i) f_i' \,\mathrm{d}C_i(\alpha, t). \quad (6.224)$$

The terms in the first line of the rhs. of (6.223) are bounded exactly as in [129, Eq. (3.124)–(3.146), (3.149)–(3.154)]. It remains to estimate the second line in the rhs. of (6.223).

The expectation of the second line of (6.223) is bounded by a constant times $n^{-1}\,\mathrm{d}t$, exactly as in (6.218). This is the only input needed to bound the terms (6.223) in [129, Eq. (3.122)-(3.123)]. Hence, in order to conclude the proof of this proposition we are left with the term in (6.224).

The quadratic variation of the term in (6.224), using the notation in [129, Eq. (3.155)–(3.157)], is given by

$$\mathrm{d}\langle M\rangle_t = \frac{1}{2n}\sum_{1\le |i|,|j|\le n}(w_i-f_i)(w_j-f_j)f_i'f_j'\,\mathbf{E}[\mathrm{d}C_i(\alpha,t)\,\mathrm{d}C_j(\alpha,t)\mid \mathcal{F}_t].$$

By 2 of Assumption (6.B) it follows that

$$\begin{aligned}\mathrm{d}\langle M\rangle_t = \frac{1}{4n^2}&\sum_{1\le |i|,|j|\le n^{\omega_A}}(w_i-f_i)(w_j-f_j)f_i'f_j'\,\mathbf{E}[\mathrm{d}C_i(\alpha,t)\,\mathrm{d}C_j(\alpha,t)\mid \mathcal{F}_t]\\ &+\frac{\alpha^2+(1-\alpha)^2}{8n^3}\sum_{n^{\omega_A}<|i|\le n}(w_i-f_i)^2(f_i')^2\,\mathrm{d}t.\end{aligned}\tag{6.225}$$

Then, by 3 of Assumption (6.B), for $|i|,|j|\le n^{\omega_A}$ we have

$$\begin{aligned}\mathbf{E}[\mathrm{d}C_i(\alpha,t)\,\mathrm{d}C_j(\alpha,t)\mid \mathcal{F}_t] &= \big[\alpha^2+(1-\alpha)^2\big]\frac{\delta_{ij}}{2n}\,\mathrm{d}t\\ &\quad+\frac{\alpha(1-\alpha)}{2n}\,\mathbf{E}\Big[\big(\mathrm{d}\mathfrak{b}_i^s\,\mathrm{d}\mathfrak{b}_j^r+\mathrm{d}\mathfrak{b}_i^r\,\mathrm{d}\mathfrak{b}_j^s\big)\,\Big|\,\mathcal{F}_t\Big],\end{aligned}\tag{6.226}$$

and that

$$\mathbf{E}\Big[\mathrm{d}\mathfrak{b}_i^s\,\mathrm{d}\mathfrak{b}_j^r\,\Big|\,\mathcal{F}_t\Big]=\mathbf{E}\Big[\big(\mathrm{d}\mathfrak{b}_i^s-\mathrm{d}\mathfrak{b}_i^r\big)\,\mathrm{d}\mathfrak{b}_j^r\,\Big|\,\mathcal{F}_t\Big]+\delta_{ij}\,\mathrm{d}t\lesssim \big(|L_{ii}(t)|^{1/2}+\delta_{ij}\big)\,\mathrm{d}t,\tag{6.227}$$

where in the last inequality we used Kunita-Watanabe inequality.

Combining (6.225)–(6.227) we finally conclude that

$$\begin{aligned}\mathrm{d}\langle M\rangle_t \le \frac{1}{8n^3}&\sum_{1\le|i|\le n}(w_i-f_i)^2(f_i')^2\,\mathrm{d}t\\ &+\frac{\alpha(1-\alpha)}{4n^3}\sum_{1\le|i|,|j|\le n^{\omega_A}}|L_{ii}(t)|^{1/2}\Big|(w_i-f_i)(w_j-f_j)f_i'f_j'\Big|\,\mathrm{d}t.\end{aligned}\tag{6.228}$$

Since $\alpha\in[0,1]$, $|L_{ii}(t)|\le n^{-\omega_Q}$ and $\omega_A\ll\omega_Q$ by (6.188) and (6.206)–(6.207), using Cauchy-Schwarz in (6.228), we conclude that

$$\mathrm{d}\langle M\rangle_t\lesssim\frac{1}{n^3}\sum_{1\le|i|\le n}(w_i-f_i)^2(f_i')^2\,\mathrm{d}t,\tag{6.229}$$

which is exactly the lhs. in [129, Eq. (3.155)], hence the high probability bound in [129, Eq. (3.155)] follows. Then the remainder of the proof of [129, Lemma 3.14] proceeds exactly in the same way.

Given (6.223) as an input, the proof of (6.221) is concluded following the proof of [129, Theorems 3.16-3.17] line by line.

□

6.7.6.4 Proof of Proposition 6.7.13

We conclude this section with the proof of Proposition 6.7.13 following [54, Section 3.6]. We remark that all the estimates above hold uniformly in $\alpha \in [0,1]$ when bounding an integrand by [129, Appendix E].

Proof of Proposition 6.7.13. For any $|i| \le n$, by (6.212), it follows that

$$s_i(t_0+t_1)-r_i(t_0+t_1) = x_i(t_1,1)-x_i(t_1,0) = \widehat{x}_i(t_1,1)-\widehat{x}_i(t_1,0)+\mathcal{O}\left(\frac{n^{\xi}t_1}{n^{\omega_l}}\right). \quad (6.230)$$

We remark that in (6.230) we ignored the scaling (6.190) since it can be removed by a simple time-rescaling (see Remark 6.7.14 for more details). Then, using that $u_i = \partial_\alpha \widehat{x}_i$ we have that

$$\widehat{x}_i(t_1,1) - \widehat{x}_i(t_1,0) = \int_0^1 u_i(t_1,\alpha)\,\mathrm{d}\alpha. \quad (6.231)$$

We recall that u is a solution of

$$\mathrm{d}\boldsymbol{u} = \mathcal{B}\boldsymbol{u}\,\mathrm{d}t + \mathrm{d}\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2\,\mathrm{d}t,$$

as defined in (6.213)–(6.214), with

$$|\xi_{2,i}(t)| \le \mathbf{1}_{\{|i|>n^{\omega_A}\}} n^C, \quad (6.232)$$

with very high probability for some constant $C > 0$ and any $0 \le t \le t_1$. Define $\boldsymbol{v} = \boldsymbol{v}(t)$ as the solution of

$$\partial_t \boldsymbol{v} = \mathcal{B}\boldsymbol{v}, \quad \boldsymbol{v}(0) = \boldsymbol{u}(0),$$

then, omitting the α-dependence from the notation, by Duhamel formula we have

$$\begin{aligned} u_i(t_1) - v_i(t_1) &= \int_0^{t_1} \sum_{|p|\le n} \mathcal{U}_{ip}(s,t_1)(\mathrm{d}\xi_{1,p}(s) + \xi_{2,p}\,\mathrm{d}s) \\ &= \int_0^{t_1} \sum_{|p|\le n^{\omega_A}} \mathcal{U}_{ip}(s,t_1)\,\mathrm{d}\xi_{1,p}(s) \\ &\quad + \int_0^{t_1} \sum_{n^{\omega_A}<|p|\le n} \mathcal{U}_{ip}(s,t_1)(\mathrm{d}\xi_{1,p}(s) + \xi_{2,p}\,\mathrm{d}s). \end{aligned} \quad (6.233)$$

In the remainder of this section we focus on the estimate of the rhs. of (6.233) for $|i| \le n^{\omega_A}/2$. Note that $\mathrm{d}\xi_{1,p}$ in (6.233) is a new term compared with [54, Eq. (3.84)]. In the remainder of this section we focus on its estimate, whilst $\xi_{2,p}$ is estimated exactly as in [54, Eq. (3.84)–(3.85)]. The term $\mathrm{d}\xi_{1,p}$ for $|p| \le n^{\omega_A}$ is estimated similarly as the term $(A_N\,\mathrm{d}B_i)/\sqrt{N}$ of [53, Eq. (4.25)] in [53, Lemma 4.2], using the notation therein.

By (6.187)–(6.188) in Assumption (6.B) and the fact that $\sqrt{2n}\,\mathrm{d}\xi_{1,p} = \mathrm{d}\mathfrak{b}_p^s - \mathrm{d}\mathfrak{b}_p^r$, it follows that the quadratic variation of the first term in the rhs. of the second equality of (6.233) is bounded by

$$n^{-1}\int_0^{t_1} \sum_{|p|,|q|\le n^{\omega_A}} \mathcal{U}_{ip}(s,t_1)\mathcal{U}_{iq}(s,t_1)L_{pq}(s)\,\mathrm{d}s \lesssim \frac{t_1\|\mathcal{U}^*\delta_i\|_1^2}{n^{1+\omega_Q}} \lesssim \frac{t_1}{n^{1+\omega_Q}}. \quad (6.234)$$

Note that in (6.234) we used that the bound $|L_{pq}(t)| \le n^{-\omega_Q}$ holds with very high probability uniformly in $t \ge 0$ when $L_{pq}(t)$ is integrated in time (see e.g. [129, Appendix E]). The rhs. of (6.234) is much smaller than the rigidity scale under the assumption $\omega_1 \ll \omega_Q$ (see (6.206)). Note that in the last inequality we used the contraction of the semigroup $\mathcal{U}$ on ℓ^1 to bound $\|\mathcal{U}^*\delta_i\|_1^2 \le 1$. Then, using Burkholder-Davis-Gundy (BDG) inequality, we conclude that

$$\left| \sup_{0\le t\le t_1} \int_0^t \sum_{|p|\le n^{\omega_A}} \mathcal{U}_{ip}(s,t)\, \mathrm{d}\xi_{1,p}(s) \right| \lesssim \sqrt{\frac{t_1}{n^{1+\omega_Q}}}, \tag{6.235}$$

with very high probability. On the other hand, using Kunita-Watanabe inequality, we bound the quadratic variation of the sum over $|p| > n^{\omega_A}$ of $\mathrm{d}\xi_{1,p}$ in (6.233) as

$$\begin{aligned} \frac{1}{n}\int_0^{t_1} \sum_{\substack{|p|>n^{\omega_A}\\ |q|>n^{\omega_A}}} \mathcal{U}_{ip}(s,t_1)\mathcal{U}_{iq}(s,t_1)\, \mathbf{E}\Big[\Big(\mathrm{d}\mathfrak{b}_p^s(s) - \mathrm{d}\mathfrak{b}_p^r(s)\Big)\Big(\mathrm{d}\mathfrak{b}_q^s(s) - \mathrm{d}\mathfrak{b}_q^r(s)\Big)\,\Big|\, \mathcal{F}_t\Big] \\ \le 4n^{-1}\int_0^{t_1}\left(\sum_{n^{\omega_A}<|p|\le n} \mathcal{U}_{ip}(s,t_1)\right)^2 \mathrm{d}s \le n^{-D}, \end{aligned} \tag{6.236}$$

for any $D > 0$ with very high probability, by finite speed of propagation (6.215) since $|i| \le n^{\omega_A/2}$ and $|p| > n^{\omega_A}$. We conclude a very high probability bound for the $\mathrm{d}\xi_{1,p}$-term in the last line of (6.233) using BDG inequality as in (6.235). This concludes the bound of the new term $\mathrm{d}\boldsymbol{\xi}_1$.

The remainder of the proof of Proposition 6.7.13 proceeds exactly in the same way of [54, Eq. (3.86)–(3.99)], hence we omit it. Since $t_f = t_0 + t_1$, choosing $\omega = \omega_1/10$, $\widehat{\omega} \le \omega/10$, the above computations conclude the proof of Proposition 6.7.13. □

6.A Proof of Lemma 6.4.9

In order to prove Lemma 6.4.9 we have to compute

$$\frac{2}{\pi^2}\int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 \partial_1\overline{\partial}_1 f(z_1)\partial_2\overline{\partial}_2\overline{g(z_2)}\Theta(z_1,z_2) \tag{6.237}$$

for compactly supported smooth functions f, g. We recall that

$$\begin{aligned} \Theta(z_1,z_2) &= \Xi(z_1,z_2) + \Lambda(z_1,z_2), \qquad \Lambda(z_1,z_2) := -\frac{1}{2}\log|1 - z_1\overline{z}_2|^2 \mathbf{1}(|z_1|,|z_2|>1), \\ \Xi(z_1,z_2) &:= -\frac{1}{2}\log|z_1 - z_2|^2[1 - \mathbf{1}(|z_1|,|z_2|>1)] + \frac{1}{2}\log|z_1|^2\mathbf{1}(|z_1|\ge 1) \\ &\qquad + \frac{1}{2}\log|z_2|^2\mathbf{1}(|z_2|\ge 1). \end{aligned} \tag{6.238}$$

In order to compute (6.237) we will perform integration by parts twice. For this purpose we split the integral in (6.237) for $\Xi(z_1,z_2)$ into the regimes $|z_1 - z_2| \ge \epsilon$ and its complement, and the integral of $\Lambda(z_1, z_2)$ into the regimes $|1 - z_1\overline{z}_2| \geq \epsilon$ and its complement. We decided to perform two different cut-offs for Ξ and Λ as a consequence of the different kind of singularity of the logarithms in their definition. By the explicit definitions in (6.238), it is easy to see that the integrals in the regimes $|z_1 - z_2| \leq \epsilon$, $|1 - z_1\overline{z}_2| \leq \epsilon$ go to zero as $\epsilon \to 0$, hence we have

$$\begin{aligned}2\mathcal{I} &:= \frac{2}{\pi^2}\int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 \partial_1\overline{\partial}_1 f(z_1)\partial_2\overline{\partial}_2\overline{g(z_2)}\Theta(z_1, z_2) \\ &= \lim_{\epsilon\to 0}\frac{2}{\pi^2}\int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 \partial_1\overline{\partial}_1 f(z_1)\partial_2\overline{\partial}_2\overline{g(z_2)} \\ &\qquad\qquad \times \Big[\Xi(z_1,z_2)\mathbf{1}(|z_1 - z_2| \geq \epsilon) + \Lambda(z_1,z_2)\mathbf{1}(|1 - z_1\overline{z}_2| \geq \epsilon)\Big].\end{aligned} \tag{6.239}$$

In order to prove Lemma 6.4.9 we write the l.h.s. of (6.239) as $\mathcal{I} + \mathcal{I}$ so that in the first integral we perform integration by parts with respect to $\partial_1, \overline{\partial}_2$ and in the second one with respect to $\overline{\partial}_1, \partial_2$. This split is motivated by the fact that

$$\overline{\partial}\overline{g}\partial f + \partial\overline{g}\overline{\partial} f = \frac{1}{2}\langle \nabla g, \nabla f\rangle,$$

which is the first term in the l.h.s. of (6.62) in Lemma 6.4.9. From now on we focus only on the integral for which we perform integration by parts with respect to $\partial_1, \overline{\partial}_2$. The computations for the other integral are exactly the same. It is well known that the distributional Laplacian of $\log|z_1 - z_2|$ is 2π the delta function in $z_1 = z_2$, more precisely, we have that

$$-\partial_1\partial_2 \log|z_1 - z_2| \,\mathrm{d}^2 z_1\,\mathrm{d}^2 z_2 = \frac{\pi}{2}\delta(z_1 - z_2), \tag{6.240}$$

in the sense of distributions. Hence, in the remainder of this section we focus on the computation of the integral of $\Lambda(z_1, z_2)$ and omit the ϵ-regularisation in the integral of Ξ.

Performing integration by parts in $\mathcal{I}$, which is defined in (6.241), with respect to $\partial_1, \overline{\partial}_2$ we get

$$\begin{aligned}&\lim_{\epsilon\to 0}\frac{1}{\pi^2}\int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 \partial_1\overline{\partial}_1 f(z_1)\partial_2\overline{\partial}_2\overline{g(z_2)}\Big[\Xi(z_1,z_2) + \Lambda(z_1,z_2)\mathbf{1}(|1 - z_1\overline{z}_2| \geq \epsilon)\Big] \\ &\quad= \lim_{\epsilon\to 0}\frac{1}{\pi^2}\int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 \overline{\partial}_1 f(z_1)\partial_2\overline{g(z_2)}\Big[\partial_1\overline{\partial}_2\Xi(z_1,z_2) + \partial_1\overline{\partial}_2\Lambda(z_1,z_2)\mathbf{1}(|1 - z_1\overline{z}_2| \geq \epsilon)\Big] \\ &\qquad+ \lim_{\epsilon\to 0} -\frac{\mathrm{i}}{2\pi^2}\int_{\mathbf{C}}\int_{\mathbf{C}} \mathrm{d}^2 z_2 \overline{\partial}_1 f\Big[\overline{\partial}_2\partial_2\overline{g}\Lambda\mathbf{1}(|1 - z_1\overline{z}_2| = \epsilon)\,\mathrm{d}\overline{z}_1 - \partial_2\overline{g}\partial_1\Lambda\mathbf{1}(|1 - z_1\overline{z}_2| = \epsilon)\,\mathrm{d}z_2\Big] \\ &\quad=: \lim_{\epsilon\to 0}\big[J_{1,\epsilon} + J_{2,\epsilon}\big].\end{aligned} \tag{6.241}$$

where in the fourth line we used Stokes theorem written symbolically in the form

$$\partial_z\mathbf{1}(|z - z_2| \geq \epsilon)\,\mathrm{d}^2 z = \frac{\mathrm{i}}{2}\mathbf{1}(|z - z_2| = \epsilon)\,\mathrm{d}\overline{z} \tag{6.242}$$

for any fixed z_2. We remark that (6.242) is understood in the sense of distributions, i.e. the equality holds when tested again smooth compactly supported test functions f, i.e.

$$-\int_{\mathbf{C}} \partial_z f(z)\mathbf{1}(|z - z_2| \geq \epsilon)\,\mathrm{d}^2 z = \frac{\mathrm{i}}{2}\int_{|z - z_2| = \epsilon} f(z)\,\mathrm{d}\overline{z}.$$

Moreover, with a slight abuse of notation in (6.241)–(6.242) by $\mathbf{1}(|z-z_2|=\epsilon)\,\mathrm{d}\overline{z}$ we denoted the clock-wise contour integral over the circle of radius ϵ around z_2. We use the notation above in the remainder of this section.

The second derivative (in the sense of the distributions) of $\Xi(z_1,z_2)$ in (6.241), using (6.240), is given by

$$\begin{aligned}
&\partial_1\overline{\partial}_2\Xi\,\mathrm{d}^2z_1\,\mathrm{d}^2z_2\\
&=\frac{\pi}{2}\delta(z_1-z_2)[1-\mathbf{1}(|z_1|,|z_2|>1)]\,\mathrm{d}^2z_1\,\mathrm{d}^2z_2-\frac{1}{8}\log|z_1-z_2|^2\mathbf{1}(|z_1|=1)\,\mathrm{d}\overline{z}_1\mathbf{1}(|z_2|=1)\,\mathrm{d}z_2\\
&\quad+\frac{\mathrm{i}}{4}\frac{1}{z_1-z_2}\mathbf{1}(|z_1|>1)\,\mathrm{d}^2z_1\mathbf{1}(|z_2|=1)\,\mathrm{d}z_2-\frac{\mathrm{i}}{4}\frac{1}{\overline{z}_1-\overline{z}_2}\mathbf{1}(|z_2|>1)\,\mathrm{d}^2z_2\mathbf{1}(|z_1|=1)\,\mathrm{d}\overline{z}_1,
\end{aligned} \tag{6.243}$$

whilst the second derivative of $\Lambda(z_1,z_2)$ by

$$\begin{aligned}
&\partial_1\overline{\partial}_2\Lambda\,\mathrm{d}^2z_1\,\mathrm{d}^2z_2\\
&=\frac{1}{2(1-z_1\overline{z}_2)^2}\mathbf{1}(|z_1|,|z_2|>1)\,\mathrm{d}^2z_1\,\mathrm{d}^2z_2+\frac{1}{8}\log|1-z_1\overline{z}_2|\mathbf{1}(|z_1|=1)\,\mathrm{d}\overline{z}_1\mathbf{1}(|z_2|=1)\,\mathrm{d}z_2\\
&\quad+\frac{\mathrm{i}}{4}\frac{\overline{z_2}}{1-z_1\overline{z}_2}\mathbf{1}(|z_1|>1)\,\mathrm{d}^2z_1\mathbf{1}(|z_2|=1)\,\mathrm{d}z_2+\frac{\mathrm{i}}{4}\frac{z_1}{1-z_1\overline{z}_2}\mathbf{1}(|z_2|>1)\,\mathrm{d}^2z_2\mathbf{1}(|z_1|=1)\,\mathrm{d}\overline{z}_1.
\end{aligned} \tag{6.244}$$

Note that

$$\begin{aligned}
\partial_1\overline{\partial}_2(\Xi+\Lambda)\,\mathrm{d}^2z_1\,\mathrm{d}^2z_2&=\frac{\pi}{2}\delta(z_1-z_2)\mathbf{1}(|z_1|,|z_2|\le 1)\,\mathrm{d}^2z_1\,\mathrm{d}^2z_2\\
&\quad+\frac{1}{2(1-z_1\overline{z}_2)^2}\mathbf{1}(|z_1|,|z_2|>1)\,\mathrm{d}^2z_1\,\mathrm{d}^2z_2,
\end{aligned}$$

hence, by (6.243)–(6.244) we conclude that

$$\lim_{\epsilon\to 0}J_{1,\epsilon}=\frac{1}{2\pi}\int_{\mathbf{D}}\overline{\partial}f\partial\overline{g}\,\mathrm{d}^2z+\lim_{\epsilon\to 0}\frac{1}{2\pi}\int_{|z_1|\ge 1}\mathrm{d}^2z_1\int_{|z_2|\ge 1}\mathrm{d}^2z_2\frac{\overline{\partial}_1f(z_1)\partial_2g(z_2)}{(1-z_1\overline{z}_2)^2}\mathbf{1}(|1-z_1\overline{z}_2|\ge\epsilon). \tag{6.245}$$

On the other hand, the integration by parts with respect to $\overline{\partial}_1,\partial_2$ gives

$$\frac{1}{2\pi}\int_{\mathbf{D}}\overline{\partial}f\partial\overline{g}\,\mathrm{d}^2z+\lim_{\epsilon\to 0}\frac{1}{2\pi}\int_{|z_1|\ge 1}\mathrm{d}^2z_1\int_{|z_2|\ge 1}\mathrm{d}^2z_2\frac{\overline{\partial}_1f(z_1)\partial_2g(z_2)}{(1-z_1\overline{z}_2)^2}\mathbf{1}(|1-z_1\overline{z}_2|\ge\epsilon). \tag{6.246}$$

Hence, summing (6.245)–(6.246) we get exactly the r.h.s. of (6.62) using that

$$\frac{1}{2\pi}\int_{\mathbf{D}}[\overline{\partial}\overline{g}\partial f+\partial\overline{g}\overline{\partial}f]\,\mathrm{d}^2z=\frac{1}{4\pi}\int_{\mathbf{D}}\langle\nabla g,\nabla f\rangle\,\mathrm{d}^2z.$$

In order to conclude the proof of Lemma 6.4.9 we prove that $|J_{2,\epsilon}|\to 0$ as $\epsilon\to 0$ in Lemma 6.A.1 and that the limit in the r.h.s. of (6.245) exists in Lemma 6.A.2.

Lemma 6.A.1. *Let $J_{2,\epsilon}$ be defined in* (6.241), *then*

$$\lim_{\epsilon\to 0}|J_{2,\epsilon}|=0. \tag{6.247}$$

Proof. For the first integral in $J_{2,\epsilon}$, using the parametrization $z_2 = r_2 e^{\mathrm{i}\theta_2}$ and $z_1 = (1 + \epsilon e^{\mathrm{i}\theta_1})/\overline{z}_2$, for any fixed z_2, we get

$$\left| \int_1^\infty \mathrm{d}r_2 \int_0^{2\pi} \mathrm{d}\theta_1 \int_0^{2\pi} \mathrm{d}\theta_2 \, \epsilon e^{\mathrm{i}(\theta_1+\theta_2)} \overline{\partial}_1 f\left(r_2^{-1} e^{\mathrm{i}\theta_2}[1+\epsilon e^{\mathrm{i}\theta_1}]\right) \overline{\partial}_2 \partial_2 g(r_2 e^{\mathrm{i}\theta_2}) \log \epsilon \right| \lesssim \epsilon \log \epsilon, \tag{6.248}$$

where we used that $\|\overline{\partial}_1 f\|_{L^\infty(\mathbf{C})}, \|\overline{\partial}_2\partial_2 g\|_{L^1(\mathbf{C})} \lesssim 1$ as a consequence of $f, g \in H_0^{2+\delta}(\Omega)$, for an open set $\Omega \subset \mathbf{C}$ such that $\overline{\mathbf{D}} \subset \Omega$.

Furthermore, using the parametrizations $z_1 = r_1 e^{\mathrm{i}\theta_1}$ and $z_2 = (1 + \epsilon e^{\mathrm{i}\theta_2})/\overline{z}_1$ for the second integral in $J_{2,\epsilon}$, we have that

$$\begin{aligned} J_{2,\epsilon} = \Bigg[\int_1^\infty \mathrm{d}r_1 \int_0^{2\pi} \mathrm{d}\theta_1 \int_0^{2\pi} \mathrm{d}\theta_2 \; & \epsilon e^{\mathrm{i}(\theta_1+\theta_2)} \overline{\partial}_1 f(r_1 e^{\mathrm{i}\theta_1}) \partial_2 \overline{g}\left(r_1^{-1} e^{\mathrm{i}\theta_1}[1+\epsilon e^{\mathrm{i}\theta_2}]\right) \\ & \times \frac{1+\epsilon e^{-\mathrm{i}\theta_2}}{\epsilon r_1 e^{-\mathrm{i}\theta_2} e^{\mathrm{i}\theta_1}} \mathbf{1}(|1+\epsilon e^{\mathrm{i}\theta_2}| > r_1) \Bigg] + \mathcal{O}(\epsilon \log \epsilon), \end{aligned} \tag{6.249}$$

where the error term comes from the integral of $\partial_1 \mathbf{1}(|z_1|, |z_2| > 1)$ and the bound in (6.248). Note that $\mathbf{1}(|1+\epsilon e^{\mathrm{i}\theta_1}| > r_1) = 0$ if $r_1 \geq 1 + 2\epsilon$, hence we can bound the first term in J_2 by

$$\int_1^{1+2\epsilon} \mathrm{d}r_1 \int_0^{2\pi} \mathrm{d}\theta_1 \int_0^{2\pi} \mathrm{d}\theta_2 \left| \overline{\partial}_1 f(r_1 e^{\mathrm{i}\theta_1}) \partial_2 \overline{g}\left(r_1^{-1} e^{\mathrm{i}\theta_1}[1+\epsilon e^{\mathrm{i}\theta_2}]\right)\right| \lesssim \epsilon, \tag{6.250}$$

since $\|\partial_2 \overline{g}\|_{L^\infty(\mathbf{C})}, \|\overline{\partial}_1 f\|_{L^1(\mathbf{C})} \lesssim 1$. Hence, we conclude that

$$J_{2,\epsilon} = \mathcal{O}\left(\epsilon + \epsilon \log \epsilon\right).$$

This concludes the proof of (6.247). □

We conclude this section proving the existence of the limit of $J_{1,\epsilon}$ as $\epsilon \to 0$. More precisely, in Lemma 6.A.2 we prove that $J_{1,\epsilon}$ is a Cauchy sequence.

Lemma 6.A.2. *Let $J_{1,\epsilon}$ be defined in* (6.241), *then for any* $0 < \epsilon' \leq \epsilon$ *we have that*

$$|J_{1,\epsilon} - J_{1,\epsilon'}| \lesssim \epsilon^\delta, \tag{6.251}$$

for some $\delta > 0$.

Proof. We only consider the integral with the second derivative of Λ. We dealt with the integral of the second derivative of $\Xi(z_1, z_2)$ already in (6.240). Define

$$I_\epsilon := \frac{1}{\pi^2} \int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 F(z_1, z_2) \Big[\partial_2 \overline{\partial}_1 \Lambda(z_1, z_2) \mathbf{1}(|1 - z_1 \overline{z}_2| \geq \epsilon) \Big], \tag{6.252}$$

where $F(z_1, z_2) := \overline{\partial}_1 f(z_1) \partial_2 \overline{g(z_2)}$ is a δ-Hölder continuous function. Then, for any $0 < \epsilon' < \epsilon$, using the change of variables $z_2 = r_2 e^{\mathrm{i}\theta_2}$ and $z_1 = (1 + r_1 e^{\mathrm{i}\theta_1})/\overline{z}_2$, we write

$$\begin{aligned} I_{\epsilon'} - I_\epsilon = \frac{1}{\pi^2} \int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 & (F(z_1, z_2) - F(\overline{z}_2^{-1}, z_2)) \\ & \times \Big[\partial_2 \overline{\partial}_1 \Lambda(z_1, z_2) \mathbf{1}(\epsilon \geq |1 - z_1 \overline{z}_2| \geq \epsilon') \Big] \\ + \frac{1}{\pi} \int_1^\infty \mathrm{d}r_2 \int_0^{2\pi} \mathrm{d}\theta_2 \int_0^{2\pi} \mathrm{d}\theta_1 & \int_{\epsilon'}^{\epsilon} \mathrm{d}r_1 \, F(r_2^{-1} e^{-\mathrm{i}\theta_2}, r_2 e^{\mathrm{i}\theta_2}) \frac{e^{2\mathrm{i}\theta_1}}{r_1 r_2}. \end{aligned} \tag{6.253}$$

Note that the integral in the second line of (6.253) is exactly zero since $e^{2\mathrm{i}\theta_1}$ the only term which depends on θ_1. On the other hand, we can bound the first integral in (6.253) by $\epsilon^{2\delta}$, with δ the Hölder exponent of F, using the fact that

$$\left|F(z_1,z_2)-F(\overline{z}_2^{-1},z_2)\right| \le \left|\frac{1}{\overline{z}_2}+\frac{r_1e^{\mathrm{i}\theta_1}}{\overline{z}_2}-\frac{1}{\overline{z}_2}\right|^{2\delta} \lesssim \left(\frac{r_1}{r_2}\right)^{2\delta}.$$

This concludes the proof of this lemma. □

6.B Derivation of the DBM for the eigenvalues of H^z

Let X be an $n\times n$ complex random matrix, let H^z be the Hermitisation of $X-z$ defined in (6.142), and define $Y^z := X-z$. We recall that $\{\lambda_i^z,-\lambda_i^z\}_{i=1}^n$ are the eigenvalues of H^z, and $\{\boldsymbol{w}_i^z,\boldsymbol{w}_{-i}^z\}_{i=1}^n$ are the corresponding orthonormal eigenvectors, i.e. for any $i,j\in[n]$ we have

$$H^z\boldsymbol{w}_{\pm i}^z=\pm\lambda_i^z, \qquad (\boldsymbol{w}_i^z)^*\boldsymbol{w}_j^z=\delta_{i,j}, \qquad (\boldsymbol{w}_i^z)^*\boldsymbol{w}_{-j}^z=0, \tag{6.254}$$

for any $i,j\in[n]$. For simplicity in the following derivation we assume that the eigenvalues are all distinct. In particular, for any $i\in[n]$, by the block structure of H^z it follows that

$$\boldsymbol{w}_{\pm i}^z=(\boldsymbol{u}_i^z,\pm\boldsymbol{v}_i^z), \qquad Y^z\boldsymbol{v}_i^z=\lambda_i^z\boldsymbol{u}_i^z, \qquad (Y^z)^*\boldsymbol{u}_i^z=\lambda_i^z\boldsymbol{v}_i^z. \tag{6.255}$$

Moreover, since $\{\boldsymbol{w}_{\pm i}^z\}_{i=1}^n$ is an orthonormal base, we conclude that

$$(\boldsymbol{u}_i^z)^*\boldsymbol{u}_i^z=(\boldsymbol{v}_i^z)^*\boldsymbol{v}_i^z=\frac{1}{2}. \tag{6.256}$$

In the following, for any fixed entry x_{ab} of X, we will use the notation

$$\dot{f}=\frac{\partial f}{\partial x_{ab}} \quad \text{or} \quad \dot{f}=\frac{\partial f}{\partial \overline{x}_{ab}}, \tag{6.257}$$

where $f=f(X)$ is a function of the matrix X. Then, we consider the flow

$$\mathrm{d}X_t=\frac{\mathrm{d}B_t}{\sqrt{n}}, \quad X_0=X, \tag{6.258}$$

where B_t is a matrix valued complex standard Brownian motion.

From now on we only consider positive indices $1\le i\le n$. We may also drop the z and t dependence to make our notation easier. For any $i,j\in[n]$, differentiating (6.254) we get

$$\dot{H}\boldsymbol{w}_i+H\dot{\boldsymbol{w}}_i=\dot{\lambda}_i\boldsymbol{w}_i+\lambda_i\dot{\boldsymbol{w}}_i, \tag{6.259}$$

$$\dot{\boldsymbol{w}}_i^*\boldsymbol{w}_j+\boldsymbol{w}_i^*\dot{\boldsymbol{w}}_j=0, \tag{6.260}$$

$$\boldsymbol{w}_i^*\dot{\boldsymbol{w}}_i+\dot{\boldsymbol{w}}_i^*\boldsymbol{w}_i=0. \tag{6.261}$$

Note that (6.261) implies that $\Re[\boldsymbol{w}_i^*\dot{\boldsymbol{w}}_i]=0$. Hence, since the eigenvectors are defined modulo a phase, we can choose eigenvectors such that $\Im[\boldsymbol{w}_i^*\dot{\boldsymbol{w}}_i]=0$ for any $t\ge 0$. Then, multiplying (6.259) by $\boldsymbol{w}_i^*$ we conclude that

$$\dot{\lambda}_i=\boldsymbol{u}_i^*\dot{Y}\boldsymbol{v}_i+\boldsymbol{v}_i^*\dot{Y}^*\boldsymbol{u}_i. \tag{6.262}$$

Moreover, multiplying (6.259) by $\boldsymbol{w}_j^*$, with $j \neq i$, and by $\boldsymbol{w}_{-j}*$, we get

$$(\lambda_i - \lambda_j)\boldsymbol{w}_j^*\dot{\boldsymbol{w}}_i = \boldsymbol{w}_j^*\dot{H}\boldsymbol{w}_i, \qquad (\lambda_i + \lambda_j)\boldsymbol{w}_{-j}^*\dot{\boldsymbol{w}}_i = \boldsymbol{w}_{-j}^*\dot{H}\boldsymbol{w}_i, \tag{6.263}$$

respectively. By (6.260)–(6.261) it follows that

$$\dot{\boldsymbol{w}}_i = \sum_{j\neq i}(\boldsymbol{w}_j^*\dot{\boldsymbol{w}}_i)\boldsymbol{w}_j + \sum_j(\boldsymbol{w}_{-j}^*\dot{\boldsymbol{w}}_i)\boldsymbol{w}_{-j}, \tag{6.264}$$

hence by (6.263) we conclude

$$\dot{\boldsymbol{w}}_i = \sum_{j\neq i}\frac{\boldsymbol{v}_j^*\dot{Y}^*\boldsymbol{u}_i + \boldsymbol{u}_j^*\dot{Y}\boldsymbol{v}_i}{\lambda_i - \lambda_j}\boldsymbol{w}_j + \sum_j\frac{\boldsymbol{u}_j^*\dot{Y}\boldsymbol{v}_i - \boldsymbol{v}_j^*\dot{Y}^*\boldsymbol{u}_i}{\lambda_i + \lambda_j}\boldsymbol{w}_{-j}. \tag{6.265}$$

By Ito's formula we have that

$$\mathrm{d}\lambda_i = \sum_{ab}\frac{\partial\lambda_i}{\partial x_{ab}}\,\mathrm{d}x_{ab} + \frac{\partial\lambda_i}{\partial\overline{x_{ab}}}\,\mathrm{d}\overline{x_{ab}} + \frac{1}{2}\sum_{ab}\sum_{kl}\frac{\partial^2\lambda_i}{\partial x_{ab}\partial\overline{x_{kl}}}\,\mathrm{d}x_{ab}\,\mathrm{d}\overline{x_{kl}} + \frac{\partial^2\lambda_i}{\partial\overline{x_{ab}}\partial x_{kl}}\,\mathrm{d}\overline{x_{ab}}\,\mathrm{d}x_{kl}. \tag{6.266}$$

Note that in (6.266) we used that $\mathrm{d}x_{ab}\,\mathrm{d}x_{ab} = \mathrm{d}\overline{x_{kl}}\,\mathrm{d}\overline{x_{kl}} = 0$. Then by (6.262)–(6.265) it follows that

$$\frac{\partial\lambda_i}{\partial x_{ab}} = u_i(a)^*v_i(b), \qquad \frac{\partial\lambda_i}{\partial\overline{x_{ab}}} = v_i(b)^*u_i(a), \tag{6.267}$$

and that

$$\frac{\partial w_i}{\partial x_{ab}}(k) = \sum_{j\neq i}\left[\frac{u_j^*(a)v_i(b)}{\lambda_i - \lambda_j}w_j(k) + \frac{u_j^*(a)v_i(b)}{\lambda_i + \lambda_j}w_{-j}(k)\right] + \frac{u_i(a)^*v_i(b)}{2\lambda_i}w_{-i}(k), \tag{6.268}$$

$$\frac{\partial w_i}{\partial\overline{x_{ab}}}(k) = \sum_{j\neq i}\left[\frac{v_j^*(b)u_i(a)}{\lambda_i - \lambda_j}w_j(k) - \frac{v_j^*(b)u_i(a)}{\lambda_i + \lambda_j}w_{-j}(k)\right] - \frac{v_i(b)^*u_i(a)}{2\lambda_i}w_{-i}(k). \tag{6.269}$$

Next, we compute

$$\begin{aligned}\frac{\partial^2\lambda_i}{\partial x_{ab}\partial\overline{x_{kl}}} &= \frac{\partial v_i^*}{\partial x_{ab}}(l)u_i(k) + v_i(l)^*\frac{\partial u_i}{\partial x_{ab}}(k)\\ &= \sum_{j\neq i}\left[\frac{v_j(b)u_i^*(a)}{\lambda_i - \lambda_j}v_j(l)^*u_i(k) + \frac{v_j(b)u_i(a)^*}{\lambda_i + \lambda_j}v_j(l)^*u_i(k)\right] + \frac{v_i(b)u_i(a)^*}{2\lambda_i}v_i(l)^*u_i(k)\\ &\quad + \sum_{j\neq i}\left[\frac{u_i^*(a)v_i(b)}{\lambda_i - \lambda_j}v_i(l)^*u_j(k) + \frac{u_j^*(a)v_i(b)}{\lambda_i + \lambda_j}v_i(l)^*u_j(k)\right] + \frac{u_i(a)^*v_i(b)}{2\lambda_i}v_\alpha(l)^*u_i(k).\end{aligned} \tag{6.270}$$

Finally, combining (6.258), (6.267), (6.266) and (6.270), we conclude (cf. [87, Eq. (5.8)])

$$\mathrm{d}\lambda_i^z = \frac{\mathrm{d}b_i^z}{\sqrt{2n}} + \frac{1}{2n}\sum_{j\neq i}\left[\frac{1}{\lambda_i^z - \lambda_j^z} + \frac{1}{\lambda_i^z + \lambda_j^z}\right]\mathrm{d}t + \frac{\mathrm{d}t}{4n\lambda_i}, \tag{6.271}$$

where we defined

$$\mathrm{d}b_i^z := \sqrt{2}(\mathrm{d}B_{ii}^z + \mathrm{d}\overline{B_{ii}^z}), \quad \mathrm{d}B_{ij}^z := \sum_{ab} \overline{u_i^z(a)}\, \mathrm{d}B_{ab} v_j^z(b), \tag{6.272}$$

where B_t is the matrix values Brownian motion in (6.258). In particular, b_i^z is a standard real Brownian motion, indeed

$$\begin{aligned}
\mathbf{E}(B_{ii}^z + \overline{B_{ii}^z})(B_{ii}^z + \overline{B_{ii}^z})^* &= \mathbf{E}\left(\sum_{ab} \overline{u_i^z(a)} B_{ab} v_i^z(b) + u_i^z(a)\overline{B}_{ab}\overline{v_i^z(b)}\right)^2 \\
&= 2\sum_{abcd} \overline{u_i^z(a)} B_{ab} v_i^z(b) u_i^z(c) \overline{B}_{cd} \overline{v_i^z(d)} \\
&= 2\sum_{abcd} \delta_{ac}\delta_{bd} \overline{u_i^z(a)} v_i^z(b) u_i^z(c) \overline{v_i^z(d)} = \frac{1}{2}.
\end{aligned}$$

Fluctuation around the circular law for random matrices with real entries 7

We extend our recent result [58] on the central limit theorem for the linear eigenvalue statistics of non-Hermitian matrices X with independent, identically distributed complex *entries to the* real *symmetry class. We find that the expectation and variance substantially differ from their complex counterparts, reflecting (i) the special spectral symmetry of real matrices onto the real axis; and (ii) the fact that real i.i.d. matrices have many real eigenvalues. Our result generalizes the previously known special cases where either the test function is analytic [152] or the first four moments of the matrix elements match the real Gaussian [126, 195]. The key element of the proof is the analysis of several weakly dependent Dyson Brownian motions (DBMs). The conceptual novelty of the real case compared with [58] is that the correlation structure of the stochastic differentials in each individual DBM is non-trivial, potentially even jeopardising its well-posedness.*

Published as G. Cipolloni et al., *Fluctuation around the circular law for random matrices with real entries*, preprint (2020), `arXiv:2002.02438`

7.1 Introduction

We consider an ensemble of $n \times n$ random matrices X with *real* i.i.d. entries of zero mean and variance $1/n$; the corresponding model with *complex* entries has been studied in [58]. According to the *circular law* [18, 103, 191] (see also [34]), the density of the eigenvalues $\{\sigma_i\}_{i=1}^n$ of X converges to the uniform distribution on the unit disk. Our main result is that the fluctuation of their linear statistics is Gaussian, i.e.

$$L_n(f) := \sum_{i=1}^{n} f(\sigma_i) - \mathbf{E} \sum_{i=1}^{n} f(\sigma_i) \sim \mathcal{N}(0, V_f) \tag{7.1}$$

converges, as $n \to \infty$, to a centred normal distribution for regular test functions f with at least $2 + \delta$ derivatives. We compute the variance V_f and the next-order deviation of the expectation $\mathbf{E}\sum_{i=1}^{n} f(\sigma_i)$ from the value $\frac{n}{\pi}\int_{|z|\le 1} f(z)$ given by the circular law. As in the complex case, both quantities depend on the fourth cumulant of the single entry distribution of X, but in the real case they also incorporate the spectral symmetry of X onto the real axis. Moreover, the expectation carries additional terms, some of them are concentrated around the real axis; a by-product of the approximately $\sqrt{n}$ real eigenvalues of X. For the Ginibre (Gaussian) case they may be computed from the explicit density [76, 77], but for general distributions they were not known before. As expected, the spectral symmetry essentially enhances V_f by a factor of two compared with the complex case but this effect is modified by an additional term involving the fourth cumulant. Previous works considered either the case of analytic test functions f [151, 152] or the (approximately) Gaussian case, i.e. when X is the real Ginibre ensemble or at least the first four moments of the matrix elements of X match the Ginibre ensemble [126, 195]. In both cases some terms in the unified formulas for the expectation and the variance vanish and thus the combined effect of the spectral symmetry, the eigenvalues on the real axis, and the role of the fourth cumulant was not detectable in these works. We remark that a CLT for polynomial statistics of only the real eigenvalues for real Ginibre matrices was proven in [179].

In [163] the limiting random field $L(f) := \lim_{n\to\infty} L_n(f)$ for complex Ginibre matrices has been identified as a projection of the *Gaussian free field (GFF)* [178]. We extended this interpretation [58] to general complex i.i.d. matrices with non-negative fourth cumulant and obtained a rank-one perturbation of the projected GFF. As a consequence of the CLT in the present paper, we find that in the real case the limiting random field is a version of the same GFF, symmetrised with respect to the real axis, reflecting the fact that complex eigenvalues of real matrices come in pairs of complex conjugates.

In general, proving CLTs for the real symmetry class is considerably harder than for the complex one. The techniques based upon the first four moment matching [126, 195] are insensitive to the symmetry class, hence these results are obtained in parallel for both real and complex ensembles. Beyond this method, however, most results on CLT for non-Hermitian matrices were restricted to the complex case [65, 95, 150, 161, 162, 164], see the introduction of [58] for a detailed history, as well as for references to the analogous CLT problem for Hermitian ensembles and log-gases. The special role that the real axis plays in the spectrum of the real case substantially complicates even the explicit formulas for the Ginibre ensemble both for the density [76] as well as for the k-point correlation functions [35, 102, 121]. Besides the complexity of the explicit formulas, there are several conceptual reasons why the real case is more involved. We now explain them since they directly motivated the new ideas in this paper compared with [58].

In [58] we started with Girko's formula [103] in the form given in [195] that relates the eigenvalues of X with resolvents of a family of $2n\times 2n$ Hermitian matrices

$$H^z := \begin{pmatrix} 0 & X-z \\ X^*-\overline{z} & 0 \end{pmatrix} \tag{7.2}$$

parametrized by $z\in\mathbf{C}$. For any smooth, compactly supported test function f we have

$$\sum_{i=1}^{n} f(\sigma_i) = -\frac{1}{4\pi}\int_{\mathbf{C}} \Delta f(z) \int_0^\infty \Im \operatorname{Tr} G^z(i\eta)\, \mathrm{d}\eta\, \mathrm{d}^2 z, \tag{7.3}$$

where $G^z(w) := (H^z - w)^{-1}$ is the resolvent of H^z. We therefore needed to understand the resolvent $G^z(i\eta)$ along the imaginary axis on all scales $\eta\in(0,\infty)$.

The main contribution to (7.3) comes from the $\eta \sim 1$ *macroscopic* regime, which is handled by proving a multi-dimensional CLT for resolvents with several z and η parameters and computing their expectation and covariance by cumulant expansion. The local laws along the imaginary axis from [11, 13] serve as a basic input (in the current work, however, we need to extend them for spectral parameters w away from the imaginary axis). The core of the argument in the real case is similar to the complex case in [58], however several additional terms have to be computed due to the difference between the real and complex cumulants. By explicit calculations, these additional terms break the rotational symmetry in the z parameter and, unlike in the complex case, the answer is not a function of $|z|$ any more. The *mesoscopic* regime $n^{-1} \ll \eta \ll 1$ is treated together with the macroscopic one; the fact that only the $\eta \sim 1$ regime contributes to (7.3) is revealed *a posteriori* after these calculations.

The scale $\eta \lesssim n^{-1}$ in (7.3) requires a very different treatment since local laws are not applicable any more and individual eigenvalues $0 \le \lambda_1^z \le \lambda_2^z \dots$ of H^z near zero substantially influence the fluctuation of $G^z(i\eta)$ (since H^z has a symmetric spectrum, we consider only positive eigenvalues). The main insight of [58] was that it is sufficient to establish that the small eigenvalues, say, λ_1^z and $\lambda_1^{z'}$, are asymptotically independent if z and z' are relatively far away, say $|z - z'| \ge n^{-1/100}$. This was achieved by exploiting the fast local equilibration mechanism of the *Dyson Brownian motion (DBM)*, which is the stochastic flow of eigenvalues $\boldsymbol{\lambda}^z(t) := \{\lambda_i^z(t)\}$ generated by adding a time-dependent Gaussian (Ginibre) component. The initial condition of this flow was chosen carefully to almost reproduce X after a properly tuned short time. We needed to follow the evolution of $\boldsymbol{\lambda}^z(t)$ for different z parameters simultaneously. These flows are correlated since they are driven by the same random source. We thus needed to study a family of DBMs, parametrized by z, with correlated driving Brownian motions. The correlation structure is given by the *overlap* of the eigenfunctions of H^z and $H^{z'}$. We could show that this overlap is small, hence the Brownian motions are essentially independent, if z and z' are far away. This step required to develop a new type of local law for *products* of resolvent, e.g. for $\operatorname{Tr} G^z(\mathrm{i}\eta)G^{z'}(\mathrm{i}\eta')$ with $\eta, \eta' \sim n^{-1+\epsilon}$. Finally, we trailed the joint evolution of $\boldsymbol{\lambda}^z(t)$ and $\boldsymbol{\lambda}^{z'}(t)$ by their independent Ginibre counterparts, showing that they themselves are asymptotically independent.

We follow the same strategy in the current paper for the real case, but we immediately face with the basic question: how do the low lying eigenvalues of H^z, equivalently the small singular values of $X - z$, behave? We do not need to compute their joint distribution, but we need to approximate them with an appropriate Ginibre ensemble. For *complex* X in [58] the approximating Ginibre ensemble was naturally complex. For *real* X there seem to be two possibilities. The key insight of our current analysis is that the small singular values of $X - z$ behave as those of a *complex* Ginibre matrix even though X is *real*, as long as z is genuinely complex (Theorem 7.2.7). In particular, we prove that the least singular value of $X - z$ belongs to the complex universality class. Moreover, we prove that the small singular values of $X - z_1$ and the ones of $X - z_2$ are asymptotically independent as long as z_1 and z_2 are far from each other.

To explain the origin of this apparent mismatch, we will derive the DBM

$$\mathrm{d}\lambda_i^z = \frac{\mathrm{d}b_i^z}{\sqrt{n}} + \frac{1}{2n}\sum_{j\neq i} \frac{1+\Lambda_{ij}^z}{\lambda_i^z - \lambda_j^z}\,\mathrm{d}t + \dots \tag{7.4}$$

for $\boldsymbol{\lambda}^z(t)$, ignoring some additional terms with negative indices coming from the spectral symmetry of H^z (see (7.133) and (7.253) for the precise equation). The correlations of the driving Brownian motions are given by

$$\mathbf{E}\, \mathrm{d}b_i^z\, \mathrm{d}b_j^{z'} = \frac{1}{2}\big[\Theta_{ij}^{z,z'} + \Theta_{ij}^{z,\overline{z}'}\big]\, \mathrm{d}t \tag{7.5}$$

with overlaps Θ, Λ defined as

$$\Theta_{ij}^{z,z'} := 4\Re\big[\langle \boldsymbol{u}_j^{z'}, \boldsymbol{u}_i^z\rangle \langle \boldsymbol{v}_i^z, \boldsymbol{v}_j^{z'}\rangle\big], \qquad \Lambda_{ij}^z := \Theta_{ij}^{z,\overline{z}}, \tag{7.6}$$

where $(\boldsymbol{u}_i^z, \boldsymbol{v}_i^z) \in \mathbf{C}^{2n}$ is the (normalized) eigenvector of H^z corresponding to the eigenvalue λ_i^z. Note that $\Theta_{ij}^{z,z} = \delta_{i,j}$, and for $j \neq i$ we have that $\Lambda_{ij}^z \approx 0$. Moreover, if z is very close to the real axis, then the eigenvectors of H^z are essentially real and $\Lambda_{ii}^z = \Theta_{ii}^{z,\overline{z}} \approx \Theta_{ii}^{z,z} = 1$. With $z = z'$, this leads to (7.4) being essentially a *real* DBM with $\beta = 1$. (We recall that the parameter $\beta = 1, 2$, customarily indicating the real or complex symmetry class of a random matrix, also expresses the ratio of the coefficient of the repulsion to the strength of the diffusion in the DBM setup.) However, if z and $\bar{z}$ are far away, i.e. z is away from the real axis, then we can show that the overlap $\Lambda^z = \Theta^{z,\overline{z}}$ is small, hence $\Lambda_{ij}^z \approx 0$ for all i, j, including $i = j$. Thus the variance of the driving Brownian motions in (7.5) with $z = z'$ is reduced by a factor of two, rendering (7.4) essentially a *complex* DBM with $\beta = 2$.

The appearance of Λ^z in (7.4) and the second term $\Theta^{z,\overline{z}'}$ in (7.5) is specific to the real symmetry class; they were not present in the complex case [58]. They have three main effects for our analysis. First, they change the symmetry class of the DBM (7.4) as we just explained. Second, due to the symmetry relation $\lambda_{-1}^z = -\lambda_1^z$ and $b_{-1}^z = -b_1^z$, the strength of the level repulsion between λ_1^z and λ_{-1}^z in (7.4) is already critically small even for $\Lambda^z = 0$, see e.g. [54, Appendix A], hence the well-posedness of (7.4) does not follow from standard results on DBM. Third, $\Theta^{z,\overline{z}}$ renders the driving Brownian motions $\boldsymbol{b}^z = \{b_i^z\}$ correlated for different indices i even for the *same* z, since Λ_{ij}^z in general is nonzero. In fact, the vector $\boldsymbol{b}^z$ is even not Gaussian, hence strictly speaking it is only a multidimensional martingale but not a Brownian motion in general. In contrast, $\Theta_{ij}^{z,z} = \delta_{i,j}$ and only the overlaps $\Theta_{ij}^{z,z'}$ for *different* $z \neq z'$ are nontrivial. Thus in the complex case [58], lacking the term $\Theta^{z,\overline{z}}$ in (7.5), the DBM (7.4) for any fixed z was the conventional DBM with independent Brownian motions and parameter $\beta = 2$ (c.f. [58, Eq. (7.15)]) and only the DBMs for *different* z's were mildly correlated. In the real case the correlations are already present within (7.4) for the *same* z due to $\Lambda^z = \Theta^{z,\overline{z}} \neq 0$.

We note that Dyson Brownian motions with nontrivial coefficients in the repulsion term have already been investigated in [53] (see also [55]) in the context of spectral universality of addition of random matrices twisted by Haar unitaries, however the driving Brownian motions were independent. The issue of well-posedness, nevertheless, has already emerged in [53] when the more critical orthogonal group ($\beta = 1$) was considered. The corresponding part of our analysis partly relies on techniques developed in [53]. We have already treated the dependence of Brownian motions for different z's in [58] for the complex case; but the more general dependence structure characteristic to the real case is a new challenge that the current work resolves.

Notations and conventions

We introduce some notations we use throughout the paper. For integers $k \in \mathbf{N}$ we use $[k] := \{1, \dots, k\}$. We write $\mathbf{H}$ for the upper half-plane $\mathbf{H} := \{z \in \mathbf{C} \mid \Im z > 0\}$,

$\mathbf{D} \subset \mathbf{C}$ for the open unit disk, and we use the notation $\mathrm{d}^2 z := 2^{-1}\mathrm{i}(\mathrm{d}z \wedge \mathrm{d}\overline{z})$ for the two dimensional volume form on $\mathbf{C}$. For positive quantities f, g we write $f \lesssim g$ and $f \sim g$ if $f \le Cg$ and $cg \le f \le Cg$, respectively, for some constants $c, C > 0$ which depend only on the *model parameters* appearing in (7.7). For any two positive real numbers $\omega_*, \omega^* \in \mathbf{R}_+$, by $\omega_* \ll \omega^*$ we denote that $\omega_* \le c\omega^*$ for some sufficiently small constant $0 < c \le 1/1000$. We denote vectors by bold-faced lower case Roman letters $\boldsymbol{x}, \boldsymbol{y}, \ldots, \in \mathbf{C}^k$, for some $k \in \mathbf{N}$, and use the notation $\mathrm{d}\boldsymbol{x} := \mathrm{d}x_1 \ldots \mathrm{d}x_k$. Vector and matrix norms, $\|\boldsymbol{x}\|$ and $\|A\|$, indicate the usual Euclidean norm and the corresponding induced matrix norm. For any $k \times k$ matrix A we set $\langle A \rangle := k^{-1} \operatorname{Tr} A$ to denote the normalized trace of A. Moreover, for vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{C}^k$ and matrices $A, B \in \mathbf{C}^{k \times k}$ we define

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum \overline{x_i} y_i, \qquad \langle A, B \rangle := \langle A^* B \rangle = \frac{1}{k} \operatorname{Tr} A^* B.$$

We will use the concept of "event with very high probability" 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 exponent which is independent of n.

7.2 Main results

We consider *real i.i.d. matrices* X, i.e. $n \times n$ matrices whose entries are independent and identically distributed as $x_{ab} \overset{\mathrm{d}}{=} n^{-1/2}\chi$ for some real random variable χ, satisfying the following:

Assumption (7.A). *We assume that* $\mathbf{E}\,\chi = 0$ *and* $\mathbf{E}\,\chi^2 = 1$. *In addition we assume the existence of high moments, i.e. that there exist constants* $C_p > 0$, *for any* $p \in \mathbf{N}$, *such that*

$$\mathbf{E}|\chi|^p \le C_p. \tag{7.7}$$

The *circular law* [18, 20, 33, 34, 101, 103, 105, 154, 191] asserts that the empirical distribution of eigenvalues $\{\sigma_i\}_{i=1}^n$ of a complex i.i.d. matrix X converges to the uniform distribution on the unit disk $\mathbf{D}$, i.e.

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n f(\sigma_i) = \frac{1}{\pi} \int_{\mathbf{D}} f(z)\, \mathrm{d}^2 z, \tag{7.8}$$

with very high probability for any continuous bounded function f. Our main result is a central limit theorem for the centred *linear statistics*

$$L_n(f) := \sum_{i=1}^n f(\sigma_i) - \mathbf{E} \sum_{i=1}^n f(\sigma_i) \tag{7.9}$$

for general real i.i.d. matrices and generic test functions f, complementing the recent central limit theorem [58] for the linear statistics of *complex i.i.d. matrices*. This CLT, formulated in Theorem 7.2.1, and its proof have two corollaries of independent interest that are formulated in Section 7.2.1 and Section 7.2.2.

In order to state the result we introduce some notations. For any function h defined on the boundary of the unit disk $\partial\mathbf{D}$ we define its Fourier transform as

$$\widehat{h}(k) = \frac{1}{2\pi} \int_0^{2\pi} h(e^{\mathrm{i}\theta}) e^{-\mathrm{i}\theta k}\, \mathrm{d}\theta, \qquad k \in \mathbf{Z}. \tag{7.10}$$

For $f, g \in H^{2+\delta}(\Omega)$ for some domain $\Omega \supset \overline{\mathbf{D}}$ we define

$$\begin{aligned} \langle g, f\rangle_{\dot H^{1/2}(\partial\mathbf{D})} &:= \sum_{k\in\mathbf{Z}} |k| \overline{\widehat{g}(k)} \widehat{f}(k), \qquad \|f\|^2_{\dot H^{1/2}(\partial\mathbf{D})} := \langle f, f\rangle_{\dot H^{1/2}(\partial\mathbf{D})}, \\ \langle g, f\rangle_{H^1_0(\mathbf{D})} &:= \langle \nabla g, \nabla f\rangle_{L^2(\mathbf{D})}, \qquad \|f\|^2_{H^1_0(\mathbf{D})} := \langle f, f\rangle_{H^1_0(\mathbf{D})}, \end{aligned} \tag{7.11}$$

where, in a slight abuse of notation, we identified f and g with their restrictions to $\partial\mathbf{D}$. We use the convention that f is extended to $\mathbf{C}$ by setting it equal to zero on Ω^c. Finally, we introduce the projection

$$(P_{\mathrm{sym}} f)(z) := \frac{f(z) + f(\overline{z})}{2}. \tag{7.12}$$

which maps functions on the complex plane to their symmetrisation with respect to the real axis.

Theorem 7.2.1 (Central Limit Theorem for linear statistics). *Let X be a real $n \times n$ i.i.d. matrix satisfying Assumption (7.A) with eigenvalues $\{\sigma_i\}_{i=1}^n$, and denote the fourth* cumulant[1] *of χ by $\kappa_4 := \mathbf{E}\,\chi^4 - 3$. Fix $\delta > 0$, let $\Omega \subset \mathbf{C}$ be open and such that $\overline{\mathbf{D}} \subset \Omega$. Then, for complex-valued test functions $f \in H^{2+\delta}(\Omega)$, the centred linear statistics $L_n(f)$, defined in* (7.9)*, converge*

$$L_n(f) \Longrightarrow L(f),$$

to complex Gaussian random variables $L(f)$ with expectation $\mathbf{E}\,L(f) = 0$ and variance $\mathbf{E}|L(f)|^2 = C(f,f) =: V_f$ and $\mathbf{E}\,L(f)^2 = C(\overline{f}, f)$, where

$$\begin{aligned} C(g,f) :=& \frac{1}{2\pi} \langle \nabla P_{\mathrm{sym}} g, \nabla P_{\mathrm{sym}} f\rangle_{L^2(\mathbf{D})} + \langle P_{\mathrm{sym}} g, P_{\mathrm{sym}} f\rangle_{\dot H^{1/2}(\partial\mathbf{D})} \\ &+ \kappa_4 \left(\frac{1}{\pi}\int_{\mathbf{D}} \overline{g(z)}\,\mathrm{d}^2 z - \frac{1}{2\pi}\int_0^{2\pi} \overline{g(e^{\mathrm{i}\theta})}\,\mathrm{d}\theta \right) \left(\frac{1}{\pi}\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z - \frac{1}{2\pi}\int_0^{2\pi} f(e^{\mathrm{i}\theta})\,\mathrm{d}\theta \right). \end{aligned} \tag{7.13}$$

For the k-th moments we have an effective convergence rate of

$$\mathbf{E}\,L_n(f)^k \overline{L_n(f)}^l = \mathbf{E}\,L(f)^k \overline{L(f)}^l + \mathcal{O}\Big(n^{-c(k+l)}\Big)$$

for some constant $c(k+l) > 0$. Moreover, the expectation in (7.9) *is given by*

$$\begin{aligned} &\mathbf{E}\sum_{i=1}^n f(\sigma_i) = E(f) + \mathcal{O}(n^{-c}) \\ &E(f) := \frac{n}{\pi}\int_{\mathbf{D}} f(z)\,\mathrm{d}^2 z + \frac{1}{4\pi}\int_{\mathbf{D}} \frac{f(\Re z) - f(z)}{(\Im z)^2}\,\mathrm{d}^2 z - \frac{\kappa_4}{\pi}\int_{\mathbf{D}} f(z)(2|z|^2 - 1)\,\mathrm{d}^2 z \\ &\qquad - \frac{1}{2\pi}\int_0^{2\pi} f(e^{\mathrm{i}\theta})\,\mathrm{d}\theta + \frac{1}{2\pi}\int_{-1}^1 \frac{f(x)}{\sqrt{1-x^2}}\,\mathrm{d}x + \frac{f(1)+f(-1)}{4} \end{aligned} \tag{7.14}$$

for some small constant $c > 0$.

Remark 7.2.2.

[1]Note that in the real case the fourth cumulant is given by $\kappa_4 = \kappa(\chi,\chi,\chi,\chi) = \mathbf{E}\,\chi^4 - 3$, while in the complex case [58] the relevant fourth cumulant was given by $\kappa(\chi,\chi,\overline{\chi},\overline{\chi}) = \mathbf{E}|\chi|^4 - 2$.

(i) Both expectation $E(f)$ and covariance $C(g,f)$ only depend on the symmetrised functions $P_{\mathrm{sym}}f$ and $P_{\mathrm{sym}}g$. Indeed, $E(f) = E(P_{\mathrm{sym}}f)$, and the coefficient of κ_4 in (7.13) *can also be written as an integral over $P_{\mathrm{sym}}f$ and $P_{\mathrm{sym}}g$.*

(ii) By polarisation, a multivariate central limit theorem as in [58, Corollary 2.4] follows immediately and any mixed k-th moments have an effective convergence rate of order $n^{-c(k)}$.

(iii) The variance $V_f = \mathbf{E}|L(f)|^2$ in Theorem 7.2.1 is strictly positive whenever f is not constant on the unit disk (see [58, Remark 2.3]).

Remark 7.2.3 (Comparison with [126] and [152])**.**

(i) The central limit theorem [126, Theorem 2] is a special case of Theorem 7.2.1. Indeed, [126, Theorem 2] implies that for real i.i.d. matrices with entries matching the real Ginibre ensemble to the fourth moment, and real-valued smooth test functions f compactly supported within the upper half of the unit disk $L_n(f)$ converge to a real Gaussian of variance

$$\frac{1}{4\pi}\langle \nabla f, \nabla f\rangle_{L^2(\mathbf{D})} = \frac{1}{2\pi}\langle \nabla P_{\mathrm{sym}}f, \nabla P_{\mathrm{sym}}f\rangle_{L^2(\mathbf{D})}, \tag{7.15}$$

where we used that $z \mapsto f(z)$ and $z \mapsto f(\overline{z})$ are assumed to have disjoint support. Due to the moment matching assumption, $\kappa_4 = 0$ in the setting of [126].

(ii) The central limit theorem [152, Corollary 2.6] is also a special case of Theorem 7.2.1. Indeed, [152, Corollary 2.6] implies that for real i.i.d. matrices and test functions f which are analytic in a neighbourhood of the unit disk and satisfy $P_{\mathrm{sym}}f\colon \overline{\mathbf{D}} \to \mathbf{R}$ the linear statistics $L_n(f)$ converge to a Gaussian of variance

$$\begin{aligned}\frac{1}{\pi}\int_{\mathbf{D}} |\partial_z f(z)|^2 \,\mathrm{d}^2 z &= \frac{1}{4\pi}\langle \nabla f, \nabla f\rangle_{L^2(\mathbf{D})} + \frac{1}{2}\langle f, f\rangle_{\dot{H}^{1/2}(\partial\mathbf{D})}\\ &= \frac{1}{2\pi}\langle \nabla P_{\mathrm{sym}}f, \nabla P_{\mathrm{sym}}f\rangle_{L^2(\mathbf{D})} + \langle P_{\mathrm{sym}}f, P_{\mathrm{sym}}f\rangle_{\dot{H}^{1/2}(\partial\mathbf{D})}.\end{aligned}$$

Here in the first step we used the analyticity of f (see [58, Eq. (2.11)]), and in the second step we used that $\langle (\nabla f)(z), (\nabla f(\bar{\cdot}))(z)\rangle = 0$ and that $\widehat{f}(k) = 0$ for $k < 0$ while $\widehat{f(\bar{\cdot})}(k) = 0$ for $k > 0$ by analyticity. We thus arrived at (7.13)*, since the coefficient of κ_4 in* (7.13) *vanishes also by analyticity of f in the setting of [152].*

Remark 7.2.4 (Comparison with the complex case)**.** *We remark that the limiting variance in the case of complex i.i.d. matrices, as studied in [58], is generally different from the real case. In the complex case $L_n(f)$ converges to a complex Gaussian with variance*

$$\begin{aligned}V_f^{(\mathbf{C})} &= V_f^{(\mathbf{C},1)} + \kappa_4 V_f^{(\mathbf{C},2)},\\ V_f^{(\mathbf{C},1)} &:= \frac{1}{4\pi}\|\nabla f\|^2_{L^2(\mathbf{D})} + \frac{1}{2}\|f\|^2_{\dot{H}^{1/2}(\partial(\mathbf{D}))}, \qquad V_f^{(\mathbf{C},2)} := |\langle f\rangle_{\mathbf{D}} - \langle f\rangle_{\partial\mathbf{D}}|^2,\end{aligned}$$

where $\langle\cdot\rangle_{\mathbf{D}}$ denotes the averaging over $\mathbf{D}$ as in (7.13)*. In contrast, in the real case the limiting variance is given by*

$$V_f^{(\mathbf{R})} = 2V^{(\mathbf{C},1)}_{P_{\mathrm{sym}}f} + \kappa_4 V_f^{(\mathbf{C},2)}.$$

Thus the variances agree exactly in the case of analytic test functions by (7.15) *and $V_f^{(\mathbf{C},2)} = 0$, while e.g. in the case of symmetric test functions, $f = P_{\mathrm{sym}}f$ and vanishing fourth cumulant $\kappa_4 = 0$ the real variance is twice as big as the complex one, $V_f^{(\mathbf{R})} = 2V_f^{(\mathbf{C})}$.*

Remark 7.2.5 (Real correction to the expected circular law)**.** *In [76, Theorem 6.2] Edelman computed the density of genuinely complex eigenvalues of the real Ginibre ensemble to be*

$$\rho_n(x+\mathrm{i}y) := \sqrt{\frac{2n}{\pi}}|y|e^{2ny^2}\operatorname{erfc}(\sqrt{2n}|y|)\frac{\Gamma(n-1,n(x^2+y^2))}{\Gamma(n-1)} \tag{7.16}$$

in terms of the upper incomplete Gamma function $\Gamma(s,x)$*. Using the large n asymptotics uniform in $z = x+\mathrm{i}y$ for the incomplete Gamma function [200, Eq. (2.2)] we obtain*

$$\rho_n(z) \approx \sqrt{\frac{2n}{\pi}}|\Im z|e^{2n(\Im z)^2}\operatorname{erfc}(\sqrt{2n}|\Im z|)\operatorname{erfc}\Big(\operatorname{sgn}(|z|-1)\sqrt{n(|z|^2-1-2\log|z|)}\Big),$$

which, using asymptotics of the error function for any fixed $|z|<1$,

$$\sqrt{\frac{2n}{\pi}}|\Im z|e^{2n(\Im z)^2}\operatorname{erfc}(\sqrt{2n}|\Im z|) \approx \frac{1}{2\pi} - \frac{1}{8n\pi(\Im z)^2},$$

gives that

$$\rho_n(z) = \frac{1}{\pi} - \frac{1}{4\pi n}\frac{1}{(\Im z)^2} + \mathcal{O}(n^{-1}),$$

in agreement with the second term in the rhs. of (7.14) *accounting for the n^{-1}-correction to the circular law away from the real axis.*

The situation very close to the real axis is much more subtle. The density of the real Ginibre eigenvalues is explicitly known [77, Corollary 4.3] and it is asymptotically uniform on $[-1,1]$, see [77, Corollary 4.5], giving a singular correction of mass of order $n^{-1/2}$ to the circular law. However, the abundance of real eigenvalues is balanced by the sparsity of genuinely complex eigenvalues in a narrow strip around the real axis — a consequence of the factor $|y|$ in (7.16)*. Since these two effects of order $n^{-1/2}$ cancel each other on the scale of our test functions f, they are not directly visible in* (7.14)*. Instead we obtain a smaller order correction of order n^{-1} specific to the real axis, in form of the second, the penultimate and the ultimate term in* (7.14)*.*

Remark 7.2.6 (Special case: Polynomial test functions)**.** *We remark that in [98, 182] exact n-dependent formulae for $\mathbf{E}\operatorname{Tr}X^k = \mathbf{E}\sum_i \sigma_i^k$ and real Ginibre X have been obtained. Translated into our scaling it follows from [98, Corollary 4] that*

$$\mathbf{E}\operatorname{Tr}X^k = \begin{cases} 1, & k \text{ even}, \\ 0, & k \text{ odd}, \end{cases} + \mathcal{O}_k(1) \tag{7.17}$$

for integers $k \geq 1$, as $n\to\infty$ (note that the trace is unnormalised). The asymptotics (7.17) *are consistent with* (7.14) *since*

$$\int_{\mathbf{D}} z^k \,\mathrm{d}^2 z = 0, \qquad \int_{-1}^{1} (e^{\mathrm{i}\theta})^k \,\mathrm{d}\theta = 0, \qquad \frac{1^k+(-1)^k}{4} = \begin{cases} \frac{1}{2}, & k \text{ even}, \\ 0, & k \text{ odd}, \end{cases}$$

and

$$\frac{1}{4\pi}\int_{\mathbf{D}} \frac{(\Re z)^k - z^k}{(\Im z)^2}\,\mathrm{d}^2 z = \begin{cases} \frac{1}{2} - 2^{-k}\binom{k-1}{k/2}, & k \text{ even}, \\ 0, & k \text{ odd}, \end{cases},$$

$$\frac{1}{2\pi}\int_{-1}^{1} \frac{x^k}{\sqrt{1-x^2}}\,\mathrm{d}x = \begin{cases} 2^{-k}\binom{k-1}{k/2}, & k \text{ even}, \\ 0, & k \text{ odd}. \end{cases}$$

7.2.1 Connection to the Gaussian free field

It has been observed in [163] that for complex Ginibre matrices the limiting random field $L(f)$ can be viewed as a projection of the *Gaussian free field (GFF)* [178]. In [58, Section 2.1] we extended this interpretation to general complex i.i.d. matrices with $\kappa_4 \geq 0$ and provided an interpretation as a rank-one perturbation of the projected GFF. The real case yields the symmetrised version of the same GFF with respect to the real axis, reflecting the fact that the complex eigenvalues of real matrices come in pairs of complex conjugates. We keep the explanation brief due to the similarity to [58, Section 2.1].

The Gaussian free field on $\mathbf{C}$ is a *Gaussian Hilbert space* of random variables $h(f)$ indexed by functions in the Sobolev space $f \in H_0^1(\mathbf{C})$ such that the map $f \mapsto h(f)$ is linear and

$$\mathbf{E}\, h(f) = 0, \quad \mathbf{E}\, \overline{h(f)} h(g) = \langle f, g \rangle_{H_0^1(\mathbf{C})} = \langle \nabla f, \nabla g \rangle_{L^2(\mathbf{C})}. \tag{7.18}$$

The Sobolev space $H_0^1(\mathbf{C}) = \overline{C_0^\infty(\mathbf{C})}^{\|\cdot\|_{H_0^1(\mathbf{C})}}$ can be orthogonally decomposed into

$$H_0^1(\mathbf{D}) \oplus H_0^1(\overline{\mathbf{D}}^c) \oplus H_0^1(\mathbf{D} \cup \overline{\mathbf{D}}^c)^\perp,$$

i.e. the H_0^1-closure of smooth functions which are compactly supported in $\mathbf{D}$ or $\overline{\mathbf{D}}^c$, and their orthogonal complement $H_0^1((\partial\mathbf{D})^c)^\perp$, the closed subspace of functions analytic outside of $\partial\mathbf{D}$ (see e.g. [178, Thm. 2.17]). With the orthogonal projection P onto the first and third of these subspaces,

$$P := P_{H_0^1(\mathbf{D})} + P_{H_0^1((\partial\mathbf{D})^c)^\perp},$$

we have (see [58, Eq. (2.13)])

$$\|Pf\|_{H_0^1(\mathbf{C})}^2 = \|f\|_{H_0^1(\mathbf{D})}^2 + 2\pi \|f\|_{\dot{H}^{1/2}(\partial\mathbf{D})}^2. \tag{7.19}$$

If $\kappa_4 \geq 0$, then L can be interpreted as

$$L = \frac{1}{\sqrt{2\pi}} P P_{\mathrm{sym}} h + \sqrt{\kappa_4}\Big(\langle \cdot \rangle_{\mathbf{D}} - \langle \cdot \rangle_{\partial\mathbf{D}} \Big) \Xi, \tag{7.20}$$

where Ξ is a standard real Gaussian, independent of h, and the projection of h is to be interpreted by duality, i.e. $(PP_{\mathrm{sym}} h)(f) := h(PP_{\mathrm{sym}} f)$, cf. [58, Eq. (2.15)]. Indeed,

$$\mathbf{E} \left| \frac{1}{\sqrt{2\pi}} h(PP_{\mathrm{sym}} f) + \sqrt{\kappa_4}(\langle f \rangle_{\mathbf{D}} - \langle f \rangle_{\partial\mathbf{D}}) \Xi \right|^2 = C(f,f),$$

as a consequence of (7.18) and (7.19).

7.2.2 Universality of the local singular value statistics of $X - z$ close to zero

As a by-product of our analysis we obtain the universality of the small singular values of $X - z$, and prove that (up to a rescaling) their distribution asymptotically agrees with the singular value distribution of a *complex* Ginibre matrix $\widetilde{X}$ if $z \notin \mathbf{R}$, even though X is a *real* i.i.d. matrix. In the following by $\{\lambda_i^z\}_{i \in [n]}$ we denote the singular values of $X - z$ in increasing order.

It is natural to express universality in terms of the k-point correlation functions $p_{k,z}^{(n)}$ which are defined implicitly by

$$\mathbf{E}\binom{n}{k}^{-1}\sum_{\{i_1,\dots,i_k\}\subset[n]} f(\lambda_{i_1}^z,\dots,\lambda_{i_k}^z)=\int_{\mathbf{R}^k} f(\boldsymbol{x})p_{k,z}^{(n)}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \tag{7.21}$$

for test functions f. The summation in (7.21) is over all the subsets of k distinct integers from $[n]$. Denote by $p_k^{(\infty,\mathbf{C})}$ the scaling limit of the k-point correlation function $p_k^{(n,\mathbf{C})}$ of the singular values of a complex $n\times n$ Ginibre matrix $\widetilde{X}$. See e.g. [96, Eqs. (2.3)–(2.4)] or [28, Eq. (1.3)] for the explicit expression of $p_k^{(\infty,\mathbf{C})}$.

Theorem 7.2.7 (Universality of small singular values of $X-z$)**.** *Fix $z\in\mathbf{C}$ with $|\Im z|\sim 1$, and $|z|\le 1-\epsilon$, for some small fixed $\epsilon>0$. Let X be an i.i.d. matrix with real entries satisfying Assumption (7.A), and denote by ρ^z the self consistent density of states of the singular values of $X-z$ (see (7.27) later). Then for any $k\in\mathbf{N}$, and for any compactly supported test function $F\in C_c^1(\mathbf{R}^k)$, it holds*

$$\int_{\mathbf{R}^k} F(\boldsymbol{x})\left[\rho^z(0)^{-k}p_{k,z}^{(n)}\left(\frac{\boldsymbol{x}}{n\rho^z(0)}\right)-p_k^{(\infty,\mathbf{C})}(\boldsymbol{x})\right]\mathrm{d}\boldsymbol{x}=\mathcal{O}\left(n^{-c(k)}\right), \tag{7.22}$$

where $c(k)>0$ is a small constant only depending on k. The implicit constant in $\mathcal{O}(\cdot)$ may depend on k, $\|F\|_{C^1}$, and C_p from (7.7).

Remark 7.2.8. *Theorem 7.2.7 states that the local statistics of the singular values of $X-z$ close to zero, for $|\Im z|\sim 1$, asymptotically agree with the ones of a complex Ginibre matrix $\widetilde{X}$, even if the entries of X are real i.i.d. random variables. It is expected that the same result holds for all (possibly n-dependent) z as long as $|\Im z|\gg n^{-1/2}$, while in the opposite regime $|\Im z|\ll n^{-1/2}$ the local statistics of the* real *Ginibre prevails with an interpolating family of new statistics which emerges for $|\Im z|\sim n^{-1/2}$.*

Besides the universality of small singular values of $X-z$, our methods also allow us to conclude the asymptotic independence of the small singular values of $X-z_1$ and those of $X-z_2$ for generic z_1,z_2. More precisely, similarly to (7.21), we define the correlation function $p_{k_1,z_1;k_2,z_2}^{(n)}$ for the singular values of $X-z_1$ and $X-z_2$ implicitly by

$$\begin{aligned}\mathbf{E}\binom{n}{k_1}^{-1}\binom{n}{k_2}^{-1}&\sum_{\substack{\{i_1,\dots,i_{k_1}\}\subset[n]\\ \{j_1,\dots,j_{k_2}\}\subset[n]}} f(\boldsymbol{\lambda}_{\boldsymbol{i}}^{z_1},\boldsymbol{\lambda}_{\boldsymbol{j}}^{z_2})\\ &=\int_{\mathbf{R}^{k_1}}\mathrm{d}\boldsymbol{x}_1\int_{\mathbf{R}^{k_2}}\mathrm{d}\boldsymbol{x}_2\, f(\boldsymbol{x}_1,\boldsymbol{x}_2)p_{k_1,z_1;k_2,z_2}^{(n)}(\boldsymbol{x}_1,\boldsymbol{x}_2),\end{aligned} \tag{7.23}$$

for any test function f, and any $k_1,k_2\in\mathbf{N}$, where we used the notations $\boldsymbol{\lambda}_{\boldsymbol{i}}^{z_1}:=(\lambda_{i_1}^{z_1},\dots,\lambda_{i_{k_1}}^{z_1})$ and $\boldsymbol{\lambda}_{\boldsymbol{j}}^{z_2}:=(\lambda_{j_1}^{z_2},\dots,\lambda_{j_{k_2}}^{z_2})$.

Theorem 7.2.9 (Asymptotic independence of small singular values of $X-z_1, X-z_2$)**.** *Let $z_1,z_2\in\mathbf{C}$ be as z in Theorem 7.2.7, and assume that $|z_1-z_2|,|z_1-\overline{z}_2|\sim 1$. Let X be an i.i.d. matrix with real entries satisfying Assumption (7.A), then for any $k_1,k_2\in\mathbf{N}$, and for* *any compactly supported test function* $F \in C_c^1(\mathbf{R}^k)$, *with* $k = k_1 + k_2$, *using the notation* $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, *with* $\boldsymbol{x}_l \in \mathbf{R}^{k_l}$, *it holds*

$$\int_{\mathbf{R}^k} F(\boldsymbol{x}) \left[\frac{p^{(n)}_{k_1,z_1;k_2,z_2}\left(\frac{\boldsymbol{x}_1}{n\rho^{z_1}}, \frac{\boldsymbol{x}_2}{n\rho^{z_2}}\right)}{(\rho^{z_1})^{k_1}(\rho^{z_2})^{k_2}} - p^{(\infty,\mathbf{C})}_{k_1}(\boldsymbol{x}_1) p^{(\infty,\mathbf{C})}_{k_2}(\boldsymbol{x}_2) \right] \mathrm{d}\boldsymbol{x} = \mathcal{O}\left(n^{-c(k)}\right), \tag{7.24}$$

where $\rho^{z_l} = \rho^{z_l}(0)$, *and* $c(k) > 0$ *is a small constant only depending on* k. *The implicit constant in* $\mathcal{O}(\cdot)$ *may depend on* k, $\|F\|_{C^1}$, *and* C_p *from* (7.7).

Remark 7.2.10. *We stated Theorem 7.2.7 for two different* z_1, z_2 *for notational simplicity. The analogous result holds for any finitely many* $z_1, \ldots, z_q$ *such that* $|z_l - z_m|, |z_l - \overline{z}_m| \sim 1$, *with* $l, m \in [q]$.

7.3 Proof strategy

The proof of Theorem 7.2.1 follows a similar strategy as the proof of [58, Theorem 2.2] with several major changes. We use Girko's formula to relate the eigenvalues of X to the resolvent of the $2n \times 2n$ matrix

$$H^z := \begin{pmatrix} 0 & X - z \\ (X-z)^* & 0 \end{pmatrix}, \tag{7.25}$$

the so called *Hermitisation* of $X - z$. We denote the eigenvalues of H^z, which come in pairs symmetric with respect to zero, by $\{\lambda^z_{\pm i}\}_{i\in[n]}$. The *local law*, see Theorem 7.3.1 below, asserts that the resolvent $G(w) = G^z(w) := (H^z - w)^{-1}$ of H^z with $\eta = \Im w \neq 0$ becomes approximately deterministic, as $n \to \infty$. Its limit is expressed via the unique solution of the scalar equation

$$-\frac{1}{m^z} = w + m^z - \frac{|z|^2}{w + m^z}, \quad \eta \Im m^z(w) > 0, \quad \eta = \Im w \neq 0, \tag{7.26}$$

which is a special case of the *matrix Dyson equation* (MDE), see e.g. [5] and (7.56) later. Note that on the imaginary axis $m^z(\mathrm{i}\eta) = \mathrm{i}\Im m^z(\mathrm{i}\eta)$. We define the *self-consistent density of states* of H^z and its extension to the upper half-plane by

$$\rho^z(E) := \rho^z(E + \mathrm{i}0), \qquad \rho^z(w) := \frac{1}{\pi}\Im m^z(w). \tag{7.27}$$

In terms of m^z the deterministic approximation to G^z is given by the $2n \times 2n$ block matrix

$$M^z(w) := \begin{pmatrix} m^z(w) & -zu^z(w) \\ -\overline{z}u^z(w) & m^z(w) \end{pmatrix}, \quad u^z(w) := \frac{m^z(w)}{w + m^z(w)}, \tag{7.28}$$

where each block is understood to be a scalar multiple of the $n \times n$ identity matrix. We note that m, u, M are uniformly bounded in z, w, i.e.

$$\|M^z(w)\| + |m^z(w)| \lesssim 1, \quad |u^z(w)| \leq |m^z(w)|^2 + |u^z(w)|^2|z|^2 < 1, \tag{7.29}$$

see e.g. [58, Eqs. (3.3)-(3.5)].

The *local law* for $G^z(w)$ in its full averaged and isotropic form has been obtained for $w \in \mathrm{i}\mathbf{R}$ in [11] for the bulk regime $|1 - |z|| \geq \epsilon$ and in [13] for the edge regime $|1 - |z|| <$

ϵ. In fact, in the companion paper [58] on the complex CLT the local law for w on the imaginary axis was sufficient. For the real CLT, however, we need its extension to general spectral parameters w in the bulk $|1-|z||\geq\epsilon$ case that we state below. We remark that tracial and entry-wise form of the local law in Theorem 7.3.1 has already been established in [44, Theorem 3.4].

Theorem 7.3.1 (Optimal local law for G). *For any $\epsilon>0$ and $z\in\mathbf{C}$ with $|1-|z||\geq\epsilon$ the resolvent G^z at $w\in\mathbf{H}$ with $\eta=\Im w$ is very well approximated by the deterministic matrix M^z in the sense that*

$$\begin{aligned} |\langle (G^z(w)-M^z(w))A\rangle| &\leq \frac{C_\epsilon\|A\|n^\xi}{n\eta},\\ |\langle \boldsymbol{x},(G^z(w)-M^z(w))\boldsymbol{y}\rangle| &\leq C_\epsilon\|\boldsymbol{x}\|\|\boldsymbol{y}\|n^\xi\Big(\frac{1}{\sqrt{n\eta}}+\frac{1}{n\eta}\Big), \end{aligned} \tag{7.30}$$

with very high probability for some $C_\epsilon\leq\epsilon^{-100}$, uniformly for $\eta\geq n^{-100}$, $|1-|z||\geq\epsilon$, and for any deterministic matrices A and vectors $\boldsymbol{x},\boldsymbol{y}$, and $\xi>0$.

Remark 7.3.2 (Cusp fluctuation averaging). *For $w\in\mathrm{i}\mathbf{R}$ we may choose $C_\epsilon=1$ by [13, Theorem 5.2] which takes into account the* cusp fluctuation averaging effect. *Since it is not necessary for the present work we refrain from adapting this technique for general w and rather present a conceptually simpler proof resulting in the ϵ-dependent bounds* (7.30).

As in [58] we express the linear statistics (7.1) of eigenvalues σ_i of X through the resolvent G^z via Girko's Hermitisation formula (7.3)

$$\begin{aligned} L_n(f) &= \frac{1}{4\pi}\int_{\mathbf{C}}\Delta f(z)\Big[\log|\det(H^z-\mathrm{i}T)|-\mathbf{E}\log|\det(H^z-\mathrm{i}T)|\Big]\,\mathrm{d}^2z\\ &\quad-\frac{n}{2\pi\mathrm{i}}\int_{\mathbf{C}}\Delta f(z)\left[\left(\int_0^{\eta_0}+\int_{\eta_0}^{\eta_c}+\int_{\eta_c}^{T}\right)\left[\langle G^z(\mathrm{i}\eta)-\mathbf{E}\,G^z(\mathrm{i}\eta)\rangle\right]\mathrm{d}\eta\right]\mathrm{d}^2z\\ &=: J_T+I_0^{\eta_0}+I_{\eta_0}^{\eta_c}+I_{\eta_c}^{T}, \end{aligned} \tag{7.31}$$

for $\eta_0=n^{-1-\delta_0}$, $\eta_c=n^{-1+\delta_1}$, and $T=n^{100}$, where J_T in (7.31) corresponds to the rhs. of the first line in (7.31) whilst $I_0^{\eta_0}, I_{\eta_0}^{\eta_c}, I_{\eta_c}^{T}$ correspond to the three different η-integrals in the second line of (7.31). Here we used that by spectral symmetry of H^z it follows that $\langle G^z(\mathrm{i}\eta)\rangle\in\mathrm{i}\mathbf{R}$ and therefore $\Im\langle G^z(\mathrm{i}\eta)\rangle=\langle G^z(\mathrm{i}\eta)\rangle/\mathrm{i}$ in order to obtain (7.31) from (7.3). The regime J_T can be trivially estimated by [58, Lemma 4.3], while the regime $I_0^{\eta_0}$ can be controlled using [196, Thm. 3.2] as in [58, Lemma 4.4] (see [58, Remark 4.5] for an alternative proof). Both contributions are negligible. For the main term $I_{\eta_c}^{T}$ we prove the following *resolvent CLT*.

Proposition 7.3.3 (CLT for resolvents). *Let $\epsilon>0$, $\eta_1,\dots,\eta_p>0$, and $z_1,\dots,z_p\in\mathbf{C}$ be such that for any $i\neq j$, $\min\{\eta_i,\eta_j\}\geq n^{\epsilon-1}|z_i-z_j|^{-2}$. Then for any $\xi>0$ the traces of the resolvents $G_i=G^{z_i}(\mathrm{i}\eta_i)$ satisfy an asymptotic Wick theorem*

$$\begin{aligned} \mathbf{E}\prod_{i\in[p]}\langle G_i-\mathbf{E}\,G_i\rangle &= \sum_{P\in\mathrm{Pairings}([p])}\prod_{\{i,j\}\in P}\mathbf{E}\langle G_i-\mathbf{E}\,G_i\rangle\langle G_j-\mathbf{E}\,G_j\rangle+\mathcal{O}(\Psi)\\ &= \frac{1}{n^p}\sum_{P\in\mathrm{Pairings}([p])}\prod_{\{i,j\}\in P}\frac{\widehat{V}_{i,j}+\kappa_4U_iU_j}{2}+\mathcal{O}(\Psi), \end{aligned} \tag{7.32}$$

where

$$\Psi := \frac{n^{\xi}}{(n\eta_*)^{1/2}} \frac{1}{\min_{i\neq j}|z_i - z_j|^4} \prod_{i\in[p]} \Big(\frac{1}{|1-|z_i||} + \frac{1}{(\Im z_i)^2} \Big) \frac{1}{n\eta_i}, \qquad \eta_* := \min_i \eta_i, \quad (7.33)$$

and $\widehat{V}_{i,j} = \widehat{V}(z_i, z_j, \eta_i, \eta_j)$ *and* $U_i = U(z_i, \eta_i)$ *are defined as*

$$\begin{aligned} \widehat{V}(z_i, z_j, \eta_i, \eta_j) &:= V(z_i, z_j, \eta_i, \eta_j) + V(z_i, \overline{z_j}, \eta_i, \eta_j) \\ V(z_i, z_j, \eta_i, \eta_j) &:= \frac{1}{2}\partial_{\eta_i}\partial_{\eta_j} \log[1 + (u_i u_j |z_i||z_j|)^2 - m_i^2 m_j^2 - 2u_i u_j \Re z_i \overline{z_j}], \\ U(z_i, \eta_i) &:= \frac{\mathrm{i}}{\sqrt{2}} \partial_{\eta_i} m_i^2, \end{aligned} \quad (7.34)$$

with $m_i = m^{z_i}(\mathrm{i}\eta_i)$ *and* $u_i = u^{z_i}(\mathrm{i}\eta_i)$ *from* (7.26)–(7.28).

Moreover, the expectation of the normalised trace of $G = G_i$ *is given by*

$$\mathbf{E}\langle G\rangle = \langle M\rangle + \mathcal{E} + \mathcal{O}\Big(\Big(\frac{1}{|1-|z||} + \frac{1}{|\Im z|^2} \Big) \Big(\frac{1}{n^{3/2}(1+\eta)} + \frac{1}{(n\eta)^2} \Big) \Big), \quad (7.35)$$

where

$$\mathcal{E} := -\frac{\mathrm{i}\kappa_4}{4n} \partial_\eta (m^4) + \frac{\mathrm{i}}{4n} \partial_\eta \log\Big(1 - u^2 + 2u^3|z|^2 - u^2(z^2 + \overline{z}^2) \Big). \quad (7.36)$$

Proposition 7.3.3 is the real analogue of [58, Proposition 3.3]. The main differences are that (i) the V-term for the variance appears in a symmetrised form with z_j and $\overline{z_j}$, (ii) the error term (7.33) deteriorates as $\Im z_i \approx 0$, and (iii) the expectation (7.35) has an additional subleading term which is even present in case $\kappa_4 = 0$ (second term in (7.36)).

Finally, in order to show that $I_{\eta_0}^{\eta_c}$ in (7.31) is negligible, we prove that $\langle G^{z_1}(\mathrm{i}\eta_1)\rangle$ and $\langle G^{z_2}(\mathrm{i}\eta_2)\rangle$ are asymptotically independent if z_1, z_2 and z_1, $\overline{z}_2$ are far enough from each other, they are far away from the real axis, they are well inside $\mathbf{D}$, and $\eta_0 \leq \eta_1, \eta_2 \leq \eta_c$. These regimes of the parameters z_1, z_2 represent the overwhelming part of the $\mathrm{d}^2 z_1\, \mathrm{d}^2 z_2$ integration in the calculation of $\mathbf{E}|I_{\eta_0}^{\eta_c}|^2$. The following proposition is the direct analogue of [58, Proposition 3.5].

Proposition 7.3.4 (Independence of resolvents with small imaginary part)**.** *Fix* $p \in \mathbf{N}$. *For any sufficiently small* $\omega_h, \omega_d > 0$ *there exist* $\omega_*, \delta_0, \delta_1$ *with* $\omega_h \ll \delta_m \ll \omega_* \ll 1$, *for* $m = 0, 1$, *such that for any choice of* $z_1, \dots, z_p$ *with*

$$|z_l| \leq 1 - n^{-\omega_h}, \ |z_l - z_m| \geq n^{-\omega_d}, \ |z_l - \overline{z}_m| \geq n^{-\omega_d}, \ |z_l - \overline{z}_l| \geq n^{-\omega_d},$$

with $l, m \in [p]$, $l \neq m$, *it follows that*

$$\mathbf{E} \prod_{l=1}^{p} \langle G^{z_l}(\mathrm{i}\eta_l)\rangle = \prod_{l=1}^{p} \mathbf{E}\langle G^{z_l}(\mathrm{i}\eta_l)\rangle + \mathcal{O}\left(\frac{n^{p(\omega_h + \delta_0) + \delta_1}}{n^{\omega_*}} \right), \quad (7.37)$$

for any $\eta_1, \dots, \eta_p \in [n^{-1-\delta_0}, n^{-1+\delta_1}]$.

As in the complex case [58], one key ingredient for both Propositions 7.3.3 and 7.3.4 is a local law for products of resolvents G_1, G_2 for $G_i = G^{z_i}(w_i)$. We remark that local laws for products of resolvents have also been derived for (generalized) Wigner matrices [89, 139] and for sample covariance matrices [56], as well as for the addition of random matrices [24].

Note that the deterministic approximation to $G_1 G_2$ is not given simply by $M_1 M_2$ where $M_i := M^{z_i}(w_i)$ from (7.28). To describe the correct approximation, as in [58, Section 5], we define the *stability operator*

$$\widehat{\mathcal{B}} = \widehat{\mathcal{B}}_{12} = \widehat{\mathcal{B}}(z_1, z_2, w_1, w_2) := 1 - M_1 \mathcal{S}[\cdot] M_2, \tag{7.38}$$

acting on the space of $2n \times 2n$ matrices. Here the linear *covariance* or *self-energy operator* $\mathcal{S}\colon \mathbf{C}^{2n\times 2n} \to \mathbf{C}^{2n\times 2n}$ is defined as

$$\mathcal{S}\left[\begin{pmatrix} A & B \\ C & D \end{pmatrix}\right] := \widetilde{\mathbf{E}} \widetilde{W} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \widetilde{W} = \begin{pmatrix} \langle D \rangle & 0 \\ 0 & \langle A \rangle \end{pmatrix}, \quad \widetilde{W} = \begin{pmatrix} 0 & \widetilde{X} \\ \widetilde{X}^* & 0 \end{pmatrix}, \quad \widetilde{X} \sim \mathrm{Gin}_{\mathbf{C}}, \tag{7.39}$$

i.e. it averages the diagonal blocks and swaps them. Here $\widetilde{\mathbf{E}}$ denotes the expectation with respect to $\widetilde{X}$, $\langle A \rangle = n^{-1}\mathrm{Tr} A$ and $\mathrm{Gin}_{\mathbf{C}}$ stands for the standard complex Ginibre ensemble. The ultimate equality in (7.39) follows directly from $\mathbf{E}\, \widetilde{x}_{ab}^2 = 0$, $\mathbf{E}|\widetilde{x}_{ab}|^2 = n^{-1}$. Note that as a matter of choice we define the stability operator (7.38) with the covariance operator $\mathcal{S}$ corresponding to the complex rather than the real Ginibre ensemble. However, to leading order there is no difference between the two and the present choice is more consistent with the companion paper [58]. The effect of this discrepancy will be estimated in a new error term (see (7.81) later).

For any deterministic matrix B we define

$$M_B^{z_1,z_2}(w_1, w_2) := \widehat{\mathcal{B}}_{12}^{-1}[M^{z_1}(w_1) B M^{z_2}(w_2)], \tag{7.40}$$

which turns out to be the deterministic approximation to $G_1 B G_2$. Indeed, from the local law for G_1, G_2, Theorem 7.3.1, and [58, Theorem 5.2] we immediately conclude the following theorem.

Theorem 7.3.5 (Local law for $G^{z_1} B G^{z_2}$). *Fix $z_1, z_2 \in \mathbf{C}$ with $|1 - |z_i|| \geq \epsilon$, for some $\epsilon > 0$ and $w_1, w_2 \in \mathbf{C}$ with $|\eta_i| := |\Im w_i| \geq n^{-1}$ such that*

$$\eta_* := \min\{|\eta_1|, |\eta_2|\} \geq n^{-1+\epsilon_*} |\widehat{\beta}_*|^{-1},$$

for some small $\epsilon_* > 0$, where $\widehat{\beta}_*$ is the, in absolute value, smallest eigenvalue of $\widehat{\mathcal{B}}_{12}$ defined in (7.38). Then, for any bounded deterministic matrix B, $\|B\| \lesssim 1$, the product of resolvents $G^{z_1} B G^{z_2} = G^{z_1}(w_1) B G^{z_2}(w_2)$ is well approximated by $M_B^{z_1,z_2} = M_B^{z_1,z_2}(w_1, w_2)$ defined in (7.40) in the sense that

$$\begin{aligned} |\langle A(G^{z_1} B G^{z_2} - M_B^{z_1,z_2})\rangle| &\leq \frac{C_\epsilon \|A\| n^\xi}{n\eta_* |\eta_1\eta_2|^{1/2} |\widehat{\beta}_*|} \left(\eta_*^{1/12} + \frac{\eta_*^{1/4}}{|\widehat{\beta}_*|} + \frac{1}{\sqrt{n\eta_*}} + \frac{1}{(|\widehat{\beta}_*| n\eta_*)^{1/4}} \right), \\ |\langle \boldsymbol{x}, (G^{z_1} B G^{z_2} - M_B^{z_1,z_2}) \boldsymbol{y} \rangle| &\leq \frac{C_\epsilon \|\boldsymbol{x}\| \|\boldsymbol{y}\| n^\xi}{(n\eta_*)^{1/2} |\eta_1\eta_2|^{1/2} |\widehat{\beta}_*|} \end{aligned} \tag{7.41}$$

for some C_ϵ with very high probability for any deterministic $A, \boldsymbol{x}, \boldsymbol{y}$ and $\xi > 0$. If $w_1, w_2 \in \mathrm{i}\mathbf{R}$ we may choose $C_\epsilon = 1$, otherwise we can choose $C_\epsilon \leq \epsilon^{-100}$.

An effective lower bound on $\Re \widehat{\beta}_*$, hence on $|\widehat{\beta}_*|$, will be given in Lemma 7.6.1 later.

The paper is organised as follows: In Section 7.4 we prove Theorem 7.2.1 by combining Propositions 7.3.3 and 7.3.4. In Section 7.5 we prove the local law for G away from the imaginary axis, Theorem 7.3.1. In Section 7.6 we prove Proposition 7.3.3, the Central Limit Theorem for resolvents using Theorem 7.3.5. In Section 7.7 we prove Proposition 7.3.4 again using Theorem 7.3.5, and conclude Theorem 7.2.7.

Note that Theorem 7.3.5, the local law for $G^{z_1}BG^{z_2}$, is used in two different contexts. Traces of $AG^{z_1}BG^{z_2}$, for some deterministic matrices $A, B \in \mathbf{C}^{2n\times 2n}$, naturally arise along the cumulant expansion for $\prod_i \langle G_i - \mathbf{E}\, G_i\rangle$ in Proposition 7.3.3. The proof of Proposition 7.3.4 is an analysis of weakly correlated DBMs, where the correlations are given by eigenvector overlaps (7.6), whose estimate is reduced to an upper bound on $\langle \Im G^{z_1} \Im G^{z_2}\rangle$.

7.4 Central limit theorem for linear statistics: Proof of Theorem 7.2.1

From Propositions 7.3.3 and 7.3.4 we conclude Theorem 7.2.1 analogously to [58, Section 4], we only describe the few minor modifications.

Proof of Theorem 7.2.1. We explain the three modifications compared with the proof of [58, Theorem 2.2]. First, there are two additional terms in in the variance (7.34) and expectation (7.36) of the resolvent CLT, compared to [58, Eqs. (3.14)-(3.15)]. These additional terms result in additional explicit terms in (7.14) and (7.13). For the expectation in (7.14) we have

$$\begin{aligned}&-\frac{1}{2\pi \mathrm{i}}\int_{\mathbf{C}} \Delta f(z)\frac{\mathrm{i}}{4n}\int_0^\infty \partial_\eta \log\Big(1-u^2+2u^3|z|^2-u^2(z^2+\overline{z}^2)\Big)\,\mathrm{d}\eta\,\mathrm{d}^2 z \\ &= \frac{1}{4\pi}\int_{\mathbf{D}}\frac{f(\Re z)-f(z)}{(\Im z)^2}\,\mathrm{d}^2 z-\frac{1}{2\pi}\int_0^{2\pi} f(e^{\mathrm{i}\theta})\,\mathrm{d}\theta+\frac{1}{2\pi}\int_{-1}^1\frac{f(x)}{\sqrt{1-x^2}}\,\mathrm{d}x+\frac{f(1)+f(-1)}{4}\end{aligned} \tag{7.42}$$

and for the variance in (7.13) we have

$$\begin{aligned}&-\frac{1}{8\pi^2}\int_{\mathbf{C}}\mathrm{d}^2 z_1\int_{\mathbf{C}}\mathrm{d}^2 z_2 \Delta f(z_1)\overline{\Delta g(z_2)}\int_0^\infty \mathrm{d}\eta_1\int_0^\infty \mathrm{d}\eta_2 V(z_1,\overline{z_2},\eta_1,\eta_2)\\ &\quad = \frac{1}{4\pi}\langle \nabla g(\bar\cdot),\nabla f\rangle_{L^2(\mathbf{D})}+\frac{1}{2}\langle g(\bar\cdot), f\rangle_{\dot H^{1/2}(\partial\mathbf{D})},\end{aligned} \tag{7.43}$$

so that together with contribution from $V(z_1, z_2, \eta_1, \eta_2)$ in (7.34) we have

$$\begin{aligned}&\frac{1}{4\pi}\langle \nabla g+\nabla g(\bar\cdot),\nabla f\rangle_{L^2(\mathbf{D})}+\frac{1}{2}\langle g+g(\bar\cdot), f\rangle_{\dot H^{1/2}(\partial\mathbf{D})}\\ &\quad=\frac{1}{2\pi}\langle \nabla P_{\mathrm{sym}} g,\nabla P_{\mathrm{sym}} f\rangle_{L^2(\mathbf{D})}+\langle P_{\mathrm{sym}} g, P_{\mathrm{sym}} f\rangle_{\dot H^{1/2}(\partial\mathbf{D})}.\end{aligned}$$

The identities (7.42)–(7.43) will be proven separately below. The other two modifications concern the error terms in (7.33) and (7.35). Namely, there is an additional factor including $(\Im z_l)^{-2}$ (cf. [58, Eqs. (3.13), (3.15)]), and, finally, (7.37) holds under the additional assumption that $|z_l - \overline{z}_m| \ge n^{-\omega_d}$, and $|z_l - \overline{z}_l| \ge n^{-\omega_d}$ (cf. [58, Proposition 3.5]). Both these issues can be handled in the same way as the constraints on $|z_l - z_m|$ have been treated in [58, Section 4] (see e.g. [58, Eq. (4.11)]). This means that we additionally exclude the regimes of negligible volume $|z_l - \overline{z}_m| < n^{-\omega_d}$ or $|z_l - \overline{z}_l| < n^{-\omega_d}$ from the $\mathrm{d}z_1 \ldots \mathrm{d}z_p$-integral in [58, Eqs. (4.10), (4.22)] using the almost optimal *a priori* bound from [58, Lemma 4.3]. □

Proof of (7.42). With the short-hand notation $z = x + \mathrm{i}y$, we compute

$$\begin{aligned} &\int_0^\infty \frac{\mathrm{i}}{4n} \partial_\eta \log\Big(1 - u^2 + 2u^3|z|^2 - u^2(z^2 + \overline{z}^2)\Big) \,\mathrm{d}\eta \\ &\quad = -\frac{\mathrm{i}}{4n} \begin{cases} \log 4 + 2\log|y|, & |z| \le 1, \\ \log|(x^2+y^2)^2 + 1 - 2(x^2 - y^2)| - \log|(x^2+y^2)^2|, & |z| > 1, \end{cases} \end{aligned} \tag{7.44}$$

using that $u = 1 + \mathcal{O}(\eta)$ for $|z| \le 1$ and $u = |z|^{-2} + \mathcal{O}(\eta)$ for $|z| > 1$, so that for (7.42) we need to compute

$$\frac{1}{4}\int_{\mathbf{C}} \Delta f(z) \big[(\log 4 + 2\log|y|)\mathbf{1}(|z| \le 1) + (\log|z-1|^2 + \log|z+1|^2 - 2\log|z|^2)\mathbf{1}(|z| \ge 1)\big] \,\mathrm{d}^2 z. \tag{7.45}$$

We may assume that f is symmetric with respect to the real axis, i.e. $f = P_{\mathrm{sym}} f$ with P_{sym} as in (7.12) since $L_n(f - P_{\mathrm{sym}} f) = 0$ by symmetry of the spectrum and therefore $L_n(f) = L_n(P_{\mathrm{sym}} f)$. Since the functions in (7.45) are singular we introduce an ϵ-regularisation which enables us to perform integration by parts. In particular, the integral in (7.45) is equal to the $\epsilon \to 0$ limit of

$$\begin{aligned} \int_{\mathbf{C}} \partial_z \partial_{\overline{z}} f(z) \big[&(\log 4 + 2\log|y|)\mathbf{1}(|z| \le 1, |y| \ge \epsilon) \\ &+ (\log|z-1|^2 + \log|z+1|^2 - 2\log|z|^2)\mathbf{1}(|z| \ge 1, |z \pm 1| \ge \epsilon)\big] \,\mathrm{d}^2 z, \end{aligned} \tag{7.46}$$

where $|z \pm 1| \ge \epsilon$ denotes that $|z-1| \ge \epsilon$ and $|z+1| \ge \epsilon$, and we used that the contribution from the regimes $|y| \le \epsilon$ and $|z \pm 1| \le \epsilon$ are negligible as $\epsilon \to 0$. In the following equalities should be understood in the $\epsilon \to 0$ limit.

Since

$$\log|z-1|^2 + \log|z+1|^2 - 2\log|z|^2 = \log 4 + 2\log|y|$$

for $|z| = 1$, when integrating by parts in (7.46), the terms where either $\mathbf{1}(|z| \le 1)$ or $\mathbf{1}(|z| > 1)$ are differentiated are equal to zero, using that

$$\partial_z \mathbf{1}(|z| \ge 1) \,\mathrm{d}^2 z = \frac{\mathrm{i}}{2} \mathbf{1}(|z| = 1) \,\mathrm{d}\overline{z}. \tag{7.47}$$

We remark that (7.47) is understood in the sense of distributions, i.e. the equality holds when tested against compactly supported test functions f:

$$-\int_{\mathbf{C}} \partial_z f(z) \mathbf{1}(|z| \ge 1) \,\mathrm{d}^2 z = \frac{\mathrm{i}}{2} \int_{|z|=1} f(z) \,\mathrm{d}\overline{z}.$$

Moreover, with a slightly abuse of notation in (7.47) by $\mathbf{1}(|z| = 1)\,\mathrm{d}\overline{z}$ we denote the clockwise contour integral over the unit circle. This notation is used in the remainder of this section.

Then, performing integration by parts with respect to $\partial_{\overline{z}}$, we conclude that (7.46) is equal to

$$-\int_{\mathbf{C}} \partial_z f(z) \left[\frac{\mathrm{i}}{y} \mathbf{1}(|z| \le 1, |y| \ge \epsilon) + \left(\frac{1}{\overline{z} - 1} + \frac{1}{\overline{z} + 1} - \frac{2}{\overline{z}}\right) \mathbf{1}(|z| \ge 1, |z \pm 1| \ge \epsilon)\right] \mathrm{d}^2 z. \tag{7.48}$$

In order to get (7.48) we used that

$$|\partial_z f(x+\mathrm{i}\epsilon) - \partial_z f(x-\mathrm{i}\epsilon)| \cdot |\log \epsilon| \lesssim \epsilon^{\delta'},$$

for some small fixed $\delta' > 0$, by $f \in H^{2+\delta}$, and similarly all the other ϵ-boundary terms tend to zero. This implies that when the $\partial_{\overline{z}}$ derivative hits the ϵ-boundary terms then these give a negligible contribution as $\epsilon \to 0$. We now consider the two terms in (7.48) separately.

Since the integral of y^{-1} over $\mathbf{D}$ is zero we can rewrite the first term in (7.48) as

$$-\int_{\mathbf{C}} \partial_z (f(z) - f(x)) \frac{\mathrm{i}}{y} \mathbf{1}(|z| \le 1, |y| \ge \epsilon) \, \mathrm{d}^2 z.$$

Then performing integration by parts we conclude that the first term in (7.48) is equal to

$$-\frac{1}{2}\int_{\mathbf{D}} \frac{f(x+\mathrm{i}y) - f(x)}{y^2} \, \mathrm{d}x \, \mathrm{d}y - \frac{\mathrm{i}}{2}\int_0^{2\pi} \frac{f(e^{\mathrm{i}\theta}) - f(\cos\theta)}{\sin\theta} e^{-\mathrm{i}\theta} \, \mathrm{d}\theta \tag{7.49}$$

where we used that

$$\left| \frac{f(x,\epsilon) - 2f(x,0) + f(x,-\epsilon)}{\epsilon} \right| \lesssim \epsilon^{\delta'},$$

to show that the terms when the ∂_z derivative hits the ϵ-boundary terms go to zero as $\epsilon \to 0$. Note that the integrals in (7.49) are absolutely convergent since f is symmetric with respect to the real axis. For the second term in (7.49) we further compute

$$\begin{aligned} \int_0^{2\pi} \frac{f(e^{\mathrm{i}\theta}) - f(\cos\theta)}{\sin\theta} e^{-\mathrm{i}\theta} \, \mathrm{d}\theta &= \int_0^{2\pi} \frac{f(e^{\mathrm{i}\theta}) - f(\cos\theta)}{\sin\theta} (\cos\theta - \mathrm{i}\sin\theta) \, \mathrm{d}\theta \\ &= -\mathrm{i}\int_0^{2\pi} \Big(f(e^{\mathrm{i}\theta}) - f(\cos\theta) \Big) \, \mathrm{d}\theta \end{aligned} \tag{7.50}$$

where we used that the term with $\cos\theta / \sin\theta$ is zero by symmetry.

With defining the domain

$$\Omega_\epsilon := \{|z| \ge 1\} \cap \{|z \pm 1| \ge \epsilon\},$$

the second term in (7.48) is equal to

$$-\int_{\Omega_\epsilon} \partial_z f(z) \left(\frac{1}{\overline{z}-1} + \frac{1}{\overline{z}+1} - \frac{2}{\overline{z}} \right) \mathrm{d}^2 z. \tag{7.51}$$

Since

$$\frac{1}{\overline{z}-1} + \frac{1}{\overline{z}+1} - \frac{2}{\overline{z}}$$

is anti-holomorphic on Ω_ϵ, performing integration by parts with respect to ∂_z in (7.51), we obtain

$$-\int_{\Omega_\epsilon} \partial_z f(z) \left(\frac{1}{\overline{z}-1} + \frac{1}{\overline{z}+1} - \frac{2}{\overline{z}} \right) \mathrm{d}^2 z = \frac{\mathrm{i}}{2}\int_{\partial\Omega_\epsilon} f(z) \left(\frac{1}{\overline{z}-1} + \frac{1}{\overline{z}+1} - \frac{2}{\overline{z}} \right) \mathrm{d}\overline{z}. \tag{7.52}$$

Taking the limit $\epsilon \to 0$ in the r.h.s. of (7.52) we conclude

$$\lim_{\epsilon\to 0} \frac{\mathrm{i}}{2}\int_{\partial\Omega_\epsilon} f(z)\left(\frac{1}{\overline{z}-1}+\frac{1}{\overline{z}+1}-\frac{2}{\overline{z}}\right)\mathrm{d}\overline{z} = \frac{\pi}{2}[f(1)+f(-1)] - \int_0^{2\pi} f(e^{\mathrm{i}\theta})\,\mathrm{d}\theta + \lim_{\epsilon\to 0}\left(\int_\epsilon^{\pi-\epsilon}+\int_{\pi+\epsilon}^{2\pi-\epsilon}\right) f(e^{\mathrm{i}\theta})\frac{e^{-2\mathrm{i}\theta}}{e^{-2\mathrm{i}\theta}-1}\,\mathrm{d}\theta. \tag{7.53}$$

The last term in (7.53) simplifies to

$$\begin{aligned}\lim_{\epsilon\to 0}\left(\int_\epsilon^{\pi-\epsilon}+\int_{\pi+\epsilon}^{2\pi-\epsilon}\right) f(e^{\mathrm{i}\theta})\frac{e^{-2\mathrm{i}\theta}}{e^{-2\mathrm{i}\theta}-1}\,\mathrm{d}\theta &= \lim_{\epsilon\to 0}\left(\int_\epsilon^{\pi-\epsilon}+\int_{\pi+\epsilon}^{2\pi-\epsilon}\right) f(e^{\mathrm{i}\theta})\Big[\frac{\mathrm{i}}{2}\frac{\cos\theta}{\sin\theta}+\frac{1}{2}\Big]\,\mathrm{d}\theta \\ &= \frac{1}{2}\int_0^{2\pi} f(e^{\mathrm{i}\theta})\,\mathrm{d}\theta,\end{aligned} \tag{7.54}$$

by symmetry. By combining (7.49)–(7.54) we conclude (7.42). □

Proof of (7.43). By change of variables $z_2 \to \overline{z_2}$ we can then write

$$\begin{aligned}&\int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 \int_0^\infty \mathrm{d}\eta_1 \int_0^\infty \mathrm{d}\eta_2 \Delta f(z_1)\Delta\overline{g(z_2)} V(z_1,\overline{z_2},\eta_1,\eta_2) \\ &\quad = \int_{\mathbf{C}} \mathrm{d}^2 z_1 \int_{\mathbf{C}} \mathrm{d}^2 z_2 \int_0^\infty \mathrm{d}\eta_1 \int_0^\infty \mathrm{d}\eta_2 \Delta f(z_1)\Delta\overline{g(\overline{z_2})} V(z_1,z_2,\eta_1,\eta_2)\end{aligned} \tag{7.55}$$

such that [58, Lemma 4.8] is applicable and (7.43) follows. □

7.5 Local law away from the imaginary axis: Proof of Theorem 7.3.1

The goal of this section is to prove a local law for $G = G^z(w)$ for z in the bulk, as stated in Theorem 7.3.1. We do not follow the precise ϵ-dependence in the proof explicitly but it can be checked from the arguments below that $C_\epsilon = \epsilon^{-100}$ clearly suffices. We denote the unique solution to the deterministic matrix equation (see e.g. [5])

$$-1 = \mathcal{S}[M]M + ZM + wM, \quad Z := \begin{pmatrix} 0 & z \\ \overline{z} & 0 \end{pmatrix}, \quad \Im M > 0, \quad \Im w > 0 \tag{7.56}$$

by $M = M^z(w)$, where we recall the definition of $\mathcal{S}$ from (7.39). The solution to (7.56) is given by (7.28). To keep notations compact, we first introduce a commonly used (see, e.g. [81]) notion of high-probability bound.

Definition 7.5.1 (Stochastic Domination). *If*

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

are families of non-negative random variables indexed by n*, and possibly some parameter* u *in a set* $U^{(n)}$*, then we say that* X *is stochastically dominated by* Y*, if for all* $\epsilon, D > 0$ *we have*

$$\sup_{u\in U^{(n)}} \mathbf{P}\left[X^{(n)}(u) > n^\epsilon Y^{(n)}(u)\right] \le n^{-D}$$

for large enough $n \geq n_0(\epsilon, D)$. In this case we use the notation $X \prec Y$. Moreover, if we have $|X| \prec Y$ for families of random variables X, Y, we also write $X = \mathcal{O}_\prec(Y)$.

Let us assume that some a-priori bounds

$$|\langle \boldsymbol{x}, (G-M)\boldsymbol{y}\rangle| \prec \Lambda, \qquad |\langle A(G-M)\rangle| \prec \xi \tag{7.57}$$

for some deterministic control functions Λ and ξ depending on w, z have already been established, uniformly in $\boldsymbol{x}, \boldsymbol{y}, A$ under the constraint $\|\boldsymbol{x}\|, \|\boldsymbol{y}\|, \|A\| \leq 1$. From the resolvent equation $1 = (W - Z - w)G$ we obtain

$$-1 = -WG + ZG + wG = \mathcal{S}[G]G + ZG + wG - \underline{WG}, \tag{7.58}$$

where we introduced the *self-renormalisation*, denoted by underlining, of a random variable of the form $Wf(W)$ for some regular function f as

$$\underline{Wf(W)} := Wf(W) - \widetilde{\mathbf{E}}\widetilde{W}(\partial_{\widetilde{W}} f)(W), \qquad \widetilde{W} = \begin{pmatrix} 0 & \widetilde{X} \\ \widetilde{X}^* & 0 \end{pmatrix}, \quad \widetilde{X} \sim \mathrm{Gin}_{\mathbf{C}}, \tag{7.59}$$

with $\widetilde{X}$ independent of X. The choice of defining the self-renormalisation in terms of the complex rather than real Ginibre ensemble has the consequence that an additional error term needs to be estimated. For real Ginibre we have

$$\mathbf{E}\, WG = -\mathbf{E}\,\mathcal{S}[G]G - \mathbf{E}\,\mathcal{T}[G]G, \quad \mathcal{T}\left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right] = \frac{1}{n}\begin{pmatrix} 0 & c^t \\ b^t & 0 \end{pmatrix},$$

but the renormalisation comprises only the $\mathcal{S}[G]$ term, i.e.

$$\underline{WG} = WG + \mathbf{E}\,\mathcal{S}[G]G,$$

thus the $\mathcal{T}$-term needs to be estimated. By the Ward identity $GG^* = G^*G = \eta^{-1}\Im G$ it follows that

$$|\langle \boldsymbol{x}, \mathcal{T}[G]G\boldsymbol{y}\rangle| \leq \frac{1}{n}\sqrt{\langle \boldsymbol{x}, GG^*\boldsymbol{x}\rangle}\sqrt{\langle \boldsymbol{y}, G^*G\boldsymbol{y}\rangle} = \frac{1}{n\eta}\sqrt{\langle \boldsymbol{x}, \Im G\boldsymbol{x}\rangle}\sqrt{\langle \boldsymbol{y}, \Im G\boldsymbol{y}\rangle} \prec \frac{\Lambda + \rho}{n\eta}, \tag{7.60}$$

where $\rho := \pi^{-1}\Im m$ from (7.27). By [84, Theorem 4.1] it follows that

$$|\langle \boldsymbol{x}, (WG + \mathcal{S}[G]G + \mathcal{T}[G]G)\boldsymbol{y}\rangle| \prec \sqrt{\frac{\rho + \Lambda}{n\eta}}, \quad |\langle A(WG + \mathcal{S}[G]G + \mathcal{T}[G]G)\rangle| \prec \frac{\rho + \Lambda}{n\eta}$$

and therefore, together with the bound (7.60) on the $\mathcal{T}$-term we obtain

$$|\langle \boldsymbol{x}, \underline{WG}\boldsymbol{y}\rangle| \prec \sqrt{\frac{\rho + \Lambda}{n\eta}}, \quad |\langle A\underline{WG}\rangle| \prec \frac{\rho + \Lambda}{n\eta}. \tag{7.61}$$

We now consider the *stability operator* $\mathcal{B} := 1 - M\mathcal{S}[\cdot]M$ which expresses the stability of (7.56) against small perturbations. Since $\mathcal{S}$ only depends on the four block traces of the input matrix, and M is a multiple of the identity matrix in each block, the operator $\mathcal{B}$ can be understood as an operator acting on 2×2 matrices after taking a *partial trace*. Henceforth for all practical purposes we may identify $\mathcal{B}$ with this four dimensional operator. Written as a 4×4 matrix, it is given by

$$\mathcal{B} = \begin{pmatrix} B_1 & 0 \\ B_2 & 1 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 - u^2|z|^2 & -m^2 \\ -m^2 & 1 - u^2|z|^2 \end{pmatrix}, \quad B_2 = \begin{pmatrix} muz & muz \\ mu\overline{z} & mu\overline{z} \end{pmatrix}, \tag{7.62}$$

with m, u defined in (7.26)–(7.28). Here the rows and columns of $\mathcal{B}$ are ordered in such a way that 2×2 matrices are mapped to vectors as in

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow \begin{pmatrix} a \\ d \\ b \\ c \end{pmatrix}.$$

We first record some spectral properties of $\mathcal{B}$ in the following lemma, the proof of which we defer to the end of the section. Note that $\mathcal{B}^*$ refers to the adjoint of $\mathcal{B}$ with respect to the scalar product $\langle A, B \rangle = (2n)^{-1} \mathrm{Tr} A^* B$, for any deterministic matrices $A, B \in \mathbf{C}^{2n \times 2n}$.

Lemma 7.5.2. *Let $w \in \mathbf{H}$, $z \in \mathbf{C}$ be bounded spectral parameters, $|w| + |z| \lesssim 1$. Then the operator $\mathcal{B}$ has the trivial eigenvalues 1 with multiplicity 2, and furthermore has two non-trivial eigenvalues, and left and right eigenvectors*

$$\begin{aligned} \mathcal{B}[E_-] &= (1 + m^2 - u^2|z|^2) E_- \quad \mathcal{B}^*[E_-] = \overline{(1 + m^2 - u^2|z|^2)} E_-, \\ \mathcal{B}[V_r] &= (1 - m^2 - u^2|z|^2) V_r, \quad \mathcal{B}^*[V_l] = \overline{(1 - m^2 - u^2|z|^2)} V_l, \end{aligned}$$

where $E_- := (E_1 - E_2)/\sqrt{2}$ and

$$E_1 := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad E_2 := \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad V_r := \begin{pmatrix} m^2 + u^2|z|^2 & -2muz \\ -2mu\overline{z} & m^2 + u^2|z|^2 \end{pmatrix}, \quad V_l := \frac{1}{\overline{\langle V_r \rangle}}. \tag{7.63}$$

Moreover, for the second non-trivial eigenvalue we have the lower bound

$$|1 - m^2 - u^2|z|^2| \gtrsim \begin{cases} \Im m, & |1 - |z|| \geq \epsilon, \\ (\Im m)^2, & |1 - |z|| < \epsilon. \end{cases} \tag{7.64}$$

Corresponding to the two non-trivial eigenvalues of $\mathcal{B}$ we define the *spectral projections*

$$\mathcal{P}_* := \langle E_-, \cdot \rangle E_-, \quad \mathcal{P} := \langle V_l, \cdot \rangle V_r, \quad \mathcal{Q}_* := 1 - \mathcal{P}_*, \quad \mathcal{Q} := 1 - \mathcal{P}_* - \mathcal{P}.$$

From (7.56) and (7.58) it follows that

$$\mathcal{B}[G - M] = M \mathcal{S}[G - M](G - M) - M \underline{WG}. \tag{7.65}$$

We now distinguish the two cases $\rho \sim 1$ and $\rho \ll 1$. In the former we obtain

$$\|\mathcal{Q}_* \mathcal{B}^{-1}\|_{\|\cdot\| \to \|\cdot\|} \lesssim \frac{1}{|1 - m^2 - u^2|z|^2|} \lesssim 1 \tag{7.66}$$

by (7.64). Since $\langle E_-, G\rangle = \langle E_-, M\rangle = 0$ by block symmetry, it follows that

$$G - M = \mathcal{Q}_*[G-M] = \mathcal{Q}_*\mathcal{B}^{-1}\mathcal{B}[G-M]$$

and thus

$$\begin{aligned}\langle \boldsymbol{x}, (G-M)\boldsymbol{y}\rangle &= \operatorname{Tr}\left[(\mathcal{Q}_*\mathcal{B}^{-1})^*[\boldsymbol{x}\boldsymbol{y}^*]\right]^* \mathcal{B}[G-M] \\ &= \sum_{i=1}^{4}\langle \boldsymbol{x}_i, (M\mathcal{S}[G-M](G-M) - M\underline{WG})\boldsymbol{y}_i\rangle \\ &= \mathcal{O}_\prec\left(\xi\Lambda + \sqrt{\frac{\rho+\Lambda}{n\eta}}\right),\end{aligned} \tag{7.67a}$$

where we used that the image of $\boldsymbol{x}\boldsymbol{y}^*$ under $(\mathcal{Q}_*\mathcal{B}^{-1})^*$ is of rank at most 4, hence it can be written as $\sum_{i=1}^4 \boldsymbol{x}_i\boldsymbol{y}_i^*$ with vectors of bounded norm. Similarly, for general matrices A we find

$$\begin{aligned}\langle A(G-M)\rangle &= \langle\left[(\mathcal{Q}\mathcal{B}^{-1})^*[A^*]\right]^* \mathcal{B}[G-M]\rangle \\ &= \langle\left[(\mathcal{Q}_*\mathcal{B}^{-1})^*[A^*]\right]^* (M\mathcal{S}[G-M](G-M) - M\underline{WG})\rangle \\ &= \mathcal{O}_\prec\left(\frac{\rho+\Lambda}{n\eta} + \xi^2\right).\end{aligned} \tag{7.67b}$$

In the complementary case $\rho \ll 1$ we similarly decompose

$$G - M = \mathcal{P}[G-M] + \mathcal{P}_*[G-M] + \mathcal{Q}[G-M] = \theta V_r + \mathcal{Q}[G-M], \quad \theta := \langle V_l, G-M\rangle. \tag{7.68}$$

Now we apply $\mathcal{B}$ to both sides of (7.68) and take the inner product with V_l to obtain

$$\langle V_l, \mathcal{B}[G-M]\rangle = (1-m^2-u^2|z|^2)\theta + \langle V_l, \mathcal{B}\mathcal{Q}[G-M]\rangle \tag{7.69}$$

from (7.65). For the spectral projection $\mathcal{Q}$ we find

$$\mathcal{B}^{-1}\mathcal{Q} = \mathcal{Q}\mathcal{B}^{-1} = \begin{pmatrix} 0 & 0 \\ B_3 & 1\end{pmatrix}, \quad B_3 = \frac{mu}{m^2+u^2|z|^2}\begin{pmatrix} z & z \\ \overline{z} & \overline{z}\end{pmatrix}. \tag{7.70}$$

Thus it follows that

$$\|\mathcal{B}^{-1}\mathcal{Q}\|_{\|\cdot\|\to\|\cdot\|} \lesssim \frac{|muz|}{|m^2+u^2|z|^2|} \lesssim 1 \tag{7.71}$$

since in the regime $\rho \ll 1$ we have $|1-m^2-u^2|z|^2| \ll 1$ due to $|\Im u^2| \ll 1$ which follows by a simple calculation.

By using (7.65) in (7.69) it follows that

$$|\theta| \prec \frac{1}{\rho}\Big(\frac{\rho+\Lambda}{n\eta} + \xi^2\Big) \tag{7.72}$$

from (7.57), (7.61) since, due to $||z|-1| \gtrsim \epsilon$, we have $|1-m^2-u^2|z|^2| \geq \rho$ according to (7.64). For general vectors $\boldsymbol{x}, \boldsymbol{y}$ it follows from (7.68), (7.72) and inserting $1 = \mathcal{B}^{-1}\mathcal{B}$

similarly to (7.67) that

$$\begin{aligned}\langle \boldsymbol{x}, (G-M)\boldsymbol{y}\rangle &= \mathcal{O}_\prec\left(\frac{\rho+\Lambda}{\rho n\eta}+\frac{\xi^2}{\rho}\right) + \langle \Big[(\mathcal{Q}\mathcal{B}^{-1})^*[\boldsymbol{x}\boldsymbol{y}^*]\Big]^* \mathcal{B}[G-M]\rangle \\ &= \mathcal{O}_\prec\left(\frac{\rho+\Lambda}{\rho n\eta}+\frac{\xi^2}{\rho}\right) + \sum_{i=1}^4 \langle \boldsymbol{x}_i, (M\mathcal{S}[G-M](G-M) - M\underline{WG})\boldsymbol{y}_i\rangle \\ &= \mathcal{O}_\prec\left(\frac{\rho+\Lambda}{\rho n\eta}+\frac{\xi^2}{\rho} + \xi\Lambda + \sqrt{\frac{\rho+\Lambda}{n\eta}}\right),\end{aligned} \tag{7.73a}$$

and

$$\begin{aligned}\langle A(G-M)\rangle &= \mathcal{O}_\prec\left(\frac{\rho+\Lambda}{\rho n\eta}+\frac{\xi^2}{\rho}\right) + \langle \Big[(\mathcal{Q}\mathcal{B}^{-1})^*[A^*]\Big]^* \mathcal{B}[G-M]\rangle \\ &= \mathcal{O}_\prec\left(\frac{\rho+\Lambda}{\rho n\eta}+\frac{\xi^2}{\rho}\right) + \langle \Big[(\mathcal{Q}\mathcal{B}^{-1})^*[A^*]\Big]^* (M\mathcal{S}[G-M](G-M) - M\underline{WG})\rangle \\ &= \mathcal{O}_\prec\left(\frac{\rho+\Lambda}{\rho n\eta}+\frac{\xi^2}{\rho}\right).\end{aligned} \tag{7.73b}$$

By using the bounds in (7.67) and (7.73) in the two complementary regimes we improve the input bound in (7.57). We can iterate this procedure and obtain

$$|\langle \boldsymbol{x}, (G-M)\boldsymbol{y}\rangle| \prec \frac{1}{n\eta} + \sqrt{\frac{\rho}{n\eta}}, \qquad |\langle A(G-M)\rangle| \prec \frac{1}{n\eta}. \tag{7.74}$$

In order to make sure the iteration yields an improvement one needs an priori bound on ξ of the form $\xi \ll 1$ since otherwise ξ^2 is difficult to control. For large η such an a priori bound is trivially available which can then be iteratively bootstrapped by monotonicity down to the optimal $\eta \gg n^{-1}$. For details on this standard argument the reader is referred to e.g. [15, Section 3.3]. Then the local law for any $\eta > 0$ readily follows by exactly the same argument as in [59, Appendix A]. This completes the proof of Theorem 7.3.1. □

Proof of Lemma 7.5.2. The fact that $\mathcal{B}$ has the eigenvalue 1 with multiplicity 2, and the claimed form of the remaining two eigenvalues and corresponding eigenvectors can be checked by direct computations. Taking the imaginary part of (7.26) we have

$$(1-|m|^2-|u|^2|z|^2)\Im m = (|m|^2+|u|^2|z|^2)\Im w, \tag{7.75}$$

which implies

$$|m|^2+|u|^2|z|^2 < 1, \qquad \lim_{\Im w\to 0}(|m|^2+|u|^2|z|^2) = 1, \quad \Re w \in \overline{\operatorname{supp}\rho} \tag{7.76}$$

as $\Im m$ and $\Im w$ have the same sign. Here $\operatorname{supp}\rho$ should be understood as the support of the self-consistent density of states, as defined in (7.27), restricted to the real axis. The second bound in (7.64) then follows from (7.76) and

$$|1-m^2-u^2|z|^2| \ge \Re(1-m^2-u^2|z|^2) = 1-(\Re m)^2+(\Im m)^2-\Re(u^2)|z|^2 \gtrsim (\Im m)^2. \tag{7.77}$$

The bound (7.77) can be improved in the case $\rho \ll 1$ if w is near a *regular edge* of ρ, i.e. where ρ locally vanishes as a square-root. According to [61, Eq. (15b)] the density ρ has two regular edges $\pm\sqrt{\mathfrak{e}_+}$ if $|z| \le 1-\epsilon$, and four regular edges in $\pm\sqrt{\mathfrak{e}_+}, \pm\sqrt{\mathfrak{e}_-}$ for $|z| \ge 1+\epsilon$, where

$$\mathfrak{e}_\pm := \frac{8(1-|z|^2)^2 \pm (1+8|z|^2)^{3/2} - 36(1-|z|^2) + 27}{8|z|^2} \gtrsim 1.$$

By the explicit form of $\mathfrak{e}_\pm$ it follows that $\mathfrak{e}_\pm \gtrsim 1$ whenever $|1-|z|| \ge \epsilon$. In contrast, if $|z| = 1$, then ρ has a *cusp* singularity in 0 where it locally vanishes like a cubic root. Near a regular edge we have $\Im m \lesssim \sqrt{\Im w}$, and therefore from (7.75)

$$(1-|m|^2-|u|^2|z|^2) \gtrsim \sqrt{\Im w} \gtrsim \Im m$$

and it follows that

$$|1-m^2-u^2|z|^2| \gtrsim \Im m,$$

proving also the first inequality in (7.64). □

7.6 CLT for resolvents: Proof of Proposition 7.3.3

The goal of this section is to prove the CLT for resolvents, as stated in Proposition 7.3.3. The proof is very similar to [58, Section 6] and we focus on the differences specific to the real case. Within this section we consider resolvents $G_1, \dots, G_p$ with $G_i = G^{z_i}(\mathrm{i}\eta_i)$ and $\eta_i \ge n^{-1}$. As a first step we recall the leading-order approximation of $G = G_i$

$$\langle G - M\rangle = -\langle \underline{WGA}\rangle + \mathcal{O}_\prec\left(\frac{1}{|\beta|(n\eta)^2}\right), \quad A := (\mathcal{B}^*)^{-1}[1]^*M \tag{7.78}$$

from [58, Eq. (6.9)], where the stability operator $\mathcal{B}$ has been defined in (7.62). Here β is the eigenvalue of $\mathcal{B}$ with eigenvector $(1,1,0,0)$ and is bounded by (see [58, Eq. (6.8b)])

$$|\beta| \gtrsim |1-|z|| + \eta^{2/3}. \tag{7.79}$$

One important input for the proof of Proposition 7.3.3 is a lower bound on the eigenvalues of the stability operator $\widehat{\mathcal{B}}$, defined in (7.38), the proof of which we defer to the end of the section. Note that the two-body stability operator $\widehat{\mathcal{B}}$ and its eigenvalues $\widehat{\beta}, \widehat{\beta}_*$ are consistently decorated by hats $\widehat{(\cdot)}$ to distinguish them from their one-body analogues $\mathcal{B}, \beta$. We will consistently equip $\mathcal{B}, \widehat{\mathcal{B}}$ and their eigenvalues, $\beta, \widehat{\beta}, \widehat{\beta}_*$ with indices when instead of M they are defined with the help of $M_i = M^{z_i}(w_i)$; e.g. $\widehat{\beta}_*^{1i}$ is the lowest eigenvalue of $\widehat{\mathcal{B}}_{1i} = \widehat{\mathcal{B}}(z_1, z_i, w_1, w_i)$ defined analogously to (7.38).

Lemma 7.6.1. *For $z_1, z_2 \in \mathbf{C}$, $w_1, w_2 \in \mathbf{C}\setminus\mathbf{R}$ such that $|z_i|, |w_i| \lesssim 1$ the two non-trivial eigenvalues $\widehat{\beta}, \widehat{\beta}_*$ of $\widehat{\mathcal{B}}$ satisfy*

$$\min\{\Re\widehat{\beta}, \Re\widehat{\beta}_*\} \gtrsim |z_1-z_2|^2 + \min\{|w_1+\overline{w_2}|, |w_1-\overline{w_2}|\}^2 + |\Im w_1| + |\Im w_2| \tag{7.80}$$

Proof of Proposition 7.3.3. The proof of Proposition 7.3.3 goes in two steps. First, we use (7.78) and a cumulant expansion in order to prove the asymptotic representation of the expectation in (7.35). In the second step we then turn to the computation of higher moments and establish an asymptotic Wick theorem in the form of (7.32).

We use the notation Δ^{ab} for the matrix $(\Delta^{ab})_{cd} = \delta_{ac}\delta_{bd}$ and decompose $W = \sum_{ab} w_{ab}\Delta^{ab}$. For each a, b we then perform a cumulant expansion and obtain

$$\mathbf{E}\langle \underline{WG}A\rangle = -\frac{1}{n}\sum_{ab}{}' \mathbf{E}\langle \Delta^{ab}G\Delta^{ab}GA\rangle + \sum_{k\geq 2}\sum_{ab}\sum_{\boldsymbol{\alpha}\in\{ab,ba\}^k} \frac{\kappa(ab,\boldsymbol{\alpha})}{k!}\mathbf{E}\,\partial_{\boldsymbol{\alpha}}\langle \Delta^{ab}GA\rangle, \tag{7.81}$$

which has an additional term compared to the complex case [58, Eq. (6.11)] since the self-renormalisation (7.59) was chosen such that it only takes the $\kappa(ab, ba) = 1$ and not the $\kappa(ab, ab) = 1$ cumulant into account. Here $\kappa(ab, cd, ef, \ldots)$ denotes the joint cumulant of the random variables $w_{ab}, w_{cd}, w_{ef}, \ldots$, and we denote partial derivatives by $\partial_{\boldsymbol{\alpha}} := \partial_{w_{\alpha_1}}\cdots\partial_{w_{\alpha_k}}$ for tuples $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$, with $\alpha_i \in [n]\times[n]$. In (7.81) we introduced the notation

$$\sum_{ab}{}' := \sum_{a\leq n}\sum_{b>n} + \sum_{a>n}\sum_{b\leq n}.$$

We note that by Assumption (7.A) the cumulants $\kappa(\alpha_1, \ldots, \alpha_k)$ satisfy the scaling

$$|\kappa(\alpha_1,\ldots,\alpha_k)| \lesssim n^{-k/2}. \tag{7.82}$$

For the second term in (7.81) we find exactly as in [58, Eqs. (6.10)-(6.13)] that

$$\sum_{k\geq 2}\sum_{ab}\sum_{\boldsymbol{\alpha}\in\{ab,ba\}^k} \frac{\kappa(ab,\boldsymbol{\alpha})}{k!}\partial_{\boldsymbol{\alpha}}\langle \Delta^{ab}GA\rangle = \frac{\mathrm{i}\kappa_4}{4n}\partial_\eta(m^4) + \mathcal{O}_\prec\Big(\frac{1}{|\beta|}\Big(\frac{1}{n^{3/2}(1+\eta)} + \frac{1}{(n\eta)^2}\Big)\Big). \tag{7.83}$$

For the first term in (7.81), which is new compared to [58, Eq. (6.11)], we rewrite

$$\frac{1}{n}\sum_{ab}{}'\langle \Delta^{ab}G\Delta^{ab}GA\rangle = \frac{1}{n}\langle GAEG^tE'\rangle = \frac{1}{n}\langle G^zAEG^{\overline{z}}E'\rangle,$$

where we used that $(G^z)^t = G^{\overline{z}}$, and the convention that formulas containing (E, E') are understood so that the matrices E, E' are summed over the assignments $(E, E') = (E_1, E_2)$ and $(E, E') = (E_2, E_1)$ with

$$E_1 := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad E_2 := \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

From the local law [58, Theorem 5.2] for products of resolvents and the bound on $|\widehat{\beta}_*|$ from Lemma 7.6.1 we can thus conclude

$$\begin{aligned} \frac{1}{n}\sum_{ab}{}'\langle \Delta^{ab}G\Delta^{ab}GA\rangle &= \frac{1}{n}\langle M^{z,\overline{z}}_{AE}E'\rangle + \mathcal{O}_\prec\Big(\frac{1}{|z-\overline{z}|^2}\frac{1}{(n\eta)^2}\Big) \\ &= \frac{m}{n}\frac{m^4 + m^2u^2|z|^2 - 2u^4|z|^4 + 2u^2(x^2-y^2)}{(1-m^2-u^2|z|^2)(1+u^4|z|^4 - m^4 - 2u^2(x^2-y^2))} \\ &\quad + \mathcal{O}_\prec\Big(\frac{1}{|z-\overline{z}|^2}\frac{1}{(n\eta)^2}\Big), \end{aligned} \tag{7.84}$$

where $z = x + \mathrm{i}y$, and the second step follows by explicitly computing the inverse

$$M^{z,\overline{z}}_{AE} = (1 - M^z\mathcal{S}[\cdot]M^{\overline{z}})^{-1}[M^zAEM^{\overline{z}}]$$

in terms of the entries of M, noting that $m^z = m^{\overline{z}}$ and $u^z = u^{\overline{z}}$. Then, using the definition $v := -\mathrm{i}m > 0$ and that

$$|z|^2u^2 + v^2 = u, \qquad u' = -\frac{2uv}{1+u-|z|^2u^2}, \qquad v^2 = u(1-|z|^2u)$$

we obtain

$$\begin{aligned} &m\frac{m^4+m^2u^2|z|^2-2u^4|z|^4+2u^2(x^2-y^2)}{(1-m^2-u^2|z|^2)(1+u^4|z|^4-m^4-2u^2(x^2-y^2))} \\ &\quad= -\frac{\mathrm{i}u'}{2}\frac{u-3|z|^2u^2+2u(x^2-y^2)}{1-u^2+2u^3|z|^2-2u^2(x^2-y^2)}. \end{aligned} \tag{7.85}$$

Now (7.35) follows from combining (7.78) and (7.81)–(7.85).

We now turn to the computation of higher moments for which we recall from (7.78) and (7.35) that

$$\begin{aligned} \prod_{i\in[p]}\langle G_i - \mathbf{E}\,G_i\rangle &= \prod_{i\in[p]}\langle G_i - M_i - \mathcal{E}_i\rangle + \mathcal{O}_\prec\Big(\frac{\psi}{n\eta}\Big) \\ &= \prod_{i\in[p]}\langle -\underline{WG_i}A_i - \mathcal{E}_i\rangle + \mathcal{O}_\prec\Big(\frac{\psi}{n\eta}\Big) \end{aligned} \tag{7.86}$$

with A_i as in (7.78) and $\mathcal{E}_i$ as in (7.36), and

$$\psi := \prod_i\Big(\frac{1}{|\beta_i|}+\frac{1}{(\Im z_i)^2}\Big)\frac{1}{n\eta_i} \le \prod_i\Big(\frac{1}{|1-|z_i||}+\frac{1}{(\Im z_i)^2}\Big)\frac{1}{n\eta_i} \tag{7.87}$$

with the bound on β_i from (7.79). We begin with the cumulant expansion of $\underline{WG_1}$ to obtain

$$\begin{aligned} &\mathbf{E}\prod_{i\in[p]}\langle -\underline{WG_i}A_i - \mathcal{E}_i\rangle \\ &\quad= \mathbf{E}\Big(\frac{1}{n}\sum_{ab}{}'\langle\Delta^{ab}G_1\Delta^{ab}G_1A_1\rangle - \langle\mathcal{E}_1\rangle\Big)\prod_{i\neq 1}\langle -\underline{WG_i}A_i - \mathcal{E}_i\rangle \\ &\qquad+ \sum_{i\neq 1}\mathbf{E}\,\widehat{\mathbf{E}}\langle\widehat{W}G_1A_1\rangle\langle\widehat{W}G_iA_i - \underline{WG_i\widehat{W}G_i}A_i\rangle\prod_{j\neq 1,i}\langle -\underline{WG_j}A_j - \mathcal{E}_j\rangle \\ &\qquad+ \sum_{k\ge 2}\sum_{ab}\sum_{\boldsymbol{\alpha}\in\{ab,ba\}^k}\frac{\kappa(ba,\boldsymbol{\alpha})}{k!}\mathbf{E}\,\partial_{\boldsymbol{\alpha}}\Big[\langle -\Delta^{ba}G_1A_1\rangle\prod_{i\neq 1}\langle -\underline{WG_i}A_i - \mathcal{E}_i\rangle\Big], \end{aligned} \tag{7.88}$$

where, compared to [58, Eq. (6.17)], the first line on the rhs. has an additional term specific to the real case, and $\widehat{W}$, as opposed to $\widetilde{W}$ in (7.59), is the Hermitisation of an independent real Ginibre matrix $\widehat{X}$ with expectation $\widehat{\mathbf{E}}$. The expansion of the third line on the rhs. of (7.88) is completely analogous to [58] since for cumulants of degree at least three nothing specific to the complex case was used. Therefore we obtain, from combining[2] [58, Eqs.

[2]Note that the definition of $\mathcal{E}$ in [58, Eq. (6.8c)] differs from (7.36) in the present paper.

(6.26), (6.29)], that

$$\sum_{k\geq 2}\sum_{ab}\sum_{\boldsymbol{\alpha}\in\{ab,ba\}^k}\frac{\kappa(ba,\boldsymbol{\alpha})}{k!}\mathbf{E}\,\partial_{\boldsymbol{\alpha}}\Big[\langle-\Delta^{ba}G_1A_1\rangle\prod_{i\neq 1}\langle-\underline{WG_i}A_i-\mathcal{E}_i\rangle\Big]$$
$$=-\frac{\mathrm{i}\kappa_4}{4n}\partial_{\eta_1}(m_1^4)\,\mathbf{E}\prod_{i\neq 1}\langle-\underline{WG_i}A_i-\mathcal{E}_i\rangle+\sum_{i\neq 1}\frac{\kappa_4U_1U_i}{2n^2}\,\mathbf{E}\prod_{j\neq 1,i}\langle-\underline{WG_j}A_j-\mathcal{E}_j\rangle \tag{7.89}$$
$$+\mathcal{O}\bigg(\frac{n^\xi\psi}{\sqrt{n\eta_*}}\bigg),$$

where

$$U_i:=-\sqrt{2}\langle M_i\rangle\langle M_iA_i\rangle=\frac{\mathrm{i}}{\sqrt{2}}\partial_{\eta_i}m_i^2.$$

Recall the definition of $\mathcal{E}_i$ in (7.36), then using (7.84)–(7.85) and (7.89) in (7.88) we thus have

$$\mathbf{E}\prod_{i\in[p]}\langle-\underline{WG_i}A_i-\mathcal{E}_i\rangle$$
$$=\sum_{i\neq 1}\mathbf{E}\Big(\frac{\kappa_4U_1U_i}{2n^2}+\widehat{\mathbf{E}}\langle\widehat{W}G_1A_1\rangle\langle\widehat{W}G_iA_i-\underline{WG_i\widehat{W}G_i}A_i\rangle\Big)\prod_{j\neq 1,i}\langle-\underline{WG_j}A_j-\mathcal{E}_j\rangle$$
$$+\mathcal{O}\bigg(\frac{n^\xi\psi}{\sqrt{n\eta_*}}\bigg). \tag{7.90}$$

It remains to consider the variance term in (7.90) for which we use the identity

$$\widehat{\mathbf{E}}\langle\widehat{W}A\rangle\langle\widehat{W}B\rangle=\frac{1}{2n^2}\langle AE(B+B^t)E'\rangle=\frac{\langle AE_1(B+B^t)E_2\rangle+\langle AE_2(B+B^t)E_1\rangle}{2n^2} \tag{7.91}$$

in order to compute

$$\widehat{\mathbf{E}}\langle\widehat{W}G_1A_1\rangle\langle\widehat{W}G_iA_i-\underline{WG_i\widehat{W}G_i}A_i\rangle$$
$$=\frac{1}{2n^2}\langle G_1A_1E(G_iA_i+A_i^tG_i^t)E'-G_1A_1E(\underline{G_iA_iWG_i}+\underline{G_i^tWA_i^tG_i^t})E'\rangle, \tag{7.92}$$

where, compared to [58, Eqs. (6.18)-(6.19)], there is an additional term with transposition. Here the self-renormalisation e.g. in $\underline{G_iA_iWG_i}$ is defined analogously to (7.59) with the derivative acting on both G_i's. For the second term in (7.92) we identify the leading order contribution using the fact that $G^z(w)^t=G^{\overline{z}}(w)$ and denoting $G_{\bar{i}}=G^{\overline{z_i}}(\mathrm{i}\eta_i)$ as

$$\langle G_1A_1E(\underline{G_iA_iWG_i}+\underline{G_i^tWA_i^tG_{\bar{i}}^t})E'\rangle$$
$$=-\langle G_1\mathcal{S}[G_1A_1EG_iA_i]G_iE'+G_1\mathcal{S}[G_1A_1EG_{\bar{i}}]A_i^tG_{\bar{i}}E'\rangle \tag{7.93}$$
$$+\langle\underline{G_1A_1EG_iA_iWG_iE'}+\underline{G_1A_1EG_{\bar{i}}WA_i^tG_{\bar{i}}E'}\rangle$$

for which we use the local law from Theorem 7.3.5 to conclude that the main terms in (7.92) are

$$\langle G_1A_1E(G_iA_i+A_i^tG_{\bar{i}})E'+G_1\mathcal{S}[G_1A_1EG_iA_i]G_iE'+G_1\mathcal{S}[G_1A_1EG_{\bar{i}}]A_i^tG_{\bar{i}}E'\rangle$$
$$=\widehat{V}_{1,i}+\mathcal{O}_\prec\bigg(\frac{1}{n|\widehat{\beta}_*^{1i}|^2\eta_*^{1i}|\eta_1\eta_i|^{1/2}}+\frac{1}{n^2|\widehat{\beta}_*^{1i}|^2(\eta_*^{1i})^2|\eta_1\eta_i|}\bigg)$$
$$\widehat{V}_{1,i}:=\langle M_{A_1E}^{z_1,z_i}A_iE'+M_{A_1EA_i^t}^{z_1,\overline{z_i}}E'+\mathcal{S}[M_{A_1E}^{z_1,z_i}A_i]M_{E'}^{z_i,z_1}+\mathcal{S}[M_{A_1E}^{z_1,\overline{z_i}}]A_i^tM_{E'}^{\overline{z_i},z_1}\rangle, \tag{7.94}$$

where $|\widehat{\beta}_*^{1i}| \gtrsim |z_1 - z_i|^2$ from Lemma 7.6.1, and $\eta_*^{1i} := \min\{\eta_1, \eta_i\}$. By an explicit computation similarly to [58, Eq. (6.23)] it follows that

$$\widehat{V}_{1,i} = V(z_1, z_i, \eta_1, \eta_i) + V(z_1, \overline{z_i}, \eta_1, \eta_i) \tag{7.95}$$

with V being exactly as in the complex case, i.e. as in (7.34). For the error term in (7.93) we claim that

$$\mathbf{E}|\langle \underline{G_1 A_1 E G_i A_i W G_i E'}\rangle|^2 + \mathbf{E}|\langle \underline{G_1 A_1 E G_{\bar{i}} W A_i^t G_{\bar{i}} E'}\rangle|^2 \lesssim \Big(\frac{1}{n\eta_1\eta_i\eta_*^{1i}}\Big)^2. \tag{7.96}$$

The CLT for resolvents, as stated in (7.32) follows from inserting (7.92)–(7.96) into (7.90), and iteration of (7.90) for the remaining product.

In order to conclude the proof of Proposition 7.3.3 it remains to prove (7.96). Introduce the shorthand notation G_{i1i} for generic finite sums of products of G_i, G_1, G_i (or $G_{\bar{i}}$ in place of G_i) with arbitrary bounded deterministic matrices, e.g. $G_i E' G_1 A_1 E G_i A_i$ appearing in the first term in (7.96). We will prove the more general claim

$$\mathbf{E}|\langle \underline{W G_{i1i}}\rangle|^2 \lesssim \Big(\frac{1}{n\eta_1\eta_i\eta_*^{1i}}\Big)^2. \tag{7.97}$$

The proof is similar to [58, Eq. (6.32)]. Therefore we focus on the differences. In the cumulant expansion of (7.97) there is an additional term compared to [58, Eq. (6.33)] given by

$$\begin{aligned} &\frac{1}{n}\sum_{ab}{}' \mathbf{E}\langle \Delta^{ab} G_i \Delta^{ab} G_{i1i} + \Delta^{ab} G_{i1} \Delta^{ab} G_{1i} + \Delta^{ab} G_{i1i} \Delta^{ab} G_i\rangle \langle \underline{W G_{i1i}}\rangle \\ &\quad = \frac{1}{n}\mathbf{E}\langle G_{1iii} + G_{i1i1}\rangle\langle \underline{W G_{i1i}}\rangle, \end{aligned} \tag{7.98}$$

where we combined two terms of type G_{1iii} into one since in our convention G_{1iii} is a shorthand notation for generic sums of products. We now perform another cumulant expansion of (7.98) to obtain

$$\begin{aligned} &\frac{1}{n}\mathbf{E}\langle G_{1iii} + G_{i1i1}\rangle\langle \underline{W G_{i1i}}\rangle \\ &= \frac{1}{n^2}\mathbf{E}\langle G_{1iii} + G_{i1i1}\rangle^2 \\ &\quad + \frac{1}{n}\mathbf{E}\,\widetilde{\mathbf{E}}\langle \widetilde{W}(G_{1iii1} + G_{iii1i} + G_{ii1ii} + G_{i1iii} + G_{1i1i1} + G_{i1i1i})\rangle\langle \widetilde{W} G_{i1i}\rangle \\ &\quad + \sum_{k\ge 2}\mathcal{O}\Big(\frac{1}{n^{(k+3)/2}}\Big)\sum_{ab}{}'\sum_{\boldsymbol{\alpha}\in\{ab,ba\}^k}\mathbf{E}\,\partial_{\boldsymbol{\alpha}}\Big[\langle G_{1iii} + G_{i1i1}\rangle\langle \Delta^{ab} G_{i1i}\rangle\Big], \end{aligned} \tag{7.99}$$

where the first line on the rhs. corresponds to the term where the remaining W acts on G_{i1i} within its own trace as in (7.98), and in the last line we used the scaling bound (7.82) for κ. In order to estimate (7.98) we recall [58, Lemma 5.8].

Lemma 7.6.2. *Let $w_1, w_2, \ldots, z_1, z_2, \ldots,$ denote arbitrary spectral parameters with $\eta_i = \Im w_i > 0$. Let $G_j = G^{z_j}(w_j)$, then with $G_{j_1\ldots j_k}$ we denote generic products of resolvents $G_{j_1}, \ldots, G_{j_k}$, or their adjoints/transpositions (in that order, each G_{j_i} appears exactly once) with bounded deterministic matrices in between, e.g. $G_{1i1} = A_1 G_1 A_2 G_i A_3 G_1 A_4$.*

(i) For $j_1, \dots j_k$ we have the isotropic bound

$$|\langle \boldsymbol{x}, G_{j_1\dots j_k}\boldsymbol{y}\rangle| \prec \|\boldsymbol{x}\|\|\boldsymbol{y}\|\sqrt{\eta_{j_1}\eta_{j_k}}\Big(\prod_{n=1}^{k}\eta_{j_n}\Big)^{-1}. \tag{7.100a}$$

(ii) For $j_1, \dots, j_k$ and any $1 \le s < t \le k$ we have the averaged bound

$$|\langle G_{j_1\dots j_k}\rangle| \prec \sqrt{\eta_{j_s}\eta_{j_t}}\Big(\prod_{n=1}^{k}\eta_{j_n}\Big)^{-1}. \tag{7.100b}$$

Since only η_1, η_i play a role within the proof of (7.96), we drop the indices from η_*^{1i} and use the notation $\eta_* = \eta_*^{1i}$. For the first term in (7.99) we use (7.100b) to obtain

$$\frac{1}{n^2}|\langle G_{1iii} + G_{i1i1}\rangle|^2 \prec \frac{1}{n^2\eta_1^2\eta_i^2\eta_*^2}. \tag{7.101}$$

Similarly for the second term we use (7.91) and again (7.100b) to bound it by

$$\begin{aligned}&\frac{1}{n}|\widetilde{\mathbf{E}}\langle\widetilde{W}(G_{1iii1} + G_{iii1i} + G_{ii1ii} + G_{i1iii} + G_{1i1i1} + G_{i1i1i})\rangle\langle\widetilde{W}G_{i1i}\rangle| \\ &\quad\prec \frac{1}{n^3\eta_1^2\eta_i^2\eta_*^3} \le \frac{1}{n^2\eta_1^2\eta_i^2\eta_*^2}\end{aligned} \tag{7.102}$$

since $\eta_* \ge 1/n$. Finally, for the last term of (7.99) we estimate

$$\left|\mathcal{O}\Big(\frac{1}{n^{(k+7)/2}}\Big)\sum_{ab}{}' \sum_c \sum_{\boldsymbol{\alpha}} \partial_{\boldsymbol{\alpha}}\Big[(G_{1iii} + G_{i1i1})_{cc}(G_{i1i})_{ba}\Big]\right| \prec \frac{1}{n^2\eta_1^2\eta_i^2\eta_*^2} \tag{7.103}$$

for any $k \ge 2$. Indeed, for $k \ge 3$ the claim (7.103) follows trivially from (7.100a) and the observation that the bound (7.100a) remains invariant under the action of derivatives. Indeed, differentiating a term like $(G_{i1i})_{ab}$ gives rise to the terms $(G_i)_{aa}(G_{i1i})_{bb}, (G_{i1})_{ab}(G_{1i})_{ab}, \dots$ for all of which (7.100a) gives the same estimate as for $(G_{i1i})_{ab}$ since the presence of an additional factor of G_1 or G_i is compensated by the fact that the same type of G appears two additional times as the first or last factor in some product. For the $k = 2$ case we observe that by parity at least one factor will be off-diagonal in the sense that it has two distinct summation indices from $\{a, b, c\}$ giving rise to an additional factor of $(n\eta_*)^{-1/2}$ by summing up one of the indices with the Ward identity. For example, for the term with $(G_{1iii})_{cc}(G_{i1})_{bb}(G_{1i})_{aa}(G_i)_{ba}$ we estimate

$$\begin{aligned}n^{-9/2}\left|\sum_{ab}{}'\sum_c (G_{1iii})_{cc}(G_{i1})_{bb}(G_{1i})_{aa}(G_i)_{ba}\right| &\prec n^{-9/2}\frac{n}{\eta_1^{3/2}\eta_i^{7/2}}\sum_{ab}{}'|(G_i)_{ba}| \\ &\le n^{-3}\frac{1}{\eta_1^{3/2}\eta_i^{7/2}}\sum_b\sqrt{\sum_a |(G_i)_{ba}|^2} \\ &= n^{-3}\frac{1}{\eta_1^{3/2}\eta_i^4}\sum_b\sqrt{(\Im G_i)_{bb}} \prec \frac{1}{n^2\eta_1^{3/2}\eta_i^4}.\end{aligned}$$

Thus, in general we obtain a bound of

$$\frac{1}{n^{3/2}}\Big(\frac{1}{\eta_1^{3/2}\eta_i^{7/2}} + \frac{1}{\eta_1^{5/2}\eta_i^{5/2}}\Big)\frac{1}{\sqrt{n\eta_*}} \lesssim \frac{1}{n^2\eta_1^2\eta_i^2\eta_*^2}.$$

By combining (7.101)–(7.103) we obtain a bound of $(n\eta_1\eta_i\eta_*)^{-2}$ on the additional term (7.98). The remaining terms can be estimated as in [58, Eq. (6.32)] and we conclude the proof of (7.97) and thereby Proposition 7.3.3. □

Proof of Lemma 7.6.1. The claim (7.80) is equivalent to the claim

$$\max\{\Re\tau,\Re\tau_*\}\le 1-c\big[|z_1-z_2|^2+\min\{|w_1+\overline{w_2}|^2,|w_1-\overline{w_2}|^2\}+|\Im w_1|+|\Im w_2|\big],\quad c>0, \tag{7.104}$$

where τ,τ_* are the eigenvalues of the matrix

$$R:=\begin{pmatrix} z_1\overline{z_2}u_1u_2 & m_1m_2\\ m_1m_2 & \overline{z_1}z_2u_1u_2\end{pmatrix}, \tag{7.105}$$

thus $\widehat{\beta}=1-\tau$, $\widehat{\beta}_*=1-\tau_*$. We first check that (7.104) holds true ineffectively, i.e. with $c=0$. We claim that

$$\max\Re\operatorname{Spec}(A)\le\lambda_{\max}\Big(\frac{A+A^*}{2}\Big):=\max\operatorname{Spec}\Big(\frac{A+A^*}{2}\Big) \tag{7.106}$$

holds for any square matrix A. Indeed, suppose that $A\boldsymbol{x}=\lambda\boldsymbol{x}$, $\|\boldsymbol{x}\|=1$ and $(A+A^*)/2\le M$ in the sense of quadratic forms. We then compute

$$0\ge\Big\langle\boldsymbol{x},\Big(\frac{A+A^*}{2}-M\Big)\boldsymbol{x}\Big\rangle=\frac{\langle\boldsymbol{x},A\boldsymbol{x}\rangle+\langle A\boldsymbol{x},\boldsymbol{x}\rangle}{2}-M=\Re\lambda-M,$$

from which (7.106) follows by choosing M to be the largest eigenvalue of $(A+A^*)/2$.

Since R is such that its entrywise real part is given by $\Re R=(R+R^*)/2$, from (7.106) we conclude the chain of inequalities

$$\max\{\Re\tau,\Re\tau_*\}\le\lambda_{\max}\begin{pmatrix}\Re(z_1\overline{z_2}u_1u_2) & \Re(m_1m_2)\\ \Re(m_1m_2) & \Re(\overline{z_1}z_2u_1u_2)\end{pmatrix} \tag{7.107a}$$

$$=(\Re u_1u_2)(\Re z_1\overline{z_2})+\sqrt{\big(|\Im u_1u_2||\Im z_1\overline{z_2}|+|\Re m_1m_2|\big)^2-2|\Im u_1u_2||\Im z_1\overline{z_2}||\Re m_1m_2|} \tag{7.107b}$$

$$\le(\Re u_1u_2)(\Re z_1\overline{z_2})+|\Im u_1u_2||\Im z_1\overline{z_2}|+|\Re m_1m_2| \tag{7.107c}$$

$$\le\Big|(\Re u_1u_2)(\Re z_1\overline{z_2})+|\Im u_1u_2||\Im z_1\overline{z_2}|\Big|+|\Re m_1m_2| \tag{7.107d}$$

$$=\sqrt{|z_1z_2u_1u_2|^2-\big(\Re u_1u_2|\Im z_1\overline{z_2}|-\Re z_1\overline{z_2}|\Im u_1u_2|\big)^2}+\sqrt{|m_1m_2|^2-[\Im m_1m_2]^2} \tag{7.107e}$$

$$\le|z_1z_2u_1u_2|+|m_1m_2| \tag{7.107f}$$

$$=\sqrt{(|u_1z_1|^2+|m_1|^2)(|u_2z_2|^2+|m_2|^2)-(|u_1z_1m_2|-|u_2z_2m_1|)^2} \tag{7.107g}$$

$$\le\sqrt{(|m_1|^2+|z_1u_1|^2)(|m_2|^2+|z_2u_2|^2)} \tag{7.107h}$$

$$\le 1, \tag{7.107i}$$

where in the last step we used (7.76).

We now assume that for some $0\le\epsilon\ll1$ we have

$$\max\{\Re\tau,\Re\tau_*\}\ge 1-\epsilon^2, \tag{7.108}$$

i.e. that all inequalities in (7.107a)–(7.107i) are in fact equalities up to an ϵ^2 error. The assertion (7.104) is then equivalent to

$$|z_1 - z_2| + \min\{|w_1 + \overline{w_2}|, |w_1 - \overline{w_2}|\} + \sqrt{|\Im w_1|} + \sqrt{|\Im w_2|} \lesssim \epsilon, \tag{7.109}$$

the proof of which we present now.

The fact that (7.107h)–(7.107i) is ϵ^2-saturated implies the saturation

$$|m_i|^2 + |z_i u_i|^2 = 1 + \mathcal{O}(\epsilon^2), \tag{7.110}$$

and, consequently,

$$|u_i| \sim 1. \tag{7.111}$$

Indeed, suppose that $|u_i| \ll 1$, then on the one hand since $u_i = u_i^2|z_i|^2 - m_i^2$, it follows that $|m_i| \ll 1$, while on the other hand $|1 - |m_i|^2| \ll 1$ from (7.110) which would be a contradiction. From (7.75) it follows that

$$|m_i|^2 + |u_i|^2|z_i|^2 \leq 1 - c\,\Im w_i,$$

from which we conclude $|\Im w_1| + |\Im w_2| \lesssim \epsilon^2$, i.e. the bound on the last two terms in (7.109). The ϵ^2-saturation of (7.107g)–(7.107h) implies that

$$\begin{aligned}\mathcal{O}(\epsilon) = |u_1 z_1 m_2| - |u_2 z_2 m_1| &= \sqrt{1 - |m_1|^2}|m_2| - \sqrt{1 - |m_2|^2}|m_1| + \mathcal{O}(\epsilon^2)\\ &= \sqrt{1 - |u_1 z_1|^2}|u_2 z_2| - \sqrt{1 - |u_2 z_2|^2}|u_1 z_1| + \mathcal{O}(\epsilon^2).\end{aligned}$$

Thus it follows that

$$|m_1| = |m_2| + \mathcal{O}(\epsilon), \qquad |z_1 u_1| = |z_2 u_2| + \mathcal{O}(\epsilon). \tag{7.112}$$

In the remainder of the proof we distinguish the cases

1. $\epsilon \ll |z_1|$ and $|m_1| \sim 1$,
2. $|z_1| \lesssim \epsilon$,
3. $|m_1| \lesssim \sqrt{\epsilon}$ and $|z_1| \sim 1$,
4. $\sqrt{\epsilon} \ll |m_1| \ll 1$ and $|z_1| \sim 1$,

where we note that this list is exhaustive since $|z_1| \ll 1$ implies $|m_1| \sim 1$ from (7.110).

In case 1 we have $|z_2| \sim |z_1|$ and $|m_1| \sim |m_2| \sim 1$ from (7.111)–(7.112). By the near-saturation of (7.107e)–(7.107f) it follows that $\Im m_1 m_2 = \mathcal{O}(\epsilon)$ and therefore with (7.112) that

$$m_1 = \pm\overline{m_2} + \mathcal{O}(\epsilon), \tag{7.113}$$

hence $|\Re m_1 m_2| \sim 1$. From the ϵ^2-saturation of (7.107b)–(7.107c) and (7.107e)–(7.107f) it then follows that

$$|\Im u_1 u_2| \left|\Im \frac{z_1 \overline{z_2}}{|z_1 z_2|}\right| = \mathcal{O}\left(\frac{\epsilon^2}{|z_1|^2}\right), \quad (\Re u_1 u_2)\left|\Im \frac{z_1 \overline{z_2}}{|z_1 z_2|}\right| = \left(\Re \frac{z_1 \overline{z_2}}{|z_1 z_2|}\right)|\Im u_1 u_2| + \mathcal{O}\left(\frac{\epsilon}{|z_1|}\right), \tag{7.114}$$

and (7.114) implies

$$|\Im u_1 u_2| + \left|\Im \frac{z_1 \overline{z_2}}{|z_1 z_2|}\right| \lesssim \frac{\epsilon}{|z_1|}. \tag{7.115}$$

Indeed, the first equality in (7.114) implies that at least one of the two factors is at most of size $\epsilon/|z_1| \ll 1$ in which case the second equality implies that the other factor satisfies the same bound since $|u_1 u_2| \sim 1$. Thus there exists some $c \in \mathbf{R}$, $|c| \sim 1$ such that $z_2 = c z_1 + \mathcal{O}(\epsilon)$ and $u_2 = \pm |c|^{-1} \overline{u_1} + \mathcal{O}(\epsilon/|z_1|)$ since the two proportionality constants c and $\pm|c|^{-1}$ are related by (7.112). On the other hand, from the MDE (7.26) we have that

$$u_2 = u_2^2 |z_2|^2 - m_2^2 = \overline{u_1}^2 |z_1|^2 - \overline{m_1}^2 + \mathcal{O}(\epsilon) = \overline{u_1} + \mathcal{O}(\epsilon) \tag{7.116}$$

and thus $|c| = 1 + \mathcal{O}(\epsilon/|z_1|)$. Finally, since (7.107c)–(7.107d) is assumed to be saturated up to an ϵ^2-error, $\Re u_1 u_2$ and $\Re z_1 \overline{z_2}$ have the same sign which, together with (7.116), fixes $c > 0$, and we conclude $z_2 = z_1 + \mathcal{O}(\epsilon)$. Finally, with

$$w_2 = \frac{m_2}{u_2} - m_2 = \pm \Big(\frac{\overline{m_1}}{\overline{u_1}} - \overline{m_1} \Big) + \mathcal{O}(\epsilon) = \pm \overline{w_1} + \mathcal{O}(\epsilon) \tag{7.117}$$

the claim (7.109) follows.

In case 2 the conclusion $z_2 = z_1 + \mathcal{O}(\epsilon)$ follows trivially from (7.112) and (7.111). Next, just as in case 1, we conclude (7.113) and therefore from (7.26) that

$$u_2 = u_2^2 |z_2|^2 - m_2^2 = -\overline{m_1}^2 + \mathcal{O}(\epsilon) = \overline{u_1} + \mathcal{O}(\epsilon),$$

and thus (7.109) follows just as in (7.117).

Finally, we consider the case $|m_i| \ll 1$, i.e. 3 and 4. If $|m_i| \ll 1$, then from (7.110), $|1 - |z_i u_i|^2| \ll 1$, and therefore from (7.26), $|1 - |u_i|| \ll 1$ and consequently $|1 - u_i |z_i|^2| = |m_i^2 / u_i| \ll 1$ and $|1 - u_i| + |1 - |z_i|^2| \ll 1$. If $|m_1| \lesssim \sqrt{\epsilon}$, then it follows from (7.112) that also $|m_2| \lesssim \sqrt{\epsilon}$. From solving the equation (7.26) for u_i we find

$$u_i = \frac{1 + \sqrt{1 + 4|z_i|^2 m_i^2}}{2|z_i|^2} = \frac{1}{|z_i|^2} + \mathcal{O}(|m_i|^2), \tag{7.118}$$

where the sign choice is fixed due to $|1 - u_i| \ll 1$.

In case 3 from $|m_i| \lesssim \sqrt{\epsilon}$ it follows that $u_i = |z_i|^{-2} + \mathcal{O}(\epsilon)$, and thus with (7.107e)–(7.107f) and $\Re u_1 u_2 \sim 1$ we can conclude

$$|\Im z_1 \overline{z_2}| = \frac{\Re z_1 \overline{z_2}}{\Re u_1 u_2} |\Im u_1 u_2| + \mathcal{O}(\epsilon) = \mathcal{O}(\epsilon), \quad |\Im u_1 u_2| = \mathcal{O}(\epsilon). \tag{7.119}$$

Together with (7.112) and the saturation of (7.107c)–(7.107d), we obtain $z_1 = z_2 + \mathcal{O}(\epsilon)$ and $u_1 = \overline{u_2} + \mathcal{O}(\epsilon)$ by the same argument as after (7.115). Equation (7.26) implies that $m_2 = \pm \overline{m_1} + \mathcal{O}(\epsilon)$ and we are able to conclude (7.109) just as in (7.117).

In case 4 from (7.112) we have $|m_2| \sim |m_1|$. By saturation of (7.107e)–(7.107f) it follows that

$$\Im \frac{m_1 m_2}{|m_1 m_2|} = \mathcal{O}\Big(\frac{\epsilon}{|m_1|} \Big)$$

and therefore, together with (7.112) we conclude that (7.113) also holds in this case. Now we use the saturation of (7.107b)–(7.107c) to conclude

$$|\Im u_1 u_2| |\Im z_1 \overline{z_2}| |\Re m_1 m_2| \lesssim \epsilon^2 \Big(|\Re m_1 m_2| + |\Im u_1 u_2| |\Im z_1 \overline{z_2}| \Big).$$

Together with the fact that $|\Im u_1 u_2||\Im z_1\overline{z_2}| \lesssim |m_i|^2 \sim |\Re m_1 m_2|$ from (7.113), (7.118), this implies $|\Im u_1 u_2||\Im z_1\overline{z_2}| \lesssim \epsilon^2$. Finally, the ϵ^2-saturation of (7.107e)–(7.107f) shows that (7.114) (with $|z_1| \sim |z_2| \sim 1$) also holds in case 4 and we are able to conclude (7.109) just like in case 1. □

7.7 Asymptotic independence of resolvents: Proof of Proposition 7.3.4

For any fixed $z \in \mathbf{C}$ let H^z be defined in (7.25). Recall that we denote the eigenvalues of H^z by $\{\lambda^z_{\pm i}\}_{i\in[n]}$, with $\lambda^z_{-i} = -\lambda^z_i$, and by $\{\boldsymbol{w}^z_{\pm i}\}_{i\in[n]}$ their corresponding orthonormal eigenvectors. As a consequence of the symmetry of the spectrum of H^z with respect to zero, its eigenvectors are of the form $\boldsymbol{w}^z_{\pm i} = (\boldsymbol{u}^z_i, \pm\boldsymbol{v}^z_i)$, for any $i \in [n]$. The eigenvectors of H^z are not well defined if H^z has multiple eigenvalues. This minor inconvenience can be easily solved by a tiny Gaussian regularization (see (7.136) and Remark 7.7.5 later).

Convention 7.7.1. *We omitted the index $i = 0$ in the definition of the eigenvalues of H^z. In the remainder of this section we always assume that all the indices are not zero, e.g we use the notation*

$$\sum_{i=-n}^{n} := \sum_{i=-n}^{-1} + \sum_{i=1}^{n},$$

and we use $|i| \leq A$, for some $A > 0$, to denote $0 < |i| \leq A$, etc.

The main result of this section is the proof of Proposition 7.3.4 which follows by Proposition 7.7.2 and the local law in Theorem 7.3.1.

Proposition 7.7.2 (Asymptotic independence of small eigenvalues of H^{z_l})**.** *Fix $p \in \mathbf{N}$, and let $\{\lambda^{z_l}_{\pm i}\}_{i=1}^n$ be the eigenvalues of H^{z_l}, with $l \in [p]$. For any $\omega_d, \omega_h, \omega_f > 0$ sufficiently small constants such that $\omega_h \ll \omega_f \ll \omega_d \ll 1$, there exist constants $\omega, \widehat{\omega}, \delta_0, \delta_1 > 0$, with $\omega_h \ll \delta_m \ll \widehat{\omega} \ll \omega \ll \omega_f$, for $m = 0, 1$, such that for any fixed $z_1, \dots, z_p \in \mathbf{C}$ so that $|z_l| \leq 1 - n^{-\omega_h}$, $|z_l - z_m|, |z_l - \overline{z}_m|, |z_l - \overline{z}_l| \geq n^{-\omega_d}$, with $l, m \in [p]$, $l \neq m$, it follows that*

$$\begin{aligned}\mathbf{E}\prod_{l=1}^p \frac{1}{n}\sum_{|i_l|\leq n^{\widehat{\omega}}} \frac{\eta_l}{(\lambda^{z_l}_{i_l})^2+\eta_l^2} &= \prod_{l=1}^p \mathbf{E}\,\frac{1}{n}\sum_{|i_l|\leq n^{\widehat{\omega}}} \frac{\eta_l}{(\lambda^{z_l}_{i_l})^2+\eta_l^2} \\ &\quad + \mathcal{O}\left(\frac{n^{\widehat{\omega}}}{n^{1+\omega}}\sum_{l=1}^p \frac{1}{\eta_l} \times \prod_{m=1}^p \left(1+\frac{n^\xi}{n\eta_m}\right) + \frac{n^{p\xi+2\delta_0}n^{\omega_f}}{n^{3/2}}\sum_{l=1}^p \frac{1}{\eta_l} + \frac{n^{p\delta_0+\delta_1}}{n^{\widehat{\omega}}}\right),\end{aligned} \tag{7.120}$$

for any $\xi > 0$, where $\eta_1, \dots, \eta_p \in [n^{-1-\delta_0}, n^{-1+\delta_1}]$ and the implicit constant in $\mathcal{O}(\cdot)$ may depend on p.

Proof of Proposition 7.3.4. Let ρ^{z_l} be the self consistent density of states of H^{z_l}, and define its quantiles $\gamma^{z_l}_i$ by

$$\frac{i}{n} = \int_0^{\gamma^{z_l}_i} \rho^{z_l}(x)\,\mathrm{d}x, \qquad i \in [n],$$

and $\gamma_{-i}^{z_l} = -\gamma_i^{z_l}$ for $i \in [n]$. Then, using the local law in Theorem 7.3.1, by standard application of Helffer-Sjöstrand formula (see e.g. [81, Lemma 7.1, Theorem 7.6] or [93, Section 5] for a detailed derivation), we conclude the following rigidity bound

$$|\lambda_i^{z_l} - \gamma_i^{z_l}| \leq \frac{n^{100\omega_h}}{n}, \qquad |i| \leq n^{1-10\omega_h}, \tag{7.121}$$

with very high probability, uniformly in $|z_l| \leq 1 - n^{-\omega_h}$. Then Proposition 7.3.4 follows by Proposition 7.7.2 and (7.121) exactly as in [58, Section 7.1]. We remark that in the current case we additionally require that $|z_l - \overline{z}_m|, |z_l - \overline{z}_l| \gtrsim n^{-\omega_d}$ compared to [58, Proposition 7.2], but this does not cause any change in the proof in [58, Section 7.1]. □

Section 7.7 is divided as follows: in Section 7.7.1 we state the main technical results needed to prove Proposition 7.7.2 and conclude its proof. In Section 7.7.2 we prove Theorem 7.2.7, which will follow by the results stated in Section 7.7.1. In Section 7.7.3 we estimate the overlaps of eigenvectors, corresponding to small indices, of H^{z_l}, H^{z_m} for $l \neq m$; this is the main input to prove the asymptotic independence in Proposition 7.7.2. In Section 7.7.4 and Section 7.7.6 we prove several technical results stated in Section 7.7.1. In Section 7.7.5 we present Proposition 7.7.16 which is a modification of the path-wise coupling of DBMs close to zero from [58, Proposition 7.14] to the case when the driving martingales in the DBM have a small correlation. This is needed to deal with the (small) correlation of $\boldsymbol{\lambda}^{z_l}$, the eigenvalues of H^{z_l}, for different l's.

7.7.1 Overview of the proof of Proposition 7.7.2

The main result of this section is the proof Proposition 7.7.2, which is essentially about the asymptotic independence of the eigenvalues $\lambda_i^{z_l}$, $\lambda_j^{z_m}$, for $l \neq m$ and small indices i and j. We do not prove this feature directly, instead we will compare $\lambda_i^{z_l}$, $\lambda_j^{z_m}$ with similar eigenvalues $\mu_i^{(l)}$, $\mu_j^{(m)}$ coming from independent Ginibre matrices, for which independence is straightforward by construction. The comparison is done by exploiting the strong local equilibration of the Dyson Brownian motion (DBM) in several steps. For convenience, we record the sequence of approximations in Figure 7.1. We remark that $z_1, \dots, z_p$ are fixed as in Proposition 7.7.2 throughout this section.

First, via a standard Green's function comparison argument (GFT) in Lemma 7.7.3 we prove that we may replace X by an i.i.d. matrix with a small Gaussian component. In the next step we make use of this Gaussian component and interpret the eigenvalues $\boldsymbol{\lambda}^z$ of H^z as the short-time evolution $\boldsymbol{\lambda}^z(t)$ of the eigenvalues of an auxiliary matrix H_t^z according to the Dyson Brownian motion. Proposition 7.7.2 is thus reduced to proving asymptotic independence of the flows $\boldsymbol{\lambda}^{z_l}(t)$ for different $l \in [p]$ after a short time $t = t_f$, a bit bigger than n^{-1}. The corresponding DBM describing the eigenvalues of H_t^z (see (7.133) later) differs from the standard DBM in two related aspects: (i) the driving martingales are weakly correlated, (ii) the interaction term has a coefficient slightly deviating from one. Note that the stochastic driving terms b_i in (7.133) are martingales but not Brownian motions (see Appendix 7.B for more details). Both effects come from the small but non-trivial overlap of the eigenvectors $\boldsymbol{w}_i^{z_l}$ with $\overline{\boldsymbol{w}}_j^{z_l}$. They also influence the well-posedness of the DBM, so an extra care is necessary. We therefore define two comparison processes. First we regularise the DBM by (i) setting the coefficient of the interaction equal to one, (ii) slightly reducing the diffusion term, and (iii) cutting off the possible large values of the correlation. The resulting process, denoted by $\mathring{\boldsymbol{\lambda}}(t)$ (see (7.141) later), will be called the *regularised DBM*. Second, we artificially remove the correlation in the driving martingales for large indices. This *partially correlated DBM*, defined in (7.146) below, will be denoted by $\widetilde{\boldsymbol{\lambda}}(t)$. We will show that in both steps the error is much smaller than the relevant scale $1/n$. After these preparations, we can directly compare the *partially correlated DBM* $\widetilde{\boldsymbol{\lambda}}(t)$ with its Ginibre counterpart $\widetilde{\boldsymbol{\mu}}(t)$ (see (7.148) later) since their distribution is the same. Finally, we remove the partial correlation in the process $\widetilde{\boldsymbol{\mu}}(t)$ by comparing it with a purely independent Ginibre DBM $\boldsymbol{\mu}(t)$, defined in (7.143) below.

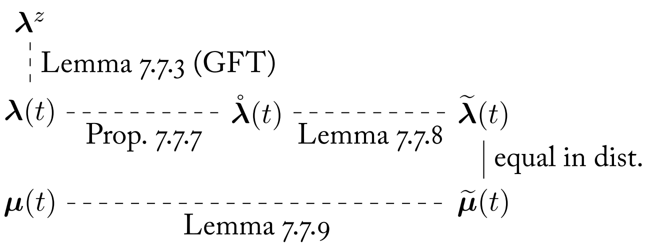


Figure 7.1: Proof overview for Proposition 7.7.2: The collections of eigenvalues $\boldsymbol{\lambda}^{z_l}$ of H^{z_l} for different l's are approximated by several stochastic processes. The processes $\boldsymbol{\mu} = \boldsymbol{\mu}^{(l)}$ are independent for different l's by definition.

Now we define these processes precisely. From now on we assume that $p = 2$ in Proposition 7.7.2 to make our presentation clearer. The case $p \geq 3$ is completely analogous. Consider the Ornstein-Uhlenbeck (OU) flow

$$d\widehat{X}_t = -\frac{1}{2}\widehat{X}_t \,\mathrm{d}t + \frac{\mathrm{d}\widehat{B}_t}{\sqrt{n}}, \qquad \widehat{X}_0 = X, \tag{7.122}$$

for a time

$$t_f := \frac{n^{\omega_f}}{n}, \tag{7.123}$$

with some small exponent $\omega_f > 0$ given as in Proposition 7.7.2, in order to add a small Gaussian component to X. Here $\widehat{B}_t$ in (7.122) is a standard matrix valued real Brownian motion, i.e. $\widehat{B}_{ab}$, $a, b \in [n]$ are i.i.d. standard real Brownian motions, independent of $\widehat{X}_0$. Then we can construct an i.i.d. matrix $\check{X}_{t_f}$ such that

$$\widehat{X}_{t_f} \stackrel{\mathrm{d}}{=} \check{X}_{t_f} + \sqrt{ct_f}U, \tag{7.124}$$

for some explicit constant $c > 0$ very close to 1, and U is a real Ginibre matrix independent of $\check{X}_{t_f}$. Using a simple Green's function comparison argument (GFT), by [58, Lemma 7.5], we conclude the following lemma.

Lemma 7.7.3. *The eigenvalues of H^{z_l} and the eigenvalues of $\widehat{H}^{z_l}_{t_f}$, with $t_f = n^{-1+\omega_f}$ obtained from replacing X by $\widehat{X}_{t_f}$, are close in the sense that for any sufficiently small $\omega_f, \delta_0, \delta_1 > 0$ it holds*

$$\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\leq n}\frac{\eta_l}{(\lambda_{i_l}(H^{z_l}))^2+\eta_l^2} = \mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\leq n}\frac{\eta_l}{(\lambda_{i_l}(\widehat{H}^{z_l}_{t_f}))^2+\eta_l^2} + \mathcal{O}\left(\frac{n^{2\xi+2\delta_0}t_f}{n^{1/2}}\sum_{l=1}^{2}\frac{1}{\eta_l}\right), \tag{7.125}$$

where $\eta_l \in [n^{-1-\delta_0}, n^{-1+\delta_1}]$.

Next, we consider the matrix flow

$$\mathrm{d}X_t = \frac{\mathrm{d}B_t}{\sqrt{n}}, \qquad X_0 = \check{X}_{t_f}, \tag{7.126}$$

and denote by H_t^z the Hermitisation of $X_t - z$. Here B_t is a real standard matrix valued Brownian motion independent of X_0 and $\widehat{B}_t$. Note that by construction X_{ct_f} is such that

$$X_{ct_f} \overset{\mathrm{d}}{=} \widehat{X}_{t_f}. \tag{7.127}$$

Denote the eigenvalues and eigenvectors of H_t^z by

$$\boldsymbol{\lambda}^z(t) = \{\lambda_{\pm i}^z(t) \mid i \in [n]\}, \quad \{\boldsymbol{w}_{\pm i}^z(t) \mid i \in [n]\} = \{(\boldsymbol{u}_i^z(t), \pm \boldsymbol{v}_i^z(t)) \mid i \in [n]\},$$

and the resolvent by $G_t^z(w) := (H_t^z - w)^{-1}$ for $w \in \mathbf{H}$. For any $\boldsymbol{w} = (\boldsymbol{u}, \boldsymbol{v})$, with $\boldsymbol{u}, \boldsymbol{v} \in \mathbf{C}^n$ define the projections $P_1, P_2 \colon \mathbf{C}^{2n} \to \mathbf{C}^n$ by

$$P_1 \boldsymbol{w} = \boldsymbol{u}, \qquad P_2 \boldsymbol{w} = \boldsymbol{v}, \tag{7.128}$$

and, for any $z, z' \in \mathbf{C}$, define the *eigenvector overlaps* by

$$\Theta_{ij}^{z,z'} = \Theta_{ij}^{z,z'}(t) := 4\Re[\langle P_1 \boldsymbol{w}_j^{z'}(t), P_1 \boldsymbol{w}_i^z(t)\rangle \langle P_2 \boldsymbol{w}_i^z(t), P_2 \boldsymbol{w}_j^{z'}(t)\rangle], \quad |i|, |j| \le n. \tag{7.129}$$

Note that by the spectral symmetry of H_t^z it holds

$$\Theta_{ij}^{z,z} = \delta_{i,j} - \delta_{i,-j}, \quad \Theta_{ij}^{z,z'} = \Theta_{ji}^{z',z}, \quad |\Theta_{ij}^{z,z'}| \le 1, \tag{7.130}$$

for any $|i|, |j| \le n$. The coefficients $\Theta_{ij}^{z,z'}(t)$ are small with high probability due to the following lemma whose proof is postponed to Section 7.7.3.

Lemma 7.7.4 (Eigenvectors overlaps are small)**.** *For any sufficiently small constants $\omega_h, \omega_d > 0$, there exists $\omega_E > 0$ so that for any $z, z' \in \mathbf{C}$ such that $|z|, |z'| \le 1 - n^{-\omega_h}$, $|z - z'| \ge n^{-\omega_d}$, we have*

$$\sup_{0 \le t \le T} \sup_{|i|,|j| \le n} \left| \Theta_{ij}^{z,z'}(t) \right| \le n^{-\omega_E}, \tag{7.131}$$

with very high probability for any fixed $T \ge 0$.

Most of the DBM analysis is performed for a fixed $z \in \{z_1, z_2\}$, with z_1, z_2 as in Proposition 7.7.2, for this purpose we introduce the notation

$$\Lambda_{ij}^z(t) := \Theta_{ij}^{z,\overline{z}}(t), \tag{7.132}$$

for any $|i|, |j| \le n$. In particular, note $\Theta_{ij}^{z,\overline{z}} = \Theta_{ij}^{\overline{z},z}$ and so that by (7.130) it follows that $\Lambda_{ij}^z(t) = \Lambda_{ji}^z(t)$.

By the derivation of the DBM in Appendix 7.B, using the fact that $\overline{\boldsymbol{w}}^z = \boldsymbol{w}^{\overline{z}}$, for $z = z_l$ with $l \in [2]$, it follows that (7.126) induces the flow

$$\mathrm{d}\lambda_i^z(t) = \frac{\mathrm{d}b_i^z}{\sqrt{n}} + \frac{1}{2n} \sum_{j \ne i} \frac{1 + \Lambda_{ij}^z(t)}{\lambda_i^z(t) - \lambda_j^z(t)} \, \mathrm{d}t, \quad \lambda_i^z(0) = \lambda_i^z, \qquad |i| \le n, \tag{7.133}$$

on the eigenvalues $\{\lambda_i^z(t)\}_{|i|\le n}$ of H_t^z. Here $\{\lambda_i^z\}_{|i|\le n}$ are the eigenvalues of the initial matrix H^z. The martingales $\{b_i^z\}_{i\in[n]}$, with $b_i^z(0)=0$, and $\Lambda_{ij}^z(t)$, the overlap of eigenvectors in (7.132), (7.129), are defined on a probability space Ω_b equipped with the filtration

$$(\mathcal{F}_{b,t})_{0\le t\le T} := \big(\sigma(X_0,(B_s)_{0\le s\le t})\big)_{0\le t\le T}, \tag{7.134}$$

where B_s is defined in (7.126). The martingale differentials in (7.133) are such that (see (7.254), (7.255))

$$\begin{aligned} \mathrm{d}b_i^z &:= \mathrm{d}B_{ii}^z + \mathrm{d}\overline{B_{ii}^z}, \quad \text{with} \quad \mathrm{d}B_{ij}^z := \langle \boldsymbol{u}_i^z, (\mathrm{d}B)\boldsymbol{v}_i^z\rangle, \quad i,j\in[n], \\ \mathbf{E}\Big[\mathrm{d}b_i^z\,\mathrm{d}b_j^z \,\Big|\, \mathcal{F}_{b,t}\Big] &= \frac{\delta_{ij}+\Lambda_{ij}^z(t)}{2}\,\mathrm{d}t, \quad i,j\in[n], \end{aligned} \tag{7.135}$$

and $\mathrm{d}b_{-i}^z = -\,\mathrm{d}b_i^z$ for $i\in[n]$. Here we used the notation Ω_b for the probability space to emphasize that is the space where the martingales $\boldsymbol{b}^z$ are defined, since in Section 7.7.1.2 we will introduce another probability space which we will denote by Ω_β.

In the remainder of this section we will apply Lemma 7.7.4 for $z=z_1, z'=z_2$ and $z=z_1, z'=\overline{z}_2$ and $z=z_l, z'=\overline{z}_l$, for $l\in[2]$, with z_1,z_2 fixed as in Proposition 7.7.2. We recall that throughout this section we assumed that $p=2$ in Proposition 7.7.2. Note that $\Lambda_{ij}^{z_1}$, $\Lambda_{ij}^{z_2}$, $\Theta_{ij}^{z_1,z_2}$, $\Theta_{ij}^{z_1,\overline{z}_2}$ with $|i|,|j|\le n$, are not well-defined if $H_t^{z_1}$, $H_t^{z_2}$ have multiple eigenvalues. This minor inconvenience can easily be resolved by a tiny regularization as in [55, Lemma 6.2] (which is the singular values counterpart of [53, Proposition 2.3]). Using this result, we may, without loss of generality, assume that the eigenvalues of $H_t^{z_l}$ are almost surely distinct for any fixed time $t\ge 0$. Indeed, if this were not the case then we replace $H_0^{z_l}$ by

$$H_{0,\mathrm{reg}}^{z_l} := \begin{pmatrix} 0 & X-z_l+e^{-n}Q \\ X^*-\overline{z}_l+e^{-n}Q^* & 0\end{pmatrix}, \tag{7.136}$$

with Q being a complex $n\times n$ Ginibre matrix independent of X, i.e. we may regularize X by adding an exponentially small Gaussian component. Then, by [55, Lemma 6.2], $H_{t,\mathrm{reg}}^{z_l}$, the evolution of $H_{0,\mathrm{reg}}^{z_l}$ along the flow (7.126), does not have multiple eigenvalues almost surely; additionally, the eigenvalues of $H_{0,\mathrm{reg}}^{z_l}$ and the ones of $H_0^{z_l}$ are exponentially close. Hence, by Fubini's theorem, $\{\Lambda_{ij}^{z_l}(t)\}_{|i|,|j|\le n}$, with $l\in[2]$, and $\{\Theta_{ij}^{z_1,z_2}(t)\}_{|i|,|j|\le n}$, $\{\Theta_{ij}^{z_1,\overline{z}_2}(t)\}_{|i|,|j|\le n}$ are well-defined for almost all $t\ge 0$; we set them equal to zero whenever they are not well defined.

Remark 7.7.5. *The perturbation of X in* (7.136) *is exponentially small, hence does not change anything in the proof of the local laws in Theorem 7.3.1 and Theorem 7.3.5 or in the Green's function comparison (GFT) argument in Lemma 7.7.3, since these proofs deal with scales much bigger than e^{-n}. This implies that any local law or GFT result which holds for $H_t^{z_l}$ then holds true for $H_{t,\mathrm{reg}}^{z_l}$ as well. Hence, in the remainder of this section we assume that [55, Lemma 6.2] holds true for $H_t^{z_l}$ (the unperturbed matrix).*

The process (7.133) is well-defined in the sense of Proposition 7.7.6, whose proof is postponed to Section 7.7.6.

Proposition 7.7.6 (The DBM in (7.133) is well-posed)**.** *Fix $z\in\{z_1,z_2\}$, and let H_t^z be defined by the flow* (7.126)*. Then the eigenvalues $\boldsymbol{\lambda}(t)$ of H_t^z are the unique strong solution to* (7.133) *on*

$[0,T]$, *for any* $T>0$, *such that* $\boldsymbol{\lambda}(t)$ *is adapted to the filtration* $(\mathcal{F}_{b,t})_{0\le t\le T}$, $\boldsymbol{\lambda}(t)$ *is* γ*-Hölder continuous for any* $\gamma\in(0,1/2)$, *and*

$$\mathbf{P}\Big(\lambda_{-n}(t)<\cdots<\lambda_{-1}(t)<0<\lambda_1(t)<\cdots<\lambda_n(t),\ \textit{for almost all}\ t\in[0,T]\Big)=1.$$

In order to prove that the term Λ_{ij}^z in (7.133) is irrelevant, we will couple the driving martingales in (7.133) with the ones of a DBM that does not have the additional term Λ_{ij}^z (see (7.141) below). For this purpose we have to consider the correlation of $\{b_i^{z_1}\}_{|i|\le n}$, $\{b_i^{z_2}\}_{|i|\le n}$ for two different $z_1,z_2\in\mathbf{C}$ as in Proposition 7.7.2. In the following we will focus only on the driving martingales with positive indices, since the ones with negative indices are defined by symmetry. The martingales $\boldsymbol{b}^{z_l}=\{b_i^{z_l}\}_{i\in[n]}$, with $l=1,2$, are defined on a common probability space equipped with the filtration $(\mathcal{F}_{b,t})_{0\le t\le T}$ from (7.134).

We consider $\boldsymbol{b}^{z_1}$, $\boldsymbol{b}^{z_2}$ jointly as a $2n$-dimensional martingale $(\boldsymbol{b}^{z_1},\boldsymbol{b}^{z_2})$. Define the naturally reordered indices

$$\mathfrak{i}=(l-1)n+i,\qquad \mathfrak{j}=(m-1)n+j,$$

with $l,m\in[2]$, $i,j\in[n]$, and $\mathfrak{i},\mathfrak{j}\in[2n]$. Then the correlation between $\boldsymbol{b}^{z_1}$, $\boldsymbol{b}^{z_2}$ is given by

$$C_{\mathfrak{i}\mathfrak{j}}(t)\,\mathrm{d}t:=\mathbf{E}\Big[\mathrm{d}b_i^{z_l}\,\mathrm{d}b_j^{z_m}\,\Big|\,\mathcal{F}_{b,t}\Big]=\frac{\Theta_{ij}^{z_l,z_m}(t)+\Theta_{ij}^{z_l,\overline{z}_m}(t)}{2}\,\mathrm{d}t\qquad \mathfrak{i},\mathfrak{j}\in[2n]. \tag{7.137}$$

Note that $C(t)$ is a positive semi-definite matrix. In particular, taking also negative indices into account, for a fixed $z\in\{z_1,z_2\}$, the family of martingales $\boldsymbol{b}^z=\{b_i^z\}_{|i|\le n}$ is such that

$$\mathbf{E}\Big[\mathrm{d}b_i^z\,\mathrm{d}b_j^z\,\Big|\,\mathcal{F}_{b,t}\Big]=\frac{\delta_{i,j}-\delta_{i,-j}+\Lambda_{ij}^z(t)}{2}\,\mathrm{d}t,\qquad |i|,|j|\le n. \tag{7.138}$$

7.7.1.1 Comparison of $\boldsymbol{\lambda}$ with the regularised process $\mathring{\boldsymbol{\lambda}}$

By Lemma 7.7.4 the overlaps $\Theta_{ij}^{z,z'}$ are typically small for any $z,z'\in\mathbf{C}$ such that $|z|,|z'|\le 1-n^{-\omega_h}$ and $|z-z'|\ge n^{-\omega_d}$. We now define their cut-off versions (see (7.140) below). We only consider positive indices, since negative indices are defined by symmetry. Throughout this section we use the convention that regularised objects will be denoted by circles. Let z_l, with $l\in[2]$ be fixed throughout Section 7.7 as in Proposition 7.7.2. Define the $2n\times 2n$ matrix $\mathring{C}(t)$ by

$$\mathring{C}_{\mathfrak{i}\mathfrak{j}}(t):=\frac{\mathring{\Theta}_{ij}^{z_l,z_m}(t)+\mathring{\Theta}_{ij}^{z_l,\overline{z}_m}(t)}{2}\qquad i,j\in[n],\quad \mathfrak{i},\mathfrak{j}\in[2n], \tag{7.139}$$

where $\mathring{\Theta}_{ij}^{z_l,z_l}=\delta_{ij}$ for $i,j\in[n]$, and

$$\begin{aligned}\mathring{\Theta}_{ij}^{z_1,z_2}(t):&=\Theta_{ij}^{z_1,z_2}(t)\cdot\mathbf{1}\left(\mathcal{A}(t)\le n^{-\omega_E}\right),\quad \mathring{\Theta}_{ij}^{z_l,\overline{z}_m}(t):=\Theta_{ij}^{z_l,\overline{z}_m}(t)\cdot\mathbf{1}\left(\mathcal{A}(t)\le n^{-\omega_E}\right),\\ \mathcal{A}(t)=\mathcal{A}^{z_1,z_2}(t)&:=\max_{|i|,|j|\le n}|\Lambda_{ij}^{z_1}(t)|+|\Lambda_{ij}^{z_2}(t)|+|\Theta_{ij}^{z_1,\overline{z}_2}(t)|+|\Theta_{ij}^{z_1,z_2}(t)|\end{aligned} \tag{7.140}$$

for any $l,m\in[2]$, recalling that $\Lambda_{ij}^{z_l}=\Theta_{ij}^{z_l,\overline{z}_l}$. Note that by Lemma 7.7.4 it follows that $\mathring{C}(t)=C(t)$ on a set of very high probability, and $\mathring{C}(t)=\frac{1}{2}I$, with I the $2n\times 2n$ identity matrix, on the complement of this set, for any $t \in [0,T]$. In particular, $\mathring{C}(t)$ is positive semi-definite for any $t \in [0,T]$, since $C(t)$, defined as a covariance in (7.137), is positive semi-definite. The purpose of the cut-off in (7.139) it is to ensure the well-posedness of the process (7.141) below.

We compare the processes $\boldsymbol{\lambda}^{z_l}(t)$ in (7.133) with the *regularised processes* $\mathring{\boldsymbol{\lambda}}^{z_l}(t)$ defined, for $z = z_l$, by

$$\mathrm{d}\mathring{\lambda}_i^z = \frac{\mathrm{d}\mathring{b}_i^z}{\sqrt{n(1+n^{-\omega_r})}} + \frac{1}{2n}\sum_{j\neq i}\frac{1}{\mathring{\lambda}_i^z - \mathring{\lambda}_j^z}\,\mathrm{d}t, \qquad \mathring{\lambda}_i^z(0) = \lambda_i^z(0), \qquad |i| \le n, \quad (7.141)$$

with $\omega_r > 0$ such that $\omega_f \ll \omega_r \ll \omega_E$. We organise the martingales $\boldsymbol{b}^{z_1}, \boldsymbol{b}^{z_2}$ with positive indices into a single $2n$-dimensional vector $\boldsymbol{b} = (\boldsymbol{b}^{z_1}, \boldsymbol{b}^{z_2})$ with a correlation structure given by (7.137). Then by Doob's martingale representation theorem [120, Theorem 18.12] there exists a standard Brownian motion $\mathfrak{w} = (\mathfrak{w}^{(1)}, \mathfrak{w}^{(2)}) \in \mathbf{R}^{2n}$ realized on an extension $(\widetilde{\Omega}_b, \widetilde{\mathcal{F}}_{b,t})$ of the original probability space $(\Omega_b, \mathcal{F}_{b,t})$ such that $\mathrm{d}\boldsymbol{b} = \sqrt{C}\,\mathrm{d}\mathfrak{w}$, with $\sqrt{C} = \sqrt{C(t)}$ the matrix square root of $C(t)$. Moreover, $\mathfrak{w}(t)$ and $C(t)$ are adapted to the filtration $\widetilde{\mathcal{F}}_{b,t}$. Then the martingales $\mathring{\boldsymbol{b}}^{z_l} = \{\mathring{b}_i^{z_l}\}_{i\in[n]}$, with $l \in [2]$, are defined by $\mathring{\boldsymbol{b}}^{z_l}(0) = 0$ and

$$\begin{pmatrix} \mathrm{d}\mathring{\boldsymbol{b}}^{z_1}(t) \\ \mathrm{d}\mathring{\boldsymbol{b}}^{z_2}(t) \end{pmatrix} := \sqrt{\mathring{C}(t)} \begin{pmatrix} \mathrm{d}\mathfrak{w}^{(1)}(t) \\ \mathrm{d}\mathfrak{w}^{(2)}(t) \end{pmatrix}, \quad (7.142)$$

where $\sqrt{\mathring{C}(t)}$ denotes the matrix square root of the positive semi-definite matrix $\mathring{C}(t)$. For negative indices we define $\mathring{b}_{-i} = -\mathring{b}_i$, with $i \in [n]$. The purpose of the additional factor $1 + n^{-\omega_r}$ in (7.141) is to ensure the well-posedness of the process, since $\mathring{\boldsymbol{b}}^z$ is a small deformation of a family of i.i.d. Brownian motions with variance $1/2$, and the well-posedness of (7.141) is already critical for those Brownian motions (it corresponds to the GOE case, i.e. $\beta = 1$). The well-posedness of the process (7.141) is proven in Appendix 7.A. The main result of this section is the following proposition, whose proof is deferred to Section 7.7.4.

Proposition 7.7.7 (The *regularised process* $\mathring{\boldsymbol{\lambda}}$ is close to $\boldsymbol{\lambda}$). *For any sufficiently small ω_d, ω_h, $\omega_f > 0$ such that $\omega_h \ll \omega_f \ll 1$ there exist small constants $\widehat{\omega}, \omega > 0$ such that $\omega_h \ll \widehat{\omega} \ll \omega \ll \omega_f$, and that for $|z_l - \overline{z}_l|, |z_l - \overline{z}_m|, |z_l - z_m| \ge n^{-\omega_d}$, $|z_l| \le 1 - n^{-\omega_h}$, with $l \neq m$, it holds*

$$|\lambda_i^{z_l}(ct_f) - \mathring{\lambda}_i^{z_l}(ct_f)| \le n^{-1-\omega}, \qquad |i| \le n^{\widehat{\omega}},$$

with very high probability, where $t_f = n^{-1+\omega_f}$ and $c > 0$ is defined in (7.124).

7.7.1.2 Definition of the partially correlated processes $\widetilde{\boldsymbol{\lambda}}$, $\widetilde{\boldsymbol{\mu}}$

The construction of the *partially correlated processes* for $\mathring{\boldsymbol{\lambda}}^{z_l}(t)$ is exactly the same as in the complex case [58, Section 7.2]; we present it here as well for completeness. We want to compare the correlated processes $\mathring{\boldsymbol{\lambda}}^{z_l}(t)$, with $l = 1, 2$, defined on a probability space $\widetilde{\Omega}_b$ equipped with a filtration $\widetilde{\mathcal{F}}_{b,t}$ with carefully constructed independent processes $\boldsymbol{\mu}^{(l)}(t)$, $l = 1, 2$ on a different probability space Ω_β equipped with a filtration $\mathcal{F}_{\beta,t}$, which is defined in (7.144) below. We choose $\boldsymbol{\mu}^{(l)}(t)$ to be a *complex Ginibre DBM*, i.e. it is given as the solution of

$$\mathrm{d}\mu_i^{(l)}(t) = \frac{\mathrm{d}\beta_i^{(l)}}{\sqrt{2n}} + \frac{1}{2n}\sum_{j\neq i}\frac{1}{\mu_i^{(l)}(t) - \mu_j^{(l)}(t)}\,\mathrm{d}t, \quad \mu_i^{(l)}(0) = \mu_i^{(l)}, \qquad |i| \le n, \quad (7.143)$$

with $\mu_i^{(l)}$ the singular values, taken with positive and negative sign, of independent complex Ginibre matrices $X^{(l)}$, and $\boldsymbol{\beta}^{(l)} = \{\beta_i^{(l)}\}_{i\in[n]}$ being independent vectors of i.i.d. standard real Brownian motions, and $\beta_{-i}^{(l)} = -\beta_i^{(l)}$ for $i \in [n]$. The filtration $\mathcal{F}_{\beta,t}$ is defined by

$$(\mathcal{F}_{\beta,t})_{0\le t\le T} := (\sigma(X^{(l)}, (\boldsymbol{\beta}_s^{(l)})_{0\le s\le t}, (\widetilde{\boldsymbol{\zeta}}_s^{(l)})_{0\le s\le t}; l \in [2]))_{0\le t\le T}, \tag{7.144}$$

with $\widetilde{\boldsymbol{\zeta}}^{(l)}$ standard real i.i.d. Brownian motions, independent of $\boldsymbol{\beta}^{(l)}$, which will be used later in the definition of the processes in (7.148).

The comparison of $\mathring{\boldsymbol{\lambda}}^{z_l}(t)$ and $\boldsymbol{\mu}^{(l)}(t)$ is done via two intermediate *partially correlated processes* $\widetilde{\boldsymbol{\lambda}}^{(l)}(t)$, $\widetilde{\boldsymbol{\mu}}^{(l)}(t)$ so that for a time $t \ge 0$ large enough $\widetilde{\lambda}_i^{(l)}(t)$, $\widetilde{\mu}_i^{(l)}(t)$ for small indices i will be close to $\mathring{\lambda}_i^{z_l}(t)$ and $\mu_i^{(l)}(t)$, respectively, with very high probability. Additionally, the processes $\widetilde{\boldsymbol{\lambda}}^{(l)}$, $\widetilde{\boldsymbol{\mu}}^{(l)}$ will be constructed such that they have the same joint distribution:

$$\left(\widetilde{\boldsymbol{\lambda}}^{(1)}(t), \widetilde{\boldsymbol{\lambda}}^{(2)}(t)\right)_{0\le t\le T} \overset{\mathrm{d}}{=} \left(\widetilde{\boldsymbol{\mu}}^{(1)}(t), \widetilde{\boldsymbol{\mu}}^{(2)}(t)\right)_{0\le t\le T}, \tag{7.145}$$

for any $T > 0$.

Fix $\omega_A > 0$ such that $\omega_h \ll \omega_A \ll \omega_f$, and for $l \in [2]$ define the process $\widetilde{\boldsymbol{\lambda}}^{(l)}(t)$ to be the solution of

$$\mathrm{d}\widetilde{\lambda}_i^{(l)}(t) = \frac{1}{2n}\sum_{j\ne i}\frac{1}{\widetilde{\lambda}_i^{(l)}(t) - \widetilde{\lambda}_j^{(l)}(t)}\,\mathrm{d}t + \begin{cases} (n(1+n^{-\omega_r}))^{-1/2}\,\mathrm{d}\mathring{b}_i^{z_l}, & |i| \le n^{\omega_A} \\ (2n)^{-1/2}\,\mathrm{d}\widetilde{b}_i^{(l)}, & n^{\omega_A} < |i| \le n, \end{cases} \tag{7.146}$$

with initial data $\widetilde{\boldsymbol{\lambda}}^{(l)}(0)$ being the singular values, taken with positive and negative sign, of independent complex Ginibre matrices $\widetilde{Y}^{(l)}$ independent of $\boldsymbol{\lambda}^{z_l}(0)$. Here $\mathrm{d}\mathring{b}_i^{z_l}$ is the martingale differential from (7.141) which is used for small indices in (7.146). For large indices we define the driving martingales to be an independent collection $\{\{\widetilde{b}_i^{(l)}\}_{i=n^{\omega_A}+1}^n \mid l \in [2]\}$ of two vector-valued i.i.d. standard real Brownian motions which are also independent of $\{\{\mathring{b}_{\pm i}^{z_l}\}_{i=1}^n \mid l \in [2]\}$, and that $\widetilde{b}_{-i}^{(l)} = -\widetilde{b}_i^{(l)}$ for $i \in [n]$. The martingales $\mathring{\boldsymbol{b}}^{z_l}$, with $l \in [2]$, and $\{\{\widetilde{b}_i^{(l)}\}_{i=n^{\omega_A}+1}^n \mid l \in [2]\}$ are defined on a common probability space that we continue to denote by $\widetilde{\Omega}_b$ with the common filtration $\widetilde{\mathcal{F}}_{b,t}$, given by

$$(\widetilde{\mathcal{F}}_{b,t})_{0\le t\le T} := (\sigma(X_0, \widetilde{Y}^{(l)}, (B_s)_{0\le s\le t}, (\widetilde{\boldsymbol{b}}^{(l)})_{0\le s\le t}; l \in [2]))_{0\le t\le T}.$$

The well-posedness of (7.146), and of (7.148) below, readily follows by exactly the same arguments as in Appendix 7.A.

Notice that $\mathring{\boldsymbol{\lambda}}(t)$ and $\widetilde{\boldsymbol{\lambda}}(t)$ differ in two aspects: the driving martingales with large indices for $\widetilde{\boldsymbol{\lambda}}(t)$ are set to be independent, and the initial conditions are different. Lemma 7.7.8 below states that these differences are negligible for our purposes (i.e. after time ct_1 the two processes at small indices are closer than the rigidity scale $1/n$). Its proof is postponed to Section 7.7.5.1. Let $\rho_{\mathrm{sc}}(E) = \frac{1}{2\pi}\sqrt{4 - E^2}$ denote the semicircle density.

Lemma 7.7.8 (The partially correlated process $\widetilde{\boldsymbol{\lambda}}$ is close to $\mathring{\boldsymbol{\lambda}}$). *Let* $\mathring{\boldsymbol{\lambda}}^{z_l}(t)$, $\widetilde{\boldsymbol{\lambda}}^{(l)}(t)$, *with* $l \in [2]$, *be the processes defined in* (7.141) *and* (7.146), *respectively. For any sufficiently small* $\omega_h, \omega_f > 0$ *such that* $\omega_h \ll \omega_f \ll 1$ *there exist constants* $\omega, \widehat{\omega} > 0$ *such that* $\omega_h \ll \widehat{\omega} \ll \omega \ll \omega_f$, *and that for* $|z_l| \le 1 - n^{-\omega_h}$ *it holds*

$$|\rho^{z_l}(0)\mathring{\lambda}_i^{z_l}(ct_f) - \rho_{\mathrm{sc}}(0)\widetilde{\lambda}_i^{(l)}(ct_f)| \le n^{-1-\omega}, \qquad |i| \le n^{\widehat{\omega}}, \tag{7.147}$$

with very high probability, where $t_f := n^{-1+\omega_f}$ *and* $c > 0$ *is defined in* (7.124).

Finally, $\widetilde{\boldsymbol{\mu}}^{(l)}(t)$, the comparison process of $\boldsymbol{\mu}^{(l)}(t)$, is given as the solution of the following DBM

$$\mathrm{d}\widetilde{\mu}_i^{(l)}(t) = \frac{1}{2n}\sum_{j\neq i}\frac{1}{\widetilde{\mu}_i^{(l)}(t)-\widetilde{\mu}_j^{(l)}(t)}\,\mathrm{d}t + \begin{cases} \big(n(1+n^{-\omega_r})\big)^{-1/2}\,\mathrm{d}\mathring{\zeta}_i^{z_l}, & |i|\le n^{\omega_A},\\ (2n)^{-1/2}\,\mathrm{d}\widetilde{\zeta}_i^{(l)}, & n^{\omega_A}<|i|\le n,\end{cases} \tag{7.148}$$

with initial data $\widetilde{\boldsymbol{\mu}}^{(l)}(0)=\boldsymbol{\mu}^{(l)}$. We now explain how to construct the driving martingales in (7.148) so that (7.145) is satisfied. For this purpose we closely follow [58, Eqs. (7.22)-(7.29)]. We only consider positive indices, since the negative indices are defined by symmetry. Define the $2n^{\omega_A}$-dimensional martingale $\underline{\mathring{\boldsymbol{b}}} := \{\{\mathring{b}_i^{z_l}\}_{i\in[n^{\omega_A}]}|l\in[2]\}$. Throughout this section underlined vectors or matrices denote their restriction to the first $i\in[n^{\omega_A}]$ indices within each l-group, i.e.

$$\boldsymbol{v}\in\mathbf{C}^{2n}\implies \underline{\boldsymbol{v}}\in\mathbf{C}^{2n^{\omega_A}},\quad \text{with}\quad \underline{v}_i := \begin{cases} v_i & \text{if}\quad i\in[n^{\omega_A}]\\ v_{i+n^{\omega_A}} & \text{if}\quad i\in n+[n^{\omega_A}].\end{cases}$$

Then we define $\underline{\mathring{C}}(t)$ as the $2n^{\omega_A}\times 2n^{\omega_A}$ positive semi-definite matrix which consists of the four blocks corresponding to index pairs $\{(i,j)\in[n^{\omega_A}]^2\}$ of the matrix $\mathring{C}(t)$ defined in (7.139). Similarly to (7.142), by Doob's martingale representation theorem, we obtain $\mathrm{d}\underline{\mathring{\boldsymbol{b}}} = (\underline{\mathring{C}})^{1/2}\,\mathrm{d}\boldsymbol{\theta}$ with $\boldsymbol{\theta}(t):=\{\{\theta_i^{(l)}(t)\}_{i\in[n^{\omega_A}]}\,|\,l\in[2]\}$ a family of i.i.d. standard real Brownian motions. We define an independent copy $\underline{\mathring{C}}^{\#}(s)$ of $\underline{\mathring{C}}(s)$ and $\underline{\boldsymbol{\beta}} := \{\{\beta_i^{(l)}\}_{i\in[n^{\omega_A}]}\,|\,l\in[2]\}$ such that $(\underline{\mathring{C}}^{\#}(t),\underline{\boldsymbol{\beta}}(t))$ has the same joint distribution as $(\underline{\mathring{C}}(t),\boldsymbol{\theta}(t))$. We then define the families $\underline{\mathring{\boldsymbol{\zeta}}} := \{\{\mathring{\zeta}_i^{z_l}\}_{i\in[n^{\omega_A}]}\,|\,l\in[2]\}$ by $\underline{\mathring{\boldsymbol{\zeta}}}(0)=0$ and

$$\mathrm{d}\underline{\mathring{\boldsymbol{\zeta}}}(t) := \left(\underline{\mathring{C}}^{\#}(t)\right)^{1/2}\mathrm{d}\underline{\boldsymbol{\beta}}(t), \tag{7.149}$$

and extend this to negative indices by $\zeta_{-i}^{z_l} = -\zeta_i^{z_l}$ for $i\in[n^{\omega_A}]$. For indices $n^{\omega_A}<|i|\le n$, instead, we choose $\{\widetilde{\zeta}_{\pm i}^{(l)}\}_{i=n^{\omega_A}+1}^n$ to be independent families (independent of each other for different l's, and also independent of $\boldsymbol{\beta}$) of i.i.d. Brownian motions defined on the same probability space Ω_β. Note that (7.145) follows by the construction in (7.149).

Similarly to Lemma 7.7.8 we also have that $\boldsymbol{\mu}(t)$ and $\widetilde{\boldsymbol{\mu}}(t)$ are close thanks to the carefully designed relation between their driving Brownian motions. The proof of this lemma is postponed to Section 7.7.5.1.

Lemma 7.7.9 (The partially correlated process $\widetilde{\boldsymbol{\mu}}$ is close to $\boldsymbol{\mu}$)**.** *For any sufficiently small $\omega_d,\omega_h,\omega_f>0$, there exist constants $\omega,\widehat{\omega}>0$ such that $\omega_h\ll\widehat{\omega}\ll\omega\ll\omega_f$, and that for $|z_l-z_m|,|z_l-\overline{z}_m|,|z_l-\overline{z}_l|\ge n^{-\omega_d}$, $|z_l|\le 1-n^{-\omega_h}$, with $l,m\in[2]$, $l\neq m$, it holds*

$$\left|\mu_i^{(l)}(ct_f)-\widetilde{\mu}_i^{(l)}(ct_f)\right|\le n^{-1-\omega},\quad |i|\le n^{\widehat{\omega}},\quad l\in[2], \tag{7.150}$$

with very high probability, where $t_f=n^{-1+\omega_f}$ and $c>0$ is defined in (7.124).

7.7.1.3 Proof of Proposition 7.7.2

In this section we conclude the proof of Proposition 7.7.2 using the comparison processes defined in Section 7.7.1.1 and Section 7.7.1.2. We recall that $p=2$ for simplicity. More precisely, we use that the processes $\boldsymbol{\lambda}^{z_l}(t)$, $\mathring{\boldsymbol{\lambda}}^{z_l}(t)$ and $\mathring{\boldsymbol{\lambda}}^{z_l}(t)$, $\widetilde{\boldsymbol{\lambda}}^{(l)}(t)$ and $\widetilde{\boldsymbol{\mu}}^{(l)}(t)$, $\boldsymbol{\mu}^{(l)}(t)$ are close path-wise at time t_1, as stated in Proposition 7.7.7, Lemma 7.7.8, and Lemma 7.7.9, respectively, choosing $\omega, \widehat{\omega}$ as the minimum of the ones in the statements of this three results. In particular, by these results and Lemma 7.7.3 we readily conclude the following lemma, whose proof is postponed to the end of this section.

Lemma 7.7.10. *Let $\boldsymbol{\lambda}^{z_l}$ be the eigenvalues of H^{z_l}, and let $\boldsymbol{\mu}^{(l)}(t)$ be the solution of (7.143). Let $\omega, \widehat{\omega}, \omega_h > 0$ given as above, and define $\nu_{z_l} := \rho_{sc}(0)/\rho^{z_l}(0)$, then for any small $\omega_f > 0$ such that $\omega_h \ll \omega_f$ there exists δ_0, δ_1 such that $\omega_h \ll \delta_m \ll \widehat{\omega}$, for $m = 0, 1$, and that*

$$\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\lambda_{i_l}^{z_l})^2+\eta_l^2} = \mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\mu_{i_l}^{(l)}(ct_f)\nu_{z_l})^2+\eta_l^2} + \mathcal{O}(\Psi), \tag{7.151}$$

where $t_f = n^{-1+\omega_f}$, $\eta_l \in [n^{-1-\delta_0}, n^{-1+\delta_1}]$, and the error term is given by

$$\Psi := \frac{n^{\widehat{\omega}}}{n^{1+\omega}}\left(\sum_{l=1}^{2}\frac{1}{\eta_l}\right)\cdot\prod_{l=1}^{2}\left(1+\frac{n^{\xi}}{n\eta_l}\right) + \frac{n^{2\xi+2\delta_0}t_f}{n^{1/2}}\sum_{l=1}^{2}\frac{1}{\eta_l} + \frac{n^{2(\delta_1+\delta_0)}}{n^{\widehat{\omega}}}. \tag{7.152}$$

We remark that Ψ in (7.152) denotes a different error term compared with the error terms in (7.33) and (7.87).

By the definition of the processes $\boldsymbol{\mu}^{(l)}(t)$ in (7.143) it follows that $\boldsymbol{\mu}^{(l)}(t)$, $\boldsymbol{\mu}^{(m)}(t)$ are independent for $l \neq m$ and so that

$$\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\mu_{i_l}^{(l)}(ct_f)\nu_{z_l})^2+\eta_l^2} = \prod_{l=1}^{2}\mathbf{E}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\mu_{i_l}^{(l)}(ct_f)\nu_{z_l})^2+\eta_l^2}. \tag{7.153}$$

Then, similarly to Lemma 7.7.10, we conclude that

$$\prod_{l=1}^{2}\mathbf{E}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\lambda_{i_l}^{z_l})^2+\eta_l^2} = \prod_{l=1}^{2}\mathbf{E}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\mu_{i_l}^{(l)}(ct_f)\nu_{z_l})^2+\eta_l^2} + \mathcal{O}(\Psi). \tag{7.154}$$

Finally, combining (7.151)–(7.154) we conclude the proof of Proposition 7.7.2. □

We remark that in order to prove (7.154) it would not be necessary to introduce the additional comparison processes $\widetilde{\boldsymbol{\lambda}}^{(l)}$ and $\widetilde{\boldsymbol{\mu}}^{(l)}$ of Section 7.7.1.2, since in (7.154) the product is outside the expectation, so one can compare the expectations one by one; the correlation between these processes for different l's plays no role. Hence, already the usual coupling (see e.g. [42, 54, 129]) between the processes $\boldsymbol{\lambda}^{z_l}(t)$, $\boldsymbol{\mu}^{(l)}(t)$ defined in (7.133) and (7.143), respectively, would be sufficient to prove (7.154). On the other hand, the comparison processes $\mathring{\boldsymbol{\lambda}}^{z_l}(t)$ are anyway needed in order to remove the coefficients Λ_{ij} (which are small with very high probability) from the interaction term in (7.133).

We conclude this section with the proof of Lemma 7.7.10.

Proof of Lemma 7.7.10. In the following, to simplify notations, we assume that the scaling factors ν_{z_l} are equal to one. First of all, we notice that the summation over the indices

$n^{\widehat{\omega}} < |i| \le n$ in (7.125) can be removed, using the eigenvalue rigidity (7.121) similarly to [58, Eqs. (7.6)-(7.7)], at a price of an additional error term $n^{2(\delta_1+\delta_0)-\widehat{\omega}}$:

$$\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\lambda_{i_l}(H^{z_l}))^2+\eta_l^2}=\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n}\frac{\eta_l}{(\lambda_{i_l}(H^{z_l}))^2+\eta_l^2}+\mathcal{O}\left(\frac{n^{2(\delta_1+\delta_0)}}{n^{\widehat{\omega}}}\right). \tag{7.155}$$

The error term is negligible by choosing δ_0, δ_1 to be such that $\omega_h \ll \delta_m \ll \widehat{\omega}$, for $m = 0, 1$. Then, from the GFT Lemma 7.7.3, and (7.127), using (7.155) again, this time for $\lambda_{i_l}^{z_l}(ct_f)$, we have that

$$\begin{aligned}\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\lambda_{i_l}(H^{z_l}))^2+\eta_l^2}&=\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\lambda_{i_l}^{z_l}(ct_f))^2+\eta_l^2}\\&\quad+\mathcal{O}\left(\frac{n^{2\xi+2\delta_0}t_f}{n^{1/2}}\sum_{l=1}^{2}\frac{1}{\eta_l}+\frac{n^{2(\delta_1+\delta_0)}}{n^{\widehat{\omega}}}\right).\end{aligned} \tag{7.156}$$

We remark that the rigidity for $\lambda_{i_l}^{z_l}(ct_f)$ is obtained by Theorem 7.3.1 exactly as in (7.121). Next, by the same computations as in [58, Lemma 7.8] by writing the difference of l.h.s. and r.h.s. of (7.157) as a telescopic sum and then using the very high probability bound from Proposition 7.7.7 we get

$$\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\lambda_{i_l}^{z_l}(ct_f))^2+\eta_l^2}=\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\mathring{\lambda}_{i_l}^{(l)}(ct_f))^2+\eta_l^2}+\mathcal{O}(\Psi). \tag{7.157}$$

Similarly to (7.157), by Lemma 7.7.8 it also follows that

$$\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\mathring{\lambda}_{i_l}^{z_l}(ct_f))^2+\eta_l^2}=\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\widetilde{\lambda}_{i_l}^{(l)}(ct_f))^2+\eta_l^2}+\mathcal{O}(\Psi). \tag{7.158}$$

By (7.145) it readily follows that

$$\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\widetilde{\lambda}_{i_l}^{(l)}(ct_f))^2+\eta_l^2}=\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\widetilde{\mu}_{i_l}^{(l)}(ct_f))^2+\eta_l^2}. \tag{7.159}$$

Moreover, by (7.150), similarly to (7.157), we conclude

$$\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\widetilde{\mu}_{i_l}^{(l)}(ct_f))^2+\eta_l^2}=\mathbf{E}\prod_{l=1}^{2}\frac{1}{n}\sum_{|i_l|\le n^{\widehat{\omega}}}\frac{\eta_l}{(\mu_{i_l}^{(l)}(ct_f))^2+\eta_l^2}+\mathcal{O}(\Psi). \tag{7.160}$$

Combining (7.156)–(7.160), we conclude the proof of (7.151). □

Finally, we conclude Section 7.7.1 by listing the scales needed in the entire Section 7.7 and explain the dependences among them.

7.7.1.4 Relations among the scales in the proof of Proposition 7.7.2

Throughout Section 7.7 various scales are characterized by exponents of n, denoted by ω's, that we will also refer to scales for simplicity.

All the scales in the proof of Proposition 7.7.2 depend on the exponents $\omega_d, \omega_h, \omega_f \ll 1$. We recall that ω_d, ω_h are the exponents such that Lemma 7.7.4 on eigenvector overlaps holds under the assumption $|z_l - z_m|, |z_l - \overline{z}_m|, |z_l - \overline{z}_l| \geq n^{-\omega_d}$, and $|z_l| \leq 1 - n^{-\omega_h}$. The exponent ω_f determines the time $t_f = n^{-1+\omega_f}$ to run the DBM so that it reaches its local equilibrium and thus to prove the asymptotic independence of $\lambda_i^{z_l}(ct_f)$ and $\lambda_j^{z_m}(ct_f)$, with $c > 0$ defined in (7.124), for small indices i, j and $l \neq m$.

The most important scales in the proof of Proposition 7.7.2 are $\omega, \widehat{\omega}, \delta_0, \delta_1, \omega_E$. The scale ω_E is determined in Lemma 7.7.4 and it controls the correlations among the driving martingales originating from the eigenvector overlaps in (7.130)–(7.132). The scale ω gives the $n^{-1-\omega}$ precision of the coupling between various processes while $\widehat{\omega}$ determines the range of indices $|i| \leq n^{\widehat{\omega}}$ for which this coupling is effective. These scales are chosen much bigger than ω_h and they are determined in Proposition 7.7.7, Lemma 7.7.8 and Lemma 7.7.9, that describe these couplings. Each of these results gives an upper bound on the scales $\omega, \widehat{\omega}$, at the end we will choose the smallest of them. Finally, δ_0, δ_1 describe the scale of the range of the η's in Proposition 7.7.2. These two scales are determined in Lemma 7.7.10, given $\omega, \widehat{\omega}$ from the previous step. Putting all these steps together, we constructed $\omega, \widehat{\omega}, \delta_0, \delta_1$ claimed in Proposition 7.7.2 and hence also in Proposition 7.3.4. These scales are related as

$$\omega_h \ll \delta_m \ll \widehat{\omega} \ll \omega \ll \omega_f \ll \omega_E \ll 1, \qquad \omega_E = 4\omega_d, \tag{7.161}$$

for $m = 0, 1$.

Along the proof of Proposition 7.7.2 four auxiliary scales, $\omega_L, \omega_A, \omega_r, \omega_c$, are also introduced. The scale ω_L describes the range of interaction in the short range approximation processes $\widehat{\boldsymbol{x}}^{z_l}(t, \alpha)$ (see (7.179) later), while ω_A is the scale for which we can (partially) couple the driving martingales of the regularized processes $\mathring{\boldsymbol{\lambda}}^{z_l}(t)$ with the driving Brownian motions of Ginibre processes $\boldsymbol{\mu}^{(l)}(t)$. The scale ω_c is a cut-off in the energy estimate in Lemma 7.7.13, see (7.187). Finally, ω_r reduces the variance of the driving martingales by a factor $(1 + n^{-\omega_r})^{-1}$ to ensure the well-posedness of the processes $\mathring{\boldsymbol{\lambda}}^{z_l}(t)$, $\widetilde{\boldsymbol{\lambda}}^{(l)}(t)$, $\widetilde{\boldsymbol{\mu}}^{(l)}$, $\boldsymbol{x}^{z_l}(t, \alpha)$ defined in (7.141), (7.146), (7.148), and (7.167), respectively. These scales are inserted in the chain (7.161) as follows

$$\omega_h \ll \omega_A \ll \omega_f \ll \omega_L \ll \omega_c \ll \omega_r \ll \omega_E. \tag{7.162}$$

Note that there are no relations required among ω_A and $\omega, \widehat{\omega}, \delta_m$.

7.7.2 Universality and independence of the singular values of $X - z_1, X - z_2$ close to zero: Proof of Theorems 7.2.7 and 7.2.9

In the following we present only the proof of Theorem 7.2.9, since the proof of Theorem 7.2.7 proceeds exactly in the same way. Universality of the joint distribution of the singular values of $X - z_1$ and $X - z_2$ follows by universality for the joint distribution of the eigenvalues of H^{z_1} and H^{z_2}, which is defined in (7.2), since the eigenvalues of H^{z_l} are exactly the singular values of $X - z_l$ taken with positive and negative sign. From now on we only consider the eigenvalues of H^{z_l}, with $z_l \in \mathbf{C}$ such that $|\Im z_l| \sim 1$, $|z_1 - z_2|, |z_1 - \overline{z}_2| \sim 1$, and $|z_l| \leq 1 - \epsilon$ for some small fixed $\epsilon > 0$.

For $l \in [2]$, denote by $\{\lambda_i^{z_l}\}_{|i|\le n}$ the eigenvalues of H^{z_l} and by $\{\lambda_i^{z_l}(t)\}_{|i|\le n}$ their evolution under the DBM flow (7.133). Define $\{\mu_i^{(l)}(t)\}_{|i|\le n}$, for $l \in [2]$, to be the solution of (7.143) with initial data $\{\mu_i^{(l)}\}_{|i|\le n}$, which are the eigenvalues of independent complex Ginibre matrices $\widetilde{X}^{(1)}, \widetilde{X}^{(2)}$. Then, defining the comparison processes $\mathring{\boldsymbol{\lambda}}^{z_l}(t)$, $\widetilde{\boldsymbol{\lambda}}^{(l)}(t)$, $\widetilde{\boldsymbol{\mu}}^{(l)}(t)$ as in Sections 7.7.1.1–7.7.1.2, and combining Proposition 7.7.7, Lemma 7.7.8, and Lemma 7.7.9, we conclude that for any sufficiently small $\omega_f > 0$ there exist $\omega, \widehat{\omega} > 0$ such that $\widehat{\omega} \ll \omega \ll \omega_f$, and that

$$|\rho^{z_l}(0)\lambda_i^{z_l}(ct_f) - \rho_{\mathrm{sc}}(0)\mu_i^{(l)}(ct_f)| \le n^{-1-\omega}, \qquad |i| \le n^{\widehat{\omega}}, \tag{7.163}$$

with very high probability, with $c > 0$ defined in (7.124).

Then, by a simple Green's function comparison argument (GFT) as in Lemma 7.7.3, using (7.163), by exactly the same computations as in the proof of [57, Proposition 3.1 in Section 7] adapted to the bulk scaling, i.e. changing $\mathfrak{b}_{r,t_1} \to 0$ and $N^{3/4} \to 2n$, using the notation therein, we conclude Theorem 7.2.9.

7.7.3 Bound on the eigenvector overlaps

In this section we prove the bound on the eigenvector overlaps, as stated in Lemma 7.7.4. For any $T > 0$, and any $t \in [0, T]$, denote by ρ_t^z the self consistent density of states (scDOS) of the Hermitised matrix H_t^z, and define its quantiles by

$$\frac{i}{n} = \int_0^{\gamma_i^z(t)} \rho_t^z(x)\,\mathrm{d}x, \qquad i \in [n], \tag{7.164}$$

and $\gamma_{-i}^z(t) = -\gamma_i^z(t)$ for $i \in [n]$. Similarly to (7.121), as a consequence of Theorem 7.3.1 and the fact that the eigenvalues of $H_t^{z_l}$ are γ-Hölder continuous in time for any $\gamma \in (0, 1/2)$ by Weyl's inequality, by standard application of Helffer-Sjöstrand formula, we conclude the following rigidity bound

$$\sup_{0\le t\le T} |\lambda_i^{z_l}(t) - \gamma_i^{z_l}(t)| \le \frac{n^{100\omega_h}}{n^{2/3}(n+1-i)^{1/3}}, \qquad i \in [n], \tag{7.165}$$

with very high probability, uniformly in $|z_l| \le 1 - n^{-\omega_h}$. A bound similar to (7.165) holds for negative indices as well. We remark that the Hölder continuity of the eigenvalues of $H_t^{z_l}$ is used to prove (7.165) uniformly in time, using a standard grid argument.

The main input to prove Lemma 7.7.4 is Theorem 7.3.5 combined with Lemma 7.6.1.

Proof of Lemma 7.7.4. Recall that $P_1 \boldsymbol{w}_i^z = \boldsymbol{u}_i^z$ and $P_2 \boldsymbol{w}_i^z = \mathrm{sign}(i)\boldsymbol{v}_i^z$, for $|i| \le n$, by (7.128). In the following we consider $z, z' \in \mathbf{C}$ such that $|z|, |z'| \le 1 - n^{-\omega_h}$, $|z - z'| \ge n^{-\omega_d}$, for some sufficiently small $\omega_h, \omega_d > 0$.

Eigenvector overlaps can be estimated by traces of products of resolvents. More precisely, for any $\eta \ge n^{-2/3+\epsilon_*}$, for some small fixed $\epsilon_* > 0$, and any $|i_0|, |j_0| \le n$, using the rigidity bound (7.165), similarly to [58, Eq. (7.43)], we have that

$$\begin{aligned} |\langle \boldsymbol{u}_{i_0}^z(t), \boldsymbol{u}_{j_0}^{z'}(t)\rangle|^2 &\lesssim \eta^2 \operatorname{Tr}(\Im G^z(\gamma_{i_0}^z(t) + \mathrm{i}\eta))E_1(\Im G^{z'}(\gamma_{j_0}^{z'}(t) + \mathrm{i}\eta))E_1, \\ |\langle \boldsymbol{v}_{i_0}^z(t), \boldsymbol{v}_{j_0}^{z'}(t)\rangle|^2 &\lesssim \eta^2 \operatorname{Tr}(\Im G^z(\gamma_{i_0}^z(t) + \mathrm{i}\eta))E_2(\Im G^{z'}(\gamma_{j_0}^{z'}(t) + \mathrm{i}\eta))E_2, \end{aligned} \tag{7.166}$$

with E_1, E_2 defined in (7.63). By Theorem 7.3.5, combined with Lemma 7.6.1, choosing $\eta = n^{-12/23}$, say, the error term in the r.h.s. of (7.41) is bounded by $n^{-1/23}n^{2\omega_d+100\omega_h}$, hence we conclude the bound in (7.131) for any fixed time $t \in [0,T]$, choosing $\omega_E = -(2\omega_d + 100\omega_h - 1/23)$, for any $\omega_h \ll \omega_d \leq 1/100$.

Moreover, the bound (7.131) holds uniformly in time by a union bound, using a standard grid argument and Hölder continuity in the form

$$\|\Im G_t^z \Im G_t^{z'} - \Im G_s^z \Im G_s^{z'}\| \lesssim n^3 \left(\|H_t^z - H_s^z\| + \|H_t^{z'} - H_s^{z'}\|\right) \lesssim n^{7/2}|t-s|^{1/2}$$

for any $s, t \in [0,T]$, where the spectral parameters in the resolvents have imaginary parts at least $\eta > 1/n$. This concludes the proof of Lemma 7.7.4. □

7.7.4 Proof of Proposition 7.7.7

Throughout this section we use the notation $z = z_l$, with $l \in [2]$, with z_1, z_2 fixed as in Proposition Proposition 7.7.7.

Remark 7.7.11. *In the remainder of this section we assume that $|z| \leq 1-\epsilon$, with some positive $\epsilon > 0$ instead of $n^{-\omega_h}$, in order to make our presentation clearer. One may follow the ϵ-dependence throughout the proofs and find that all the estimates deteriorate with some fixed ϵ^{-1} power, say ϵ^{-100}. Thus, when $|z| \leq 1 - n^{-\omega_h}$ is assumed, we get an additional factor $n^{100\omega_h}$ but this does not play any role since ω_h is the smallest exponent (e.g. see Proposition 7.7.7) in the analysis of the processes* (7.133), (7.141).

The proof of Proposition 7.7.7 consists of several parts that we first sketch. The process $\mathring{\boldsymbol{\lambda}}^z(t)$ differs from $\boldsymbol{\lambda}^z(t)$ in three aspects: (i) the coefficients $\Lambda_{ij}^z(t)$ in the SDE (7.133) for $\boldsymbol{\lambda}^z(t)$ are removed; (ii) large values of the correlation of the driving martingales is cut off, and (iii) the martingale term is slightly reduced by a factor $(1+n^{\omega_r})^{-1/2}$. We deal with these differences in two steps. The substantial step is the first one, from Section 7.7.4.1 to Section 7.7.4.4, where we handle (i) by interpolation, using short range approximation and energy method. This is followed by a more technical second step in Section 7.7.4.5, where we handle (ii) and (iii) using a stopping time controlled by a well chosen Lyapunov function to show that the correlation typically remains below the cut-off level.

A similar analysis has been done in [53, Section 4] (which has been used in the singular value setup in [55, Eq. (3.13)]) but our more complicated setting requires major modifications. In particular, (7.133) has to be compared to [53, Eq. (4.1)] with $\mathrm{d}M_i = 0$, $Z_i = 0$, and identifying Λ_{ij}^z with γ_{ij}, using the notations therein. One major difference is that we now have a much weaker estimate $|\Lambda_{ij}^z| \leq n^{-\omega_E}$ than the bound $|\gamma_{ij}| \leq n^{-1+a}$, for some small fixed $a > 0$, used in [53]. We therefore need to introduce an additional cut-off function $\boldsymbol{\chi}$ in the energy estimate in Section 7.7.4.4.

7.7.4.1 Interpolation process

In order to compare the processes $\boldsymbol{\lambda}^z$ and $\mathring{\boldsymbol{\lambda}}^z$ from (7.133) and (7.141) we start with defining an interpolation process, for any $\alpha \in [0,1]$, as

$$\mathrm{d}x_i^z(t,\alpha) = \frac{\mathrm{d}\mathring{b}_i^z}{\sqrt{n(1+n^{-\omega_r})}} + \frac{1}{2n}\sum_{j\neq i} \frac{1+\alpha\mathring{\Lambda}_{ij}^z(t)}{x_i^z(t,\alpha) - x_j^z(t,\alpha)}\,\mathrm{d}t, \quad x_i^z(0,\alpha) = \lambda_i^z(0), \tag{7.167}$$

for any $|i| \leq n$. We recall that $\omega_f \ll \omega_r \ll \omega_E$. We use the notation $x_i^z(t,\alpha)$ instead of $z_i(t,\alpha)$ as in [53, Eq. (4.12)] to stress the dependence of $x_i^z(t,\alpha)$ on $z \in \mathbf{C}$. The well-posedness of the process (7.167) is proven in Appendix 7.A for any fixed $\alpha \in [0,1]$. In particular, the particles keep their order $x_i^z(t,\alpha) < x_{i+1}^z(t,\alpha)$. Additionally, by Lemma 7.A.2 it follows that the differentiation with respect to α of the process $\boldsymbol{x}^z(t,\alpha)$ is well-defined.

Note that the process $\boldsymbol{x}^z(t,\alpha)$ does not fully interpolate between $\mathring{\boldsymbol{\lambda}}^z(t)$ and $\boldsymbol{\lambda}^z(t)$; it handles only the removal of the $\mathring{\Lambda}_{ij}$ term. Indeed, it holds $\boldsymbol{x}^z(t,0) = \mathring{\boldsymbol{\lambda}}^z(t)$ for any $t \in [0,T]$, but $\boldsymbol{x}^z(t,1)$ is not equal to $\boldsymbol{\lambda}^z(t)$. Thus we will proceed in two steps as already explained:

1. The process $\boldsymbol{x}^z(t,\alpha)$ does not change much in $\alpha \in [0,1]$ for particles close to zero (by Lemma 7.7.13 below), i.e. $x_i^z(t,1) - x_i^z(t,0)$ is much smaller than the rigidity scale $1/n$ for small indices;
2. The process $\boldsymbol{x}^z(t,1)$ is very close to $\boldsymbol{\lambda}^z(t)$ for all indices (see Lemma 7.7.14 below).

We start with the analysis of the interpolation process $\boldsymbol{x}^z(t,\alpha)$, then in Section 7.7.4.5 we state and prove Lemma 7.7.14.

7.7.4.2 Local law for the interpolation process

In order to analyse the interpolation process $\boldsymbol{x}^z(t,\alpha)$, we first need to establish a local law for the Stieltjes transform of the empirical particle density. This will be used for a rigidity estimate to identify the location of $x_i(t,\alpha)$ with a precision $n^{-1+\epsilon}$, for some small $\epsilon > 0$, that is above the final target precision but it is needed as an a priori bound. Note that, unlike for $\boldsymbol{\lambda}^z(t)$, for $\boldsymbol{x}^z(t,\alpha)$ there is no obvious matrix ensemble behind this process, so local law and rigidity have to be proven directly from its defining equation (7.167).

Define the Stieltjes transform of the empirical particle density by

$$m_n(w,t,\alpha) = m_n^z(w,t,\alpha) := \frac{1}{2n} \sum_{|i| \leq n} \frac{1}{x_i^z(t,\alpha) - w}, \tag{7.168}$$

and denote the Stieltjes transform of ρ^z, the *self-consistent density of states (scDOS)* of H^z, by $m^z(w)$. Moreover, we denote the Stieltjes transform of ρ_t^z, the free convolution of ρ^z with the semicircular flow up to time t, by $m_t^z(w)$. Using the definition of the quantiles $\gamma_i^z(t)$ in (7.164), by Theorem 7.3.1 we have that

$$\begin{aligned} \sup_{|\Re w| \leq 10 c_1} \sup_{n^{-1+\gamma} \leq \Im w \leq 10} \sup_{\alpha \in [0,1]} |m_n(w,0,\alpha) - m^z(w)| &\leq \frac{n^\xi C_\epsilon}{n \Im w}, \\ \sup_{|i| \leq 10 c_2 n} \sup_{\alpha \in [0,1]} |x_i^z(0,\alpha) - \gamma_i^z(0)| &\leq \frac{C_\epsilon n^\xi}{n}, \end{aligned} \tag{7.169}$$

with very high probability for any $\xi > 0$, uniformly in $|z| \leq 1 - \epsilon$, for some small fixed $c_1, c_2, \gamma > 0$. We recall that $C_\epsilon \leq \epsilon^{-100}$. The rigidity bound in the second line of (7.169) follows by a standard application of Helffer-Sjöstrand formula.

In Lemma 7.7.12 we prove that (7.169) holds true uniformly in $0 \leq t \leq t_f$. For its proof, similarly to [53, Section 4.5], we follow the analysis of [114, Section 3.2] using (7.169) as an input.

Lemma 7.7.12 (Local law and rigidity). *Fix* $|z| \leq 1 - \epsilon$, *and assume that* (7.169) *holds with some* $\gamma, c_1, c_2, C_\epsilon > 0$, *then*

$$\begin{aligned} \sup_{|\Re w| \leq 10 c_1} \sup_{n^{-1+\gamma} \leq \Im w \leq 10} \sup_{\alpha \in [0,1]} \sup_{0 \leq t \leq t_f} |m_n^z(w,t,\alpha) - m_t(w)| &\leq \frac{C_\epsilon n^\xi}{n \Im w}, \\ \sup_{|i| \leq 10 c_2 n} \sup_{\alpha \in [0,1]} \sup_{0 \leq t \leq t_f} |x_i^z(t,\alpha) - \gamma_i^z(t)| &\leq \frac{C_\epsilon n^\xi}{n}, \end{aligned} \tag{7.170}$$

with very high probability for any $\xi > 0$, *with* $\gamma_i^z(t) \sim i/n$ *for* $|i| \leq 10 c_2 n$ *and* $t \in [0, t_f]$.

Proof. Differentiating (7.168), by (7.167) and Itô's formula, we get

$$\begin{aligned} \mathrm{d} m_n = {} & m_n (\partial_w m_n) \,\mathrm{d}t - \frac{1}{2 n^{3/2} \sqrt{1 + n^{-\omega_r}}} \sum_{|i| \leq n} \frac{\mathrm{d} \mathring{b}_i}{(x_i - w)^2} v \\ & + \frac{\alpha}{4n^2} \sum_{|i|,|j| \leq n} \frac{\mathring{\Lambda}_{ij}}{(x_i - w)^2 (x_j - w)} \,\mathrm{d}t \\ & + \frac{1}{4n^2} \sum_{|i| \leq n} \frac{\left[1 - \alpha - n^{-\omega_r} (1 + n^{-\omega_r})^{-1}\right] \mathring{\Lambda}_{ii}}{(x_i - w)^3} \,\mathrm{d}t. \end{aligned} \tag{7.171}$$

Note that by (7.139)–(7.140) it follows that

$$\mathring{\Lambda}_{ij}(t) = \Lambda_{ij}(t), \qquad (\mathring{b}_i(s))_{0 \leq s \leq t} = (b_i(s))_{0 \leq s \leq t}, \tag{7.172}$$

with very high probability uniformly in $0 \leq t \leq t_f$, where Λ_{ij} and $(b_i(s))_{0 \leq s \leq t}$ are defined in (7.129)–(7.132) and (7.134)–(7.135), respectively.

The equation (7.171) is the analogue of [114, Eq. (3.20)] with some differences. First, the last two terms are new and need to be estimated, although the penultimate term in (7.171) already appeared in [53, Eq. (4.62)] replacing $\mathring{\Lambda}_{ij}$ by $\widehat{\gamma}_{ij}$, using the notation therein. Second, the martingales in the second term in the r.h.s. of (7.171) are correlated. Hence, in order to apply the results in [114, Section 3.2] we prove that these additional terms are bounded as in [53, Eq. (4.64)]. Note that in [53, Eq. (4.64)] the corresponding term to the penultimate term in the r.h.s. of (7.171) is estimated using that $\widehat{\gamma}_{ij} \leq n^{-1+a}$, for some small $a > 0$. In our case, however, the bound on $|\mathring{\Lambda}|$ is much weaker and a crude estimate by absolute value is not affordable. We will use (7.172) and then the explicit form of Λ_{ij} in (7.129)–(7.132), that enables us to perform the two summations and write this term as the trace of the product of two operators (see (7.176) later).

Since $|\mathring{\Lambda}_{ii}| \leq n^{-\omega_E}$ by its definition below (7.140), the last term in (7.171) is easily bounded by

$$\left| \frac{1}{4n^2} \sum_{|i| \leq n} \frac{(1 - \alpha - n^{-\omega_r}(1 + n^{-\omega_r})^{-1}) \mathring{\Lambda}_{ii}}{(x_i - w)^3} \right| \leq \frac{\Im m_n(w)}{n^{1+\omega_E} (\Im w)^2}. \tag{7.173}$$

Next, we proceed with the estimate of the penultimate term in (7.171). Define the operators

$$T(t,\alpha) := \sum_{|i| \leq n} f(x_i(t,\alpha)) \boldsymbol{w}_i(t) [\boldsymbol{w}_i(t)]^*, \qquad S(t,\alpha) := \sum_{|i| \leq n} g(x_i(t,\alpha)) \overline{\boldsymbol{w}}_i(t) [\overline{\boldsymbol{w}}_i(t)]^*, \tag{7.174}$$

where $\{\boldsymbol{w}_i(t)\}_{|i|\le n}$ are the orthonormal eigenvectors in the definition of $\Lambda_{ij}(t)$ in (7.129), and for any fixed $w\in\mathbf{H}$ the functions $f,g\colon\mathbf{R}\to\mathbf{C}$ are defined as

$$f(x):=\frac{1}{(x-w)^2},\qquad g(x):=\frac{1}{x-w}. \tag{7.175}$$

Then, using the definitions (7.174)–(7.175) and (7.172), we bound the last term in the first line of (7.171) as

$$\begin{aligned}\left|\frac{\alpha}{4n^2}\sum_{|i|,|j|\le n}\frac{\mathring{\Lambda}_{ij}}{(x_i-w)^2(x_j-w)}\,\mathrm{d}t\right|&\\ &=\left|\frac{\alpha}{2n^2}\left[\mathrm{Tr}(P_1TP_2P_2SP_1)+\overline{\mathrm{Tr}(P_1TP_2P_2SP_1)}\right]\right|\\ &\lesssim\frac{1}{n^2}\left[\Im w\,\mathrm{Tr}[P_1TP_2(P_1TP_2)^*]+\frac{\mathrm{Tr}[P_1SP_2(P_1SP_2)^*]}{\Im w}\right]\\ &\lesssim\frac{1}{n^2}\left[\Im w\sum_{|i|\le n}|f(x_i)|^2+\frac{1}{\Im w}\sum_{|i|\le n}|g(x_i)|^2\right]\lesssim\frac{\Im m_n(w)}{n(\Im w)^2},\end{aligned} \tag{7.176}$$

with very high probability uniformly in $0\le t\le t_f$. Note that in the first equality of (7.176) we used that $\mathring{\Lambda}_{ij}(t)=\Lambda_{ij}(t)$ for any $0\le t\le t_f$ with very high probability by (7.172).

Finally, in order to conclude the proof, we estimate the martingale term in (7.171). For this purpose, using that $\mathbf{E}[\mathrm{d}\mathring{b}_i\,\mathrm{d}\mathring{b}_j\,|\,\mathcal{F}_{b,t}]=(\delta_{i,j}-\delta_{i,-j}+\mathring{\Lambda}_{ij})/2\,\mathrm{d}t$ and proceeding similarly to (7.176), we estimate its quadratic variation by

$$\begin{aligned}\frac{1}{4n^3(1+n^{-\omega_r})}\sum_{|i|,|j|\le n}\frac{\mathbf{E}[\mathrm{d}\mathring{b}_i\,\mathrm{d}\mathring{b}_j\mid\mathcal{F}_{b,t}]}{(x_i-w)^2(x_j-\overline{w})^2}&\\ &=\frac{1}{8n^3(1+n^{-\omega_r})}\sum_{|i|\le n}\frac{1}{|x_i-w|^4}\,\mathrm{d}t\\ &\quad+\frac{1}{8n^3(1+n^{-\omega_r})}\sum_{|i|\le n}\frac{1}{(x_i+w)^2(x_i-\overline{w})^2}\,\mathrm{d}t\\ &\quad+\frac{1}{8n^3(1+n^{-\omega_r})}\sum_{|i|,|j|\le n}\frac{\mathring{\Lambda}_{ij}}{(x_i-w)^2(x_j-\overline{w})^2}\,\mathrm{d}t\\ &\lesssim\frac{\Im m_n(w)}{n^2(\Im w)^3}+\frac{1}{n^3}\,\mathrm{Tr}[P_1TP_2(P_1TP_2)^*]\,\mathrm{d}t\\ &\lesssim\frac{\Im m_n(w)}{n^2(\Im w)^3},\end{aligned} \tag{7.177}$$

where the operator T is defined in (7.174), and in the penultimate inequality we used that $\mathring{\Lambda}_{ij}(t)=\Lambda_{ij}(t)$ for any $0\le t\le t_f$ with very high probability.

Combining (7.173), (7.176), and (7.177) we immediately conclude the proof of the first bound in (7.170) using the arguments of [114, Section 3.2]. The rigidity bound in the second line of (7.170) follows by a standard application of Helffer-Sjöstrand (see also below (7.169)). □

7.7.4.3 Short range approximation

Since the main contribution to the dynamics of $x_i^z(t,\alpha)$ comes from the nearby particles, in this section we introduce a *short range approximation* process $\widehat{\boldsymbol{x}}^z(t,\alpha)$, which will very well approximate the original process $\boldsymbol{x}^z(t,\alpha)$ (see (7.182) below). The actual interpolation analysis comparing $\alpha = 0$ and $\alpha = 1$ will then be performed on the short range process $\widehat{\boldsymbol{x}}^z(t,\alpha)$ in Section 7.7.4.4.

Fix $\omega_L > 0$ so that $\omega_f \ll \omega_L \ll \omega_E$, and define the index set

$$\mathcal{A} := \{(i,j) \,|\, |i-j| \le n^{\omega_L}\} \cup \{(i,j) \,|\, |i|,|j| > 5c_2 n\}, \tag{7.178}$$

with $c_2 > 0$ defined in (7.170). We remark that in [53, Eq. (4.69)] the notation ω_l is used instead of ω_L; we decided to change this notation in order to not create confusion with ω_l defined in [58, Eq. (7.67)]. Then we define the short range approximation $\widehat{\boldsymbol{x}}^z(t,\alpha)$ of the process $\boldsymbol{x}^z(t,\alpha)$ by

$$\begin{aligned} \mathrm{d}\widehat{x}_i^z(t,\alpha) &= \frac{\mathrm{d}\mathring{b}_i^z}{\sqrt{n}} + \frac{1}{2n}\sum_{\substack{j:(i,j)\in\mathcal{A},\\ j\ne i}} \frac{1+\alpha\mathring{\Lambda}_{ij}(t)}{\widehat{x}_i^z(t,\alpha)-\widehat{x}_j^z(t,\alpha)}\,\mathrm{d}t + \frac{1}{2n}\sum_{\substack{j:(i,j)\notin\mathcal{A},\\ j\ne i}} \frac{1}{x_i^z(t,0)-x_j^z(t,0)}\,\mathrm{d}t, \\ \widehat{x}_i^z(0,\alpha) &= x_i^z(0,\alpha), \qquad |i|\le n. \end{aligned} \tag{7.179}$$

The well-posedness of the process (7.179) follows by nearly identical computations as in the proof of Proposition 7.A.1.

In order to check that the *short range approximation* $\widehat{\boldsymbol{x}}^z(t,\alpha)$ is close to the process $\boldsymbol{x}^z(t,\alpha)$, defined in (7.167), we start with a trivial bound on $|x_i^z(t,\alpha) - x_i^z(t,0)|$ (see (7.180) below) to estimate the difference of particles far away from zero in (7.181), for which we do not have the rigidity bound in (7.170). Notice that by differentiating (7.167) in α and estimating $|\mathring{\Lambda}_{ij}|$ trivially by $n^{-\omega_E}$, it follows that

$$\sup_{0\le t\le t_f}\sup_{|i|\le n}\sup_{\alpha\in[0,1]} |x_i^z(t,\alpha)-x_i^z(t,0)| \lesssim n^{-\omega_E/2}, \tag{7.180}$$

similarly to [53, Lemma 4.3].

By the rigidity estimate (7.170), the weak global estimate (7.180) to estimate the contribution of the far away particles for which we do not know rigidity, and the bound $|\mathring{\Lambda}_{ij}| \le n^{-\omega_E}$ from (7.140) it follows that

$$\left| \frac{1}{2n}\sum_{\substack{j:(i,j)\notin\mathcal{A},\\ j\ne i}} \frac{1}{x_i^z(t,0)-x_j^z(t,0)} - \frac{1}{2n}\sum_{\substack{j:(i,j)\notin\mathcal{A},\\ j\ne i}} \frac{1+\alpha\mathring{\Lambda}_{ij}(t)}{x_i^z(t,\alpha)-x_j^z(t,\alpha)} \right| \lesssim n^{-\omega_E/2} + n^{-\omega_L+\xi}, \tag{7.181}$$

for any $\xi > 0$ with very high probability uniformly in $0 \le t \le t_f$. Hence, by exactly the same computations as in [129, Lemma 3.8], it follows that

$$\sup_{\alpha\in[0,1]}\sup_{|i|\le n}\sup_{0\le t\le t_f} |x_i^z(t,\alpha)-\widehat{x}_i^z(t,\alpha)| \le \frac{n^{2\omega_f}}{n}\left(\frac{1}{n^{\omega_E/2}}+\frac{1}{n^{\omega_L}}\right). \tag{7.182}$$

Note that (7.182) implies that the second estimate in (7.170) holds with x_i^z replaced by $\widehat{x}_i^z$. In order to conclude the proof of Proposition 7.7.7 in the next section we differentiate in the process $\widehat{\boldsymbol{x}}^z$ in α and study the deterministic (discrete) PDE we obtain from (7.179) after the α-derivation. Note that the α-derivative of $\widehat{\boldsymbol{x}}^z$ is well defined by Lemma 7.A.2.

7.7.4.4 Energy estimate

Define $v_i = v_i^z(t,\alpha) := \partial_\alpha \widehat{x}_i^z(t,\alpha)$, for any $|i| \le n$. In the remainder of this section we may omit the z-dependence since the analysis is performed for a fixed $z \in \mathbf{C}$ such that $|z| \le 1-\epsilon$, for some small fixed $\epsilon > 0$. By (7.179) it follows that $\boldsymbol{v}$ is the solution of the equation

$$\partial_t v_i = -(B\boldsymbol{v})_i + \xi_i, \quad v_i(0) = 0, \qquad |i| \le n, \tag{7.183}$$

where

$$(B\boldsymbol{v})_i := \sum_{j:(i,j)\in\mathcal{A}} B_{ij}(v_j - v_i),\ B_{ij} = B_{ij}(t,\alpha) := \frac{1+\alpha\mathring{\Lambda}_{ij}(t)}{2n(\widehat{x}_i(t,\alpha)-\widehat{x}_j(t,\alpha))^2}\mathbf{1}((i,j)\in\mathcal{A}), \tag{7.184}$$

and

$$\xi_i = \xi_i(t,\alpha) := \frac{1}{2n}\sum_{j:(i,j)\in\mathcal{A}} \frac{\mathring{\Lambda}_{ij}(t)}{\widehat{x}_i(t,\alpha)-\widehat{x}_j(t,\alpha)}.$$

Before proceeding with the optimal estimate of the ℓ^∞-norm of $\boldsymbol{v}$ in (7.186), we give the following crude bound

$$\sup_{|i|\le n}\sup_{0\le t\le t_f}\sup_{\alpha\in[0,1]} |v_i(t,\alpha)| \lesssim 1, \tag{7.185}$$

that will be needed as an a priori estimate for the more precise result later. The bound (7.185) immediately follows by exactly the same computations as in [53, Lemma 4.7] using that $|\mathring{\Lambda}_{ij}| \le n^{-\omega_E}$.

The main technical result to prove 1 towards Proposition 7.7.7 is the following lemma. In particular, after integration in α, Lemma 7.7.13 proves that the processes $\boldsymbol{x}^z(t,1)$ and $\boldsymbol{x}^z(t,0)$ are closer than the rigidity scale $1/n$.

Lemma 7.7.13. *For any small $\omega_f > 0$ there exist small constants $\omega, \widehat{\omega} > 0$ such that $\widehat{\omega} \ll \omega \ll \omega_f$ and*

$$\sup_{\alpha\in[0,1]}\sup_{|i|\le n^{\widehat{\omega}}}\sup_{0\le t\le t_f} |v_i(t)| \le n^{-1-\omega}, \tag{7.186}$$

with very high probability.

This lemma is based upon the finite speed of propagation mechanism for the dynamics (7.183) [91, Lemma 9.6]. Our proof follows [39, Lemma 6.2] that introduced a carefully chosen special cut-off function.

Proof. In order to bound $|v_i(t)|$ for small indices we will bound $\|\boldsymbol{v}\boldsymbol{\chi}\|_\infty$ for an appropriate cut-off vector $\boldsymbol{\chi}$ supported at a few coordinates around zero. More precisely, we will use an energy estimate to control $\|\boldsymbol{v}\boldsymbol{\chi}\|_2$ and then we use the trivial bound $\|\boldsymbol{v}\boldsymbol{\chi}\|_\infty \le \|\boldsymbol{v}\boldsymbol{\chi}\|_2$. This bound would be too crude without the cut-off.

Let $\varphi(x)$ be a smooth cut-off function which is equal to zero for $|x| \geq 1$, it is equal to one if $|x| \leq 1/2$. Fix a small constant $\omega_c > 0$ such that $\omega_f \ll \omega_L \ll \omega_c \ll \omega_E$, and define

$$\chi(x) := e^{-2xn^{1-\omega_c}} \varphi((2c_2)^{-1}x), \tag{7.187}$$

for any $x > 0$, with the constant $c_2 > 0$ defined in (7.170). It is trivial to see that χ is Lipschitz, i.e.

$$|\chi(x) - \chi(y)| \lesssim e^{-(x\wedge y)n^{1-\omega_c}} |x-y| n^{1-\omega_c}, \tag{7.188}$$

for any $x, y \geq 0$, and that

$$|\chi(x) - \chi(y)| \lesssim e^{-(x+y)n^{1-\omega_c}} |x-y| n^{1-\omega_c}, \tag{7.189}$$

if additionally $|x-y| \leq n^{\omega_c}/(2n)$. Finally we define the vector $\boldsymbol{\chi}$ by

$$\chi_i = \chi(\widehat{x}_i) := e^{-2|\widehat{x}_i| n^{1-\omega_c}} \varphi((2c_1)^{-1}\widehat{x}_i). \tag{7.190}$$

Note that χ_i is exponentially small if $n^{3\omega_c/2} \leq |i| \leq n$ by rigidity (7.170) and the fact that $\gamma_i^z \sim i/n$, for $n^{3\omega_c/2} \leq |i| \leq 10c_2 n$. We remark that the lower bound $n^{3\omega_c/2}$ on $|i|$ is arbitrary, since χ_i is exponentially small for any $|i|$ much bigger than n^{ω_c}. Moreover, as a consequence of (7.170) we have that

$$\widehat{x}_i \sim \frac{i}{n} \quad \text{for} \quad n^{\xi} \leq |i| \leq 10c_2 n, \tag{7.191}$$

with very high probability for any $\xi > 0$.

By (7.183) it follows that

$$\begin{aligned}
\partial_t \|\boldsymbol{v}\boldsymbol{\chi}\|_2^2 = \partial_t \sum_{|i|\leq n} v_i^2 \chi_i^2 &= -2\sum_i \chi_i^2 v_i (Bv)_i + \frac{1}{n} \sum_{(i,j)\in\mathcal{A}} \frac{\chi_i^2 v_i \mathring{\Lambda}_{ij}}{\widehat{x}_i - \widehat{x}_j} \\
&= -\sum_{(i,j)\in\mathcal{A}} B_{ij}(v_i\chi_i - v_j\chi_j)^2 + \frac{1}{2n} \sum_{(i,j)\in\mathcal{A}} \frac{(v_i\chi_i - v_j\chi_j)\mathring{\Lambda}_{ij}}{\widehat{x}_i - \widehat{x}_j}\chi_i \\
&\quad + \sum_{(i,j)\in\mathcal{A}} B_{ij} v_i v_j (\chi_i - \chi_j)^2 + \frac{1}{2n} \sum_{(i,j)\in\mathcal{A}} \frac{(\chi_i - \chi_j)\mathring{\Lambda}_{ij}}{\widehat{x}_i - \widehat{x}_j} v_j \chi_j,
\end{aligned} \tag{7.192}$$

where, in order to symmetrize the sums, we used that the operator B and the set $\mathcal{A}$ are symmetric, i.e. $B_{ij} = B_{ji}$ (see (7.184)) and $(i,j) \in \mathcal{A} \Leftrightarrow (j,i) \in \mathcal{A}$, and that $\mathring{\Lambda}_{ij} = \mathring{\Lambda}_{ji}$.

We start estimating the terms in the second line of the r.h.s. of (7.192). The most critical term is the first one because of the $(\widehat{x}_i - \widehat{x}_j)^{-2}$ singularity of B_{ij}. We write this term as

$$\sum_{(i,j)\in\mathcal{A}} B_{ij} v_i v_j (\chi_i - \chi_j)^2 = \left(\sum_{\substack{(i,j)\in\mathcal{A}, \\ |i-j|\leq n^{\omega_L}}} + \sum_{\substack{(i,j)\in\mathcal{A}, \\ |i-j|> n^{\omega_L}}} \right) B_{ij} v_i v_j (\chi_i - \chi_j)^2. \tag{7.193}$$

Then, using (7.189), $\|\boldsymbol{v}\|_\infty \lesssim 1$ by (7.185), $|\mathring{\Lambda}_{ij}| \le n^{-\omega_E}$ by (7.140), the rigidity (7.191), and that $\omega_L \ll \omega_c$, we bound the first sum by

$$\begin{aligned}
&\left|\sum_{\substack{(i,j)\in\mathcal{A},\\ |i-j|\le n^{\omega_L}}} B_{ij}v_iv_j(\chi_i-\chi_j)^2\right| \\
&\quad\lesssim \frac{1}{n}\sum_{\substack{(i,j)\in\mathcal{A},\\ |i-j|\le n^{\omega_L}}} \frac{1+|\mathring{\Lambda}_{ij}|}{(\widehat{x}_i-\widehat{x}_j)^2}|v_iv_j|\frac{n^2|\widehat{x}_i-\widehat{x}_j|^2}{n^{2\omega_c}}e^{-2(|\widehat{x}_i|+|\widehat{x}_j|)n^{1-\omega_c}} \\
&\quad\lesssim n^{1-2\omega_c}\left(\sum_{|i|,|j|\le n^{3\omega_c/2}} + \sum_{\substack{|i|\le n^{3\omega_c/2},|j|\ge n^{3\omega_c/2},\\ |i-j|\le n^{\omega_L}}}\right)|v_i||v_j|e^{-2(|\widehat{x}_i|+|\widehat{x}_j|)n^{1-\omega_c}} \\
&\quad\lesssim n^{1-\omega_c/2}\|\boldsymbol{v}\boldsymbol{\chi}\|_2^2+e^{-\frac{1}{2}n^{\omega_c/2}},
\end{aligned} \tag{7.194}$$

with very high probability. In the last inequality we trivially inserted φ to reproduce $\boldsymbol{\chi}$, using that $\varphi((2c_2)^{-1}|\widehat{x}_i|)=\varphi((2c_2)^{-1}|\widehat{x}_j|)=1$ with very high probability uniformly in $0\le t\le t_f$ if $|i|,|j|\le c_2 n$ by the rigidity estimate in (7.191).

Define the set

$$\mathcal{A}_1 := \{(i,j)\,|\,|i|,|j|\ge 5c_2 n\}\cap\{(i,j)\,|\,|i-j|>n^{\omega_L}\} = \mathcal{A}\cap\{(i,j)\,|\,|i-j|>n^{\omega_l}\},$$

which is symmetric. The second sum in (7.193), using (7.188), (7.185), and rigidity from (7.191), is bounded by

$$\left|\sum_{(i,j)\in\mathcal{A}_1} B_{ij}v_iv_j(\chi_i-\chi_j)^2\right| \lesssim n^{1-2\omega_c}\sum_{(i,j)\in\mathcal{A}_1} e^{-2(|\widehat{x}_i|\wedge|\widehat{x}_j|)n^{1-\omega_c}} \le e^{-n/2}, \tag{7.195}$$

with very high probability.

Next, we consider the second term in the second line of the r.h.s. of (7.192). Using (7.189), and that $|\mathring{\Lambda}_{ij}|\le n^{-\omega_E}$, proceeding similarly to (7.194)–(7.195), we bound this term as

$$\begin{aligned}
\left|\frac{1}{n}\sum_{(i,j)\in\mathcal{A}}\frac{(\chi_i-\chi_j)\mathring{\Lambda}_{ij}}{\widehat{x}_i-\widehat{x}_j}v_j\chi_j\right| &\lesssim \left|\frac{1}{n}\sum_{\substack{(i,j)\in\mathcal{A}\\ |i-j|\le n^{\omega_L}}}\frac{(\chi_i-\chi_j)\mathring{\Lambda}_{ij}}{\widehat{x}_i-\widehat{x}_j}v_j\chi_j\right| \\
&\quad+\left|\frac{1}{n}\sum_{(i,j)\in\mathcal{A}_1}\frac{(\chi_i-\chi_j)\mathring{\Lambda}_{ij}}{\widehat{x}_i-\widehat{x}_j}v_j\chi_j\right| \\
&\lesssim \sum_{\substack{(i,j)\in\mathcal{A}\\ |i-j|\le n^{\omega_L}}}\frac{|\mathring{\Lambda}_{ij}|}{|\widehat{x}_i-\widehat{x}_j|}\frac{|\widehat{x}_i-\widehat{x}_j|}{n^{\omega_c}}|v_j|\chi_j e^{-(|\widehat{x}_i|+|\widehat{x}_j|)n^{1-\omega_c}}+e^{-n/2} \\
&\lesssim \frac{1}{n^{\omega_c+\omega_E}}\sum_{|i|,|j|\le n^{3\omega_c/2}}|v_j|\chi_j+e^{-\frac{1}{2}n^{\omega_c/2}} \\
&\lesssim \frac{1}{n^{\omega_c/4+\omega_E}}\|\boldsymbol{v}\boldsymbol{\chi}\|_2+e^{-\frac{1}{2}n^{\omega_c/2}},
\end{aligned} \tag{7.196}$$

with very high probability uniformly in $0 \le t \le t_f$.

Finally, we consider the first line in the r.h.s. of (7.192). Since $1 + \alpha\mathring{\Lambda}_{ij} \ge 1/2$, we conclude that

$$\begin{aligned}\left|\frac{1}{n}\sum_{(i,j)\in\mathcal{A}}\frac{(v_i\chi_i - v_j\chi_j)\mathring{\Lambda}_{ij}}{\widehat{x}_i - \widehat{x}_j}\chi_i\right| &\le \frac{1}{C}\sum_{(i,j)\in\mathcal{A}} B_{ij}(v_i\chi_i - v_j\chi_j)^2 + \frac{C}{n}\sum_{(i,j)\in\mathcal{A}}|\mathring{\Lambda}_{ij}|^2\chi_i^2 \\ &\le \frac{1}{C}\sum_{(i,j)\in\mathcal{A}} B_{ij}(v_i\chi_i - v_j\chi_j)^2 + \frac{C}{n}\sum_{|i|,|j|\le n^{3\omega_c/2}}|\mathring{\Lambda}_{ij}|^2\chi_i^2 \\ &\quad + e^{-\frac{1}{2}n^{\omega_c/2}} \\ &\le \frac{1}{C}\sum_{(i,j)\in\mathcal{A}} B_{ij}(v_i\chi_i - v_j\chi_j)^2 + \frac{n^{3\omega_c}}{n^{1+2\omega_E}},\end{aligned} \tag{7.197}$$

for some large $C > 0$. The error term in the r.h.s. of (7.197) is affordable since $\omega_c \ll \omega_E$.

Hence, combining (7.192)–(7.197), we conclude that

$$\partial_t\|\boldsymbol{v\chi}\|_2^2 \lesssim -\frac{1}{2}\sum_{(i,j)\in\mathcal{A}} B_{ij}(v_i\chi_i - v_j\chi_j)^2 + n^{1-\omega_c/2}\|\boldsymbol{v\chi}\|_2^2 + n^{-\omega_c/4-\omega_E}\|\boldsymbol{v\chi}\|_2 + \frac{n^{3\omega_c}}{n^{1+2\omega_E}}, \tag{7.198}$$

with very high probability uniformly in $0 \le t \le t_f$. Then, ignoring the negative first term, integrating (7.198) from 0 to $t_f = n^{-1+\omega_f}$, and using that $n^{1-\omega_c/2}t_f = n^{\omega_f-\omega_c/2}$ with $\omega_f \ll \omega_c \ll \omega_E$, we get

$$\sup_{0\le t\le t_f}\|\boldsymbol{v\chi}\|_2^2 \le \frac{n^{3\omega_c}t_f}{n^{1+2\omega_E}}.$$

Hence, using the bound

$$\sup_{0\le t\le t_f}\sup_{|i|\le n^{\widehat{\omega}}}|v_i(t)| \le \sup_{0\le t\le t_f}\|\boldsymbol{v\chi}\|_2 \le \sqrt{\frac{n^{3\omega_c}t_f}{n^{1+2\omega_E}}},$$

we conclude (7.186) for some $\omega, \widehat{\omega} > 0$ such that $\widehat{\omega} \ll \omega \ll \omega_f \ll \omega_L \ll \omega_c \ll \omega_E$. □

With this proof we completed the main 1 in the proof of Proposition 7.7.7, the analysis of the interpolation process $\boldsymbol{x}^z(t,\alpha)$.

7.7.4.5 The processes $\boldsymbol{\lambda}(t)$ and $\boldsymbol{x}^z(t,1)$ are close

In 2 towards the proof of Proposition 7.7.7, we now prove that the processes $\boldsymbol{\lambda}(t)$ and $\boldsymbol{x}^z(t,1)$ are very close for any $t \in [0, t_f]$:

Lemma 7.7.14. *Let $\boldsymbol{\lambda}^z(t)$, $\boldsymbol{x}^z(t,1)$ be defined in* (7.133) *and* (7.167)*, respectively, and let* $t_f = n^{-1+\omega_f}$*, then*

$$\sup_{|i|\le n}\sup_{0\le t\le t_f}|x_i^z(t,1) - \lambda_i^z(t)| \lesssim \frac{n^{\omega_f}}{n^{1+\omega_r}}. \tag{7.199}$$

with very high probability.

Proof of Proposition 7.7.7. Proposition 7.7.7 follows by exactly the same computations as in [53, Section (4.10)], combining (7.199), (7.182), (7.185)–(7.186). □

Proof of Lemma 7.7.14. The proof of this lemma closely follows [53, Lemma 4.2]. We remark that in our case $\mathrm{d}M_i = Z_i = 0$ compared to [53, Lemma 4.2], using the notation therein. Recall the definitions of $C(t), \Lambda_{ij}^{z_l}(t), \Theta_{ij}^{z_1,z_2}(t), \Theta_{ij}^{z_1,\overline{z}_2}(t)$ and $\mathring{C}(t), \mathring{\Lambda}_{ij}^{z_l}(t), \mathring{\Theta}_{ij}^{z_1,z_2}(t), \mathring{\Theta}_{ij}^{z_1,\overline{z}_2}(t)$ in (7.137), (7.129),(7.132) and (7.139)–(7.140), respectively. In the following we may omit the z-dependence. Introduce the stopping times

$$\tau_1 := \inf\Big\{t \geq 0 \,\Big|\, \exists |i|, |j| \leq n; l \in [2] \text{ s.t. } |\Lambda_{ij}^{z_l}(t)| + |\Theta_{ij}^{z_1,z_2}(t)| + |\Theta_{ij}^{z_1,\overline{z}_2}(t)| > n^{-\omega_E}\Big\}, \tag{7.200}$$

$$\tau_2 := \inf\{t \geq 0 \mid \exists |i| \leq n \text{ s.t. } |x_i(t,1)| + |\lambda_i(t)| > 2R\}, \tag{7.201}$$

for some large $R > 0$, and

$$\tau := \tau_1 \wedge \tau_2 \wedge t_f. \tag{7.202}$$

Note that $|\lambda_i(t)| \leq R$ with very high probability, since $\boldsymbol{\lambda}(t)$ are the eigenvalues of H_t^z, whose norm is typically bounded. Furthermore, by (7.180) and the fact that the process $\boldsymbol{x}(t,0)$ stays bounded by [114, Section 3] it follows that $|x_i(t,\alpha)| \leq R$ for any $t \in [0,t_f]$ and $\alpha \in [0,1]$. We remark that the analysis in [114, Section 3] is done for a process of the form (7.167), with $\alpha = 0$, when it has i.i.d. driving Brownian motions, but the same results apply for our case as well since the correlation in (7.139) does not play any role (see (7.177)). This, together with Lemma 7.7.13 applied for $z = z_1, z' = z_2$ and $z = z_1, z' = \overline{z}_2$ and $z = z_l, z' = \overline{z}_l$, implies that

$$\tau = t_f$$

with very high probability. In particular, $\mathring{\Theta}_{ij}(t) = \Theta_{ij}(t)$ for any $t \leq \tau$, hence

$$C(t) = \mathring{C}(t) \tag{7.203}$$

for any $t \leq \tau$.

In the remainder of the proof, omitting the time- and z-dependence, we use the notation $\boldsymbol{x} = \boldsymbol{x}^z(t,1)$, $\boldsymbol{\lambda} = \boldsymbol{\lambda}(t)$. Define

$$u_i := \lambda_i - x_i, \qquad |i| \leq n,$$

then, as a consequence of (7.203), subtracting (7.133) and (7.167), it follows that

$$\mathrm{d}u_i = \sum_{j \neq i} B_{ij}(u_j - u_i)\,\mathrm{d}t + \frac{A_n}{\sqrt{n}}\,\mathrm{d}b_i, \tag{7.204}$$

for any $0 \leq t \leq \tau$, where

$$B_{ij} = \frac{1 + \Lambda_{ij}}{2n(\lambda_i - \lambda_j)(x_i - x_j)} > 0, \tag{7.205}$$

since $|\Lambda_{ij}(t)| = |\mathring{\Lambda}_{ij}(t)| \le n^{-\omega_E}$, and

$$A_n = \frac{1}{\sqrt{1+n^{-\omega_r}}} - 1 = \mathcal{O}(n^{-\omega_r}). \tag{7.206}$$

Let $\nu := n^{1+\omega_r}$, and define the Lyapunov function

$$F(t) := \frac{1}{\nu}\log\left(\sum_{|i|\le n} e^{\nu u_i(t)}\right). \tag{7.207}$$

By Itô's lemma, for any $0 \le t \le \tau$, we have that

$$\begin{aligned} \mathrm{d}F = {} & \frac{1}{\sum_{|i|\le n} e^{\nu u_i}} \sum_{|i|\le n} e^{\nu u_i} \sum_{j\ne i} B_{ij}(u_j - u_i)\,\mathrm{d}t + \frac{n^{-1/2}A_n}{\sum_{|i|\le n} e^{\nu u_i}} \sum_{|i|\le n} e^{\nu u_i}\,\mathrm{d}b_i \\ & + \frac{n^{-1}\nu A_n^2}{4\sum_{|i|\le n} e^{\nu u_i}} \sum_{|i|\le n} e^{\nu u_i}(1+\Lambda_{ii})\,\mathrm{d}t - \frac{4n^{-1}\nu A_n^2}{\left(\sum_{|i|\le n} e^{\nu u_i}\right)^2} \sum_{|i|,|j|\le n} e^{\nu u_i} e^{\nu u_j}\,\mathbf{E}\Big[\mathrm{d}b_i\,\mathrm{d}b_j \,\Big|\, \widetilde{\mathcal{F}}_{b,t}\Big]. \end{aligned} \tag{7.208}$$

Note that the first term in the r.h.s. of (7.208) is negative since the map $x \mapsto e^{\nu x}$ is increasing. The second and third term in the r.h.s. of (7.208), using that $1 + \Lambda_{ii} \le 2$, are bounded exactly as in [53, Eqs. (4.37)–(4.38)] by

$$\frac{n^{\xi} t_f^{1/2}}{n^{1/2+\omega_r}} + \frac{t_f \nu}{n^{1+2\omega_r}},$$

with very high probability for any $\xi > 0$.

Note that

$$\sum_{|i|,|j|\le n} e^{\nu u_i} e^{\nu u_j}\,\mathbf{E}\Big[\mathrm{d}b_i\,\mathrm{d}b_j \,\Big|\, \widetilde{\mathcal{F}}_{b,t}\Big] \ge 0,$$

hence, the last term in the r.h.s. of (7.208) is always non positive. This implies that

$$\sup_{0\le t\le t_f} F(t) \le F(0) + \frac{t_f \nu A_n^2}{n} + \frac{n^{\xi} t_f^{1/2} A_n}{n^{1/2}},$$

for any $\xi > 0$. Then, since

$$F(0) = \frac{\log(2n)}{n^{1+\omega_r}}, \qquad F(t) \ge \sup_{|i|\le n} u_i(t),$$

we conclude the upper bound in (7.199). Then noticing that $u_{-i} = -u_i$ for $i \in [n]$, we conclude the lower bound as well. □

7.7.5 Path-wise coupling close to zero: Proof of Lemmata 7.7.8–7.7.9

This section is the main technical result used in the proof of Lemmata 7.7.8–7.7.9. In Proposition 7.7.16 we will show that the points with small indices in the two processes become very close to each other on a certain time scale $t_f = n^{-1+\omega_f}$, for any small $\omega_f > 0$.

The main result of this section (Proposition 7.7.16) is stated for general deterministic initial data $\boldsymbol{s}(0)$ satisfying a certain regularity condition (see Definition 7.7.15 later) even if for its applications in the proof of Proposition 7.7.2 we only consider initial data which are eigenvalues of i.i.d. random matrices. The initial data $\boldsymbol{r}(0)$, without loss of generality, are assumed to be the singular values of a Ginibre matrix (see also below (7.210) for a more detailed explanation). For notational convenience we formulate the result for two general processes $\boldsymbol{s}$ and $\boldsymbol{r}$ and later we specialize them to our application.

Fix a small constant $0 < \omega_r \ll 1$, and define the processes $s_i(t)$, $r_i(t)$ to be the solution of

$$\mathrm{d}s_i(t) = \sqrt{\frac{1}{2n(1+n^{-\omega_r})}}\,\mathrm{d}\mathfrak{b}_i^s(t) + \frac{1}{2n}\sum_{j\neq i}\frac{1}{s_i(t)-s_j(t)}\,\mathrm{d}t, \qquad 1 \le |i| \le n, \quad (7.209)$$

and

$$\mathrm{d}r_i(t) = \sqrt{\frac{1}{2n(1+n^{-\omega_r})}}\,\mathrm{d}\mathfrak{b}_i^r(t) + \frac{1}{2n}\sum_{j\neq i}\frac{1}{r_i(t)-r_j(t)}\,\mathrm{d}t, \qquad 1 \le |i| \le n, \quad (7.210)$$

with initial data $s_i(0) = s_i$, $r_i(0) = r_i$, where $\boldsymbol{s} = \{s_{\pm i}\}_{i\in[n]}$ and $\boldsymbol{r} = \{r_{\pm i}\}_{i\in[n]}$ are two independent sets of particles such that $s_{-i} = -s_i$ and $r_{-i} = -r_i$ for $i \in [n]$. The driving martingales $\{\mathfrak{b}_i^s\}_{i\in[n]}$, $\{\mathfrak{b}_i^r\}_{i\in[n]}$ in (7.209)–(7.210) are two families satisfying Assumption (7.B) below, and they are such that $\mathfrak{b}_{-i}^s = -\mathfrak{b}_i^s$, $\mathfrak{b}_{-i}^r = -\mathfrak{b}_i^r$ for $i \in [n]$. The coefficient $(1+n^{-\omega_r})^{-1/2}$ ensures the well-posedness of the processes (7.209)–(7.210) (see Appendix 7.A), but it does not play any role in the proof of Proposition 7.7.16 below.

For convenience we also assume that $\{r_{\pm i}\}_{i=1}^n$ are the singular values of $\widetilde{X}$, with $\widetilde{X}$ a Ginibre matrix. This is not a restriction; indeed, once a process with general initial data $\boldsymbol{s}$ is shown to be close to the reference process with Ginibre initial data, then processes with any two initial data will be close.

On the correlation structure between the two families of i.i.d. Brownian motions $\{\mathfrak{b}_i^s\}_{i=1}^n$, $\{\mathfrak{b}_i^r\}_{i=1}^n$ and the initial data $\{s_{\pm i}\}_{i\in[n]}$ we make the following assumptions.

Assumption (7.B). *Fix $\omega_K, \omega_Q > 0$ such that $\omega_K \ll \omega_r \ll \omega_Q \ll 1$, with ω_r defined in* (7.209)–(7.210), *and define the n-dependent parameter $K = K_n = n^{\omega_K}$. Suppose that the families $\{\mathfrak{b}_{\pm i}^s\}_{i=1}^n$, $\{\mathfrak{b}_{\pm i}^r\}_{i=1}^n$ in* (7.209)–(7.210) *are realised on a common probability space with a common filtration $\mathcal{F}_t$. Let*

$$L_{ij}(t)\,\mathrm{d}t := \mathbf{E}\Big[\big(\mathrm{d}\mathfrak{b}_i^s(t) - \mathrm{d}\mathfrak{b}_i^r(t)\big)\big(\mathrm{d}\mathfrak{b}_j^s(t) - \mathrm{d}\mathfrak{b}_j^r(t)\big)\,\Big|\,\mathcal{F}_t\Big] \quad (7.211)$$

denote the covariance of the increments conditioned on $\mathcal{F}_t$. The processes satisfy the following assumptions:

1. *The two families of martingales $\{\mathfrak{b}_i^s\}_{i=1}^n$, $\{\mathfrak{b}_i^r\}_{i=1}^n$ are such that*

$$\mathbf{E}\Big[\mathrm{d}\mathfrak{b}_i^{q_1}(t)\,\mathrm{d}\mathfrak{b}_j^{q_2}(t)\,\Big|\,\mathcal{F}_t\Big] = \big[\delta_{ij}\delta_{q_1q_2} + \Xi_{ij}^{q_1,q_2}(t)\big]\,\mathrm{d}t, \quad |\Xi_{ij}^{q_1,q_2}(t)| \le n^{-\omega_Q}, \quad (7.212)$$

for any $i,j \in [n]$, $q_1,q_2 \in \{s,r\}$. The quantities in (7.212) *for negative i,j-indices are defined by symmetry.*

2. *The subfamilies* $\{\mathfrak{b}^s_{\pm i}\}_{i=1}^K$, $\{\mathfrak{b}^r_{\pm i}\}_{i=1}^K$ *are very strongly dependent in the sense that for any* $|i|, |j| \le K$ *it holds*
$$|L_{ij}(t)| \le n^{-\omega_Q} \tag{7.213}$$
with very high probability for any fixed $t \ge 0$.

Definition 7.7.15 ((g, G)-regular points [58, Definition 7.12]). *Fix a very small* $\nu > 0$, *and choose* g, G *such that*
$$n^{-1+\nu} \le g \le n^{-2\nu}, \qquad G \le n^{-\nu}.$$
A set of $2n$*-points* $\boldsymbol{s} = \{s_i\}_{|i|\le n}$ *on* $\mathbf{R}$ *is called* (g, G)-regular *if there exist constants* $c_\nu, C_\nu > 0$ *such that*
$$c_\nu \le \frac{1}{2n}\Im \sum_{i=-n}^{n} \frac{1}{s_i - (E + \mathrm{i}\eta)} \le C_\nu, \tag{7.214}$$
for any $|E| \le G$, $\eta \in [g, 10]$, *and if there is a constant* C_s *large enough such that* $\|\boldsymbol{s}\|_\infty \le n^{C_s}$. *Moreover,* $c_\nu, C_\nu \sim 1$ *if* $\eta \in [g, n^{-2\nu}]$ *and* $c_\nu \ge n^{-100\nu}$, $C_\nu \le n^{100\nu}$ *if* $\eta \in [n^{-2\nu}, 10]$.

Let $\rho_{\mathrm{fc},t}(E)$ be the scDOS of the particles $\{s_{\pm i}(t)\}_{i\in[n]}$ that is given by the semicircular flow acting on the scDOS of the initial data $\{s_{\pm i}(0)\}_{i\in[n]}$, see [129, Eqs. (2.5)–(2.6)].

Proposition 7.7.16 (Path-wise coupling close to zero). *Let the processes* $\boldsymbol{s}(t) = \{s_{\pm i}(t)\}_{i\in[n]}$, $\boldsymbol{r}(t) = \{r_{\pm i}(t)\}_{i\in[n]}$ *be the solutions of* (7.209) *and* (7.210), *respectively, and assume that the driving martingales in* (7.209)–(7.210) *satisfy Assumption (7.B) for some* $\omega_K, \omega_Q > 0$. *Additionally, assume that* $\boldsymbol{s}(0)$ *is* (g, G)*-regular in the sense of Definition 7.7.15 and that* $\boldsymbol{r}(0)$ *are the singular values of a Ginibre matrix. Then for any small* $\omega_f, \nu > 0$ *such that* $\nu \ll \omega_K \ll \omega_f \ll \omega_Q$ *and that* $gn^\nu \le t_f \le n^{-\nu}G^2$, *there exist constants* $\omega, \widehat{\omega} > 0$ *such that* $\nu \ll \widehat{\omega} \ll \omega \ll \omega_f$, *and*
$$|\rho_{\mathrm{fc},t_1}(0)s_i(t_f) - \rho_{\mathrm{sc}}(0)r_i(t_f)| \le n^{-1-\omega}, \qquad |i| \le n^{\widehat{\omega}}, \tag{7.215}$$
with very high probability, where $t_f := n^{-1+\omega_f}$.

Proof. The proof of Proposition 7.7.16 is nearly identical to the proof of [58, Proposition 7.14], which itself follows the proof of fixed energy universality in [42, 129], adapted to the block structure (7.25) in [54] (see also [40] for a different technique to prove universality, adapted to the block structure in [208]). We will not repeat the whole proof, just explain the modification. The only difference of Proposition 7.7.16 compared to [58, Proposition 7.14] is that here we allow the driving martingales in (7.209)–(7.210) to have a (small) correlation (compare Assumption (7.B) with a non zero $\Xi_{ij}^{q_1,q_2}$ to [58, Assumption 7.11]). The additional pre-factor $(1 + n^{-\omega_r})^{-1/2}$ does not play any role.

The correlation of the driving martingales in (7.209)–(7.210) causes a difference in the estimate of [58, Eq. (7.83)]. In particular, the bound on
$$\mathrm{d}M_t = \frac{1}{2n}\sum_{|i|\le n}(w_i - f_i)f_i' \,\mathrm{d}C_i(t, \alpha), \qquad \mathrm{d}C_i(t, \alpha) := \frac{\alpha\,\mathrm{d}\mathfrak{b}^s + (1-\alpha)\,\mathrm{d}\mathfrak{b}^r}{\sqrt{2n(1 + n^{-\omega_r})}}, \tag{7.216}$$
using the notation in [58, Eq. (7.83)], will be slightly different. In the remainder of the proof we present how [58, Eqs. (7.83)-(7.87)] changes in the current setup. Using that by [129, Eqs. (3.119)–(3.120)] we have
$$|f_i| + |f_i'| + |w_i| \le n^{-D}, \qquad n^{\omega_A} < |i| \le n, \tag{7.217}$$

for $\omega_A = \omega_K$ (with ω_K defined in Assumption (7.B)), and for any $D > 0$ with very high probability, we bound the quadratic variation of (7.216) by

$$\mathrm{d}\langle M\rangle_t = \frac{1}{4n^2} \sum_{1\le|i|,|j|\le n^{\omega_A}} (w_i - f_i)(w_j - f_j) f_i' f_j' \, \mathbf{E}[\mathrm{d}C_i(\alpha,t)\,\mathrm{d}C_j(\alpha,t) \mid \mathcal{F}_t] + \mathcal{O}\left(n^{-100}\right). \tag{7.218}$$

We remark that here we estimated the regime when $|i|$ or $|j|$ are larger than n^{ω_A} differently compared to [58, Eq. (7.84)], since, unlike in [58, Eq. (7.84)], $\mathbf{E}[\mathrm{d}C_i(t,\alpha)\,\mathrm{d}C_j(t,\alpha) \mid \mathcal{F}_t] \neq \delta_{ij}$, hence here we anyway need to estimate the double sum using (7.217).

Then, by 1–2 of Assumption (7.B), for $|i|, |j| \le n^{\omega_A}$ we have

$$\begin{aligned} \mathbf{E}[\mathrm{d}C_i(t,\alpha)\,\mathrm{d}C_j(t,\alpha) \mid \mathcal{F}_t] &= \frac{\delta_{ij} + \alpha^2 \Xi_{ij}^{s,s}(t) + (1-\alpha)^2 \Xi_{ij}^{r,r}(t)}{2n(1+n^{-\omega_r})}\,\mathrm{d}t \\ &\quad + \frac{\alpha(1-\alpha)}{2n(1+n^{-\omega_r})}\,\mathbf{E}\left[(\mathrm{d}\mathfrak{b}_i^s\,\mathrm{d}\mathfrak{b}_j^r + \mathrm{d}\mathfrak{b}_i^r\,\mathrm{d}\mathfrak{b}_j^s) \,\middle|\, \mathcal{F}_t\right], \end{aligned} \tag{7.219}$$

and that

$$\begin{aligned} \left|\mathbf{E}\left[\mathrm{d}\mathfrak{b}_i^s\,\mathrm{d}\mathfrak{b}_j^r \,\middle|\, \mathcal{F}_t\right]\right| &= \left|\mathbf{E}\left[(\mathrm{d}\mathfrak{b}_i^s - \mathrm{d}\mathfrak{b}_i^r)\,\mathrm{d}\mathfrak{b}_j^r \,\middle|\, \mathcal{F}_t\right] + (\delta_{ij} + \Xi_{ij}^{r,r}(t))\,\mathrm{d}t\right| \\ &\lesssim (|L_{ii}(t)|^{1/2} + |\Xi_{ij}^{r,r}(t)| + \delta_{ij})\,\mathrm{d}t, \end{aligned} \tag{7.220}$$

where in the last step we used Kunita-Watanabe inequality for the quadratic variation $(\mathrm{d}\mathfrak{b}_i^s - \mathrm{d}\mathfrak{b}_i^r)\,\mathrm{d}\mathfrak{b}_j^r$.

Combining (7.218)–(7.220), and adding back the sum over $n^{\omega_A} < |i| \le n$ of $(w_i - f_i)^2 (f_i')^2$ at the price of an additional error $\mathcal{O}(n^{-100})$, omitting the t-dependence, we finally conclude that

$$\begin{aligned} \mathrm{d}\langle M\rangle_t &\lesssim \frac{1}{n^3} \sum_{1\le|i|\le n} (w_i - f_i)^2 (f_i')^2\,\mathrm{d}t \\ &\quad + \frac{1}{n^3} \sum_{|i|,|j|\le n^{\omega_A}} \left(|L_{ii}|^{1/2} + |\Xi_{ij}^{s,s}| + |\Xi_{ij}^{r,r}|\right) \left|(w_i - f_i)(w_j - f_j) f_i' f_j'\right| \mathrm{d}t + \mathcal{O}\left(n^{-100}\right). \end{aligned} \tag{7.221}$$

Since $|L_{ii}| + |\Xi_{ij}^{q_1,q_2}| \le n^{-\omega_Q}$, for any $|i|, |j| \le n$, $q_1, q_2 \in \{s, r\}$, and $\omega_A = \omega_K \ll \omega_Q$ by (7.212)–(7.213), using Cauchy-Schwarz in (7.221), we conclude that

$$\mathrm{d}\langle M\rangle_t \lesssim \frac{1}{n^3} \sum_{1\le|i|\le n} (w_i - f_i)^2 (f_i')^2\,\mathrm{d}t + \mathcal{O}\left(n^{-100}\right), \tag{7.222}$$

which is exactly the same bound as in [58, Eq. (7.88)] (except for the tiny error $\mathcal{O}(n^{-100})$ that is negligible). Proceeding exactly as in [58], we conclude the proof of Proposition 7.7.16. □

7.7.5.1 Proof of Lemma 7.7.8 and Lemma 7.7.9

The fact that the processes $\mathring{\boldsymbol{\lambda}}(t)$, $\widetilde{\boldsymbol{\lambda}}(t)$ and $\widetilde{\boldsymbol{\mu}}(t)$, $\boldsymbol{\mu}(t)$ satisfy the hypotheses of Proposition 7.7.16 for the choices $\nu = \omega_h$, $\omega_K = \omega_A$, $\omega_Q = \omega_E$, and $\Xi_{ij}^{q_1,q_2} = \Theta_{ij}^{z_1,\overline{z}_2}$ follows by Lemma 7.7.4 applied for $z = z_1, z' = z_2$ and $z = z_1, z' = \overline{z}_2$ and $z = z_l, z' = \overline{z}_l$, and exactly the same computations as in [58, Section 7.5]. We remark that the processes $\boldsymbol{\mu}^{(l)}(t)$ do not have the additional coefficient $(1+n^{-\omega_r})$ in the driving Brownian motions, but this does not play any role in the application of Proposition 7.7.16 since it causes an error term $n^{-1-\omega_r}$ that is much smaller then the bound $n^{-1-\omega}$ in (7.150). Then, by Proposition 7.7.16, the results in Lemma 7.7.8 and Lemma 7.7.9 immediately follow. □

7.7.6 Proof of Proposition 7.7.6

First of all we notice that $\boldsymbol{\lambda}(t)$ is γ-Hölder continuous for any $\gamma \in (0,1/2)$ by Weyl's inequality. Then the proof of Proposition 7.7.6 consists of two main steps, (i) proving that the eigenvalues $\boldsymbol{\lambda}(t)$ are a strong solution of (7.133) as long as there are no collisions, and (ii) proving that there are no collisions for almost all $t \in [0,T]$.

The proof that the eigenvalues $\boldsymbol{\lambda}(t)$ are a solution of (7.133) is deferred to Appendix 7.B. The fact that there are no collisions for almost all $t \in [0,T]$ is ensured by [55, Lemma 6.2] following nearly the same computations as in [53, Theorem 5.2] (see also [55, Theorem 6.3] for its adaptation to the 2×2 block structure). The only difference in our case compared to the proof of [53, Theorem 5.2] is that the martingales $\mathrm{d}M_i(t)$ (cf. [53, Eq. (5.4)]) are defined as

$$\mathrm{d}M_i(t) := \frac{\mathrm{d}b_i^z(t)}{\sqrt{n}}, \qquad |i| \le n, \tag{7.223}$$

with $\{b_i^z\}_{i\in[n]}$ having non trivial covariance (7.135). This fact does not play any role in that proof, since the only information about $\mathrm{d}\boldsymbol{M} = \{\mathrm{d}M_i\}_{|i|\le n}$ used in [53, Theorem 5.2] is that it has bounded quadratic variation and that $\boldsymbol{M}(t)$ is γ-Hölder continuous for any $\gamma \in (0,1/2)$, which is clearly the case for $\mathrm{d}\boldsymbol{M}$ defined in (7.223). □

7.A The interpolation process is well defined

We recall that the eigenvectors of H^z are of the form $\boldsymbol{w}_{\pm i}^z = (\boldsymbol{u}_i^z, \pm \boldsymbol{v}_i^z)$ for any $i \in [n]$, as a consequence of the symmetry of the spectrum of H^z with respect to zero. Consider the matrix flow

$$\mathrm{d}X_t = \frac{\mathrm{d}B_t}{\sqrt{n}}, \qquad X_0 = X, \tag{7.224}$$

with B_t being a standard real matrix valued Brownian motion. Let H_t^z denote the Hermitisation of $X_t - z$, and $\{\boldsymbol{w}_i^z(t)\}_{|i|\le n}$ its eigenvectors. We recall that the eigenvectors $\{\boldsymbol{w}_i^z(t)\}_{|i|\le n}$ are almost surely well defined, since H_t^z does not have multiple eigenvalues almost surely by (7.136). We set the eigenvectors equal to zero where they are not well defined. Recall the definitions of the coefficients $\Lambda_{ij}^z(t)$, $\mathring{\Lambda}_{ij}^z(t)$ from (7.129), (7.132) and (7.140), respectively. Set

$$\Delta_n := \Big\{ (x_i)_{|i|\le n} \in \mathbf{R}^{2n} \,\Big|\, 0 < x_1 < \dots < x_n,\, x_{-i} = -x_i,\, \forall i \in [n] \Big\},$$

and let $C(\mathbf{R}_+, \Delta_n)$ be the space of continuous functions $f : \mathbf{R}_+ \to \Delta_n$. Let $\omega_E > 0$ be the exponent in (7.140), and let $\omega_r > 0$ be such that $\omega_r \ll \omega_E$. In this appendix we prove that for any $\alpha \in [0,1]$ the system of SDEs

$$\mathrm{d}x_i^z(t,\alpha) = \frac{\mathrm{d}\mathring{b}_i^z(t)}{\sqrt{n(1+n^{-\omega_r})}} + \frac{1}{2n}\sum_{j\ne i} \frac{1+\alpha\mathring{\Lambda}_{ij}^z(t)}{x_i^z(t,\alpha) - x_j^z(t,\alpha)}\,\mathrm{d}t, \quad x_i^z(0,\alpha) = x_i(0), \tag{7.225}$$

for $|i| \leq n$, with $\boldsymbol{x}(0) \in \Delta_n$, admits a strong solution for any $t \geq 0$. For $T > 0$, by (7.139), the martingales $\{\mathring{b}_i^z\}_{|i|\leq[n]}$, defined on a filtration $(\widetilde{\mathcal{F}}_{b,t})_{0\leq t\leq T}$, are such that $\mathring{b}_{-i}^z = -\mathring{b}_i^z$ for $i \in [n]$, and that

$$\mathbf{E}\Big[\mathrm{d}\mathring{b}_i^z \,\mathrm{d}\mathring{b}_j^z \,\Big|\, \widetilde{\mathcal{F}}_{b,t}\Big] = \frac{\delta_{i,j} - \delta_{i,-j} + \mathring{\Lambda}_{ij}^z(t)}{2}\,\mathrm{d}t, \qquad |i|,|j| \leq n. \tag{7.226}$$

The main result of this section is Proposition 7.A.1 below. Its proof follows closely [53, Proposition 5.4], which is inspired by the proof of [17, Lemma 4.3.3]. We nevertheless present the proof of Proposition 7.A.1 for completeness, explaining the differences compared with [53, Proposition 5.4] as a consequence of the correlation in (7.226).

Proposition 7.A.1. *Fix any $z \in \mathbf{C}$, and let $\boldsymbol{x}(0) \in \Delta_n$. Then for any fixed $\alpha \in [0,1]$ there exists a unique strong solution $\boldsymbol{x}(t,\alpha) = \boldsymbol{x}^z(t,\alpha) \in C(\mathbf{R}_+, \Delta_n)$ to the system of SDE (7.225) with initial condition $\boldsymbol{x}(0)$.*

We will mostly omit the z-dependence since the analysis of (7.225) is done for any fixed $z \in \mathbf{C}$; in particular, we will use the notation $\mathring{\Lambda}_{ij} = \mathring{\Lambda}_{ij}^z$. By (7.129), (7.132) and (7.140) it follows that $\mathring{\Lambda}_{ij}(t) = \mathring{\Lambda}_{ji}(t)$, and that $|\mathring{\Lambda}_{ij}(t)| \leq n^{-\omega_E}$, for any $t \geq 0$.

Proof. We follow the notations used in the proof of [53, Proposition 5.4] to make the comparison clearer. Moreover, we do not keep track of the n-dependence of the constants, since throughout the proof n is fixed. By a simple time rescaling, we rewrite the process (7.225) as

$$\mathrm{d}x_i(t,\alpha) = \mathrm{d}\mathring{b}_i(t) + \frac{1}{2}\sum_{j\neq i} \frac{1+\theta_{ij}(t)}{x_i(t,\alpha) - x_j(t,\alpha)}\,\mathrm{d}t, \qquad |i| \leq n, \tag{7.227}$$

where $\theta_{ij}(t) := \alpha\mathring{\Lambda}_{ij}(1+n^{-\omega_r}) + n^{-\omega_r}$ is such that $\theta_{ij}(t) = \theta_{ji}(t)$. Note that $c_1 \leq \theta_{ij}(t) \leq c_2$ for any $t \geq 0$ and $\alpha \in [0,1]$, with $c_1 = n^{-\omega_r}/2$, $c_2 = 1$. For any $\epsilon > 0$ define the bounded Lipschitz function $\phi_\epsilon \colon \mathbf{R} \to \mathbf{R}$ as

$$\phi_\epsilon(x) := \begin{cases} x^{-1}, & |x| \geq \epsilon, \\ \epsilon^{-2}x, & |x| < \epsilon, \end{cases}$$

that cuts off the singularity of x^{-1} at zero.

Introduce the system of cut-off SDEs

$$\mathrm{d}x_i^\epsilon(t,\alpha) = \mathrm{d}\mathring{b}_i(t) + \frac{1}{2}\sum_{j\neq i}(1+\theta_{ij}(t))\phi_\epsilon(x_i^\epsilon(t,\alpha) - x_j^\epsilon(t,\alpha))\,\mathrm{d}t, \quad |i| \leq n, \tag{7.228}$$

which admits a unique strong solution (see e.g. [122, Theorem 2.9 of Section 5]) as a consequence of ϕ_ϵ being Lipschitz and the fact that $\mathrm{d}\mathring{\boldsymbol{b}} = (\mathring{C})^{1/2}\,\mathrm{d}\mathfrak{w}$ (see (7.142)). Define the stopping times

$$\tau_\epsilon = \tau_\epsilon(\alpha) := \inf\left\{ t \;\middle|\; \min_{|i|,|j|\leq n} \left| x_i^\epsilon(t,\alpha) - x_j^\epsilon(t,\alpha)\right| \leq \epsilon \quad \text{or} \quad \|\boldsymbol{x}^\epsilon(t,\alpha)\|_\infty \geq \epsilon^{-1} \right\}. \tag{7.229}$$

By strong uniqueness we have that $\boldsymbol{x}^{\epsilon_2}(t,\alpha) = \boldsymbol{x}^{\epsilon_1}(t,\alpha)$ for any $t \in [0,\tau_{\epsilon_2}]$ if $0 < \epsilon_1 < \epsilon_2$. Note that $\tau_{\epsilon_2} \leq \tau_{\epsilon_1}$ for $\epsilon_1 < \epsilon_2$, thus the limit $\tau = \tau(\alpha) := \lim_{\epsilon\to 0}\tau_\epsilon(\alpha)$ exists, and $\boldsymbol{x}(t,\alpha) := \lim_{\epsilon\to 0} \boldsymbol{x}^\epsilon(t,\alpha)$ defines a strong solution to (7.227) on $[0,\tau)$. Moreover, by continuity in time, $\boldsymbol{x}(t,\alpha)$ remains ordered as $0 < x_1(t,\alpha) < \cdots < x_n(t,\alpha)$ and $x_{-i}(t,\alpha) = -x_i(t,\alpha)$ for $i \in [n]$. Additionally, for the square of the ℓ^2-norm $\|\boldsymbol{x}\|_2^2 = \sum_i x_i^2$ a simple calculation shows that

$$\mathrm{d}\|\boldsymbol{x}(t,\alpha)\|_2^2 = \frac{1}{2}\left(\sum_{j\neq i}(1+\theta_{ij}) + \sum_{|i|,|j|\leq n} \mathring{\Lambda}_{ij}\right)\mathrm{d}t + \mathrm{d}M_1, \tag{7.230}$$

with $\mathrm{d}M_1$ being a martingale term. This implies that $\mathbf{E}\|\boldsymbol{x}(t\wedge s)\|_2^2 \leq c(1+t)$ for any stopping time $s < \tau$ and for any $t \geq 0$, where c depends on n.

Let $a > 0$ be a large constant that we will choose later in the proof, and define a_k recursively by $a_0 := a$, $a_{k+1} := a_k^5$ for $k \geq 0$. Consider the Lyapunov function

$$f(\boldsymbol{x}) := -2\sum_{k\neq l} a_{|k-l|} \log|x_k - x_l|. \tag{7.231}$$

Then by Itô's formula we get

$$\mathrm{d}f(\boldsymbol{x}) = A(\boldsymbol{x}(t,\alpha))\,\mathrm{d}t + \mathrm{d}M_2(t), \tag{7.232}$$

with

$$\begin{aligned} A(\boldsymbol{x}(t,\alpha)) :=\; & -2\sum_{l\neq i, j\neq i} \frac{(1+\theta_{ij})a_{|i-l|}}{(x_i(t,\alpha)-x_l(t,\alpha))(x_i(t,\alpha)-x_j(t,\alpha))} + \sum_{|i|\leq n}\frac{a_{|2i|}}{(2x_i(t,\alpha))^2} \\ & + \sum_{j\neq i}\frac{a_{|i-j|}(1+\mathring{\Lambda}_{ii}(t)-\mathring{\Lambda}_{ij}(t))}{(x_i(t,\alpha)-x_j(t,\alpha))^2}, \end{aligned} \tag{7.233}$$

where $\mathrm{d}M_2$ is a martingale given by

$$\mathrm{d}M_2(t) = -2\sum_{j\neq i}\frac{a_{|i-j|}\,\mathrm{d}\mathring{b}_i(t)}{x_i(t,\alpha)-x_j(t,\alpha)}.$$

In the following we will often omit the time dependence. Note that the term in (7.233) containing $\mathring{\Lambda}_{ii} - \mathring{\Lambda}_{ij}$ is new compared to [53, Eq. (5.39)], since it comes from the correlation of the martingales $\{\mathring{b}_i\}_{|i|\leq n}$, whilst in [53, Eq. (5.39)] i.i.d. Brownian motions have been considered. In the remainder of the proof we show that the term $\mathring{\Lambda}_{ii} - \mathring{\Lambda}_{ij}$ is negligible using the fact that $|\mathring{\Lambda}_{ij}| \leq n^{-\omega_E}$, and so that this term can be absorbed in the negative term coming from the first sum in the r.h.s. of (7.233) for $l = j$.

We now prove that $A(\boldsymbol{x}(t,\alpha)) \leq 0$ if $a > 0$ is sufficiently large. Firstly, we write $A(\boldsymbol{x}(t,\alpha))$ as

$$\begin{aligned} A(\boldsymbol{x}(t,\alpha)) = & -2\sum_{\substack{l\neq i, j\neq i\\ j\neq l}}\frac{(1+\theta_{ij})a_{|i-l|}}{(x_i-x_l)(x_i-x_j)} - \sum_{j\neq\pm i}\frac{a_{|i-j|}(1+2\theta_{ij}-\mathring{\Lambda}_{ii}+\mathring{\Lambda}_{ij})}{(x_i-x_j)^2} \\ & -2\sum_{|i|\leq n}\frac{a_{|2i|}(\theta_{-i,i}-\mathring{\Lambda}_{ii})}{(2x_i)^2}. \end{aligned} \tag{7.234}$$

Then, using that the first sum in (7.234) is non-positive for $(i-l)(i-j) > 0$, and that $c_1 \le \theta_{ij} \le c_2$, with $c_1 = n^{-\omega_r}$, we bound $A(\boldsymbol{x}(t,\alpha))$ as follows

$$\begin{aligned} A(\boldsymbol{x}(t,\alpha)) \le -2(1+c_2) \sum_{|i|\le n} \sum_{(i-l)(i-j)<0} \frac{a_{|i-l|}}{(x_i-x_l)(x_i-x_j)} - c_1 \sum_{j\neq i} \frac{a_{|i-j|}}{(x_i-x_j)^2} \\ - \sum_{j\neq \pm i} \frac{a_{|i-j|}}{(x_i-x_j)^2}. \end{aligned} \tag{7.235}$$

In (7.235) we used that

$$\theta_{ij} - \mathring{\Lambda}_{ii} + \mathring{\Lambda}_{ij} \ge \frac{c_1}{2}, \qquad \theta_{-i,i} - \mathring{\Lambda}_{ii} \ge \frac{c_1}{2},$$

since $\theta_{ij} \ge c_1 = n^{-\omega_r}$ and $|\mathring{\Lambda}_{ij}| \le n^{-\omega_E}$, where $\omega_r \ll \omega_E$. This shows that the correlations of the martingales $\{\mathring{b}_i\}_{|i|\le n}$ is negligible. Note that the r.h.s. of (7.235) has exactly the same form as [53, Eq. (5.42)], since the third term in (7.235) is non-positive. Hence, following exactly the same computations as in [53, Eqs. (5.43)–(5.46)], choosing $a > n^{10}$, we conclude that

$$A(\boldsymbol{x}(t,\alpha)) \le \left[\frac{2(1+c_2)}{a} - c_1\right] \sum_{j\neq i} \frac{a_{|i-j|}}{(x_i-x_j)^2}, \tag{7.236}$$

which is negative for a sufficiently large.

Fix $a > 0$ large enough so that $A(\boldsymbol{x}(t,\alpha)) \le 0$, then for any stopping time $s < \tau$, and any $t \ge 0$ we have

$$\mathbf{E}[f(\boldsymbol{x}(t\wedge s,\alpha))] \le \mathbf{E}[f(\boldsymbol{x}(0,\alpha))]. \tag{7.237}$$

Hence, by [53, Eqs. (5.48)–(5.49)], using that $\mathbf{E}\|\boldsymbol{x}(t\wedge\tau_\epsilon)\|_2^2 \le c(1+t)$, it follows that

$$\log(\epsilon^{-1})\,\mathbf{P}(\tau_\epsilon < t) \le c,$$

and so that $\mathbf{P}(\tau < t) = 0$, letting $\epsilon \to 0$. Since $t \ge 0$ is arbitrary, this implies that $\mathbf{P}(\tau < +\infty) = 0$, i.e. (7.227) has a unique strong solution on $(0,\infty)$ such that $\boldsymbol{x}(t,\alpha) \in \Delta_n$ for any $t \ge 0$ and $\alpha \in [0,1]$. □

Additionally, by a similar argument as in [53, Proposition 5.5], we conclude the following lemma.

Lemma 7.A.2. *Let $\boldsymbol{x}(t,\alpha)$ be the unique strong solution of (7.225) with initial data $\boldsymbol{x}(0,\alpha) \in \Delta_n$, for any $\alpha \in [0,1]$, and assume that there exists $L > 0$ such that $\|\boldsymbol{x}(0,\alpha_1) - \boldsymbol{x}(0,\alpha_2)\|_2 \le L|\alpha_1-\alpha_2|$, for any $\alpha_1,\alpha_2 \in [0,1]$. Then $\boldsymbol{x}(t,\alpha)$ is Lipschitz in $\alpha \in [0,1]$ for any $t \ge 0$ on an event Ω such that $\mathbf{P}(\Omega) = 1$, and its derivative satisfies*

$$\begin{aligned} \partial_\alpha x_i(t,\alpha) = \partial_\alpha x_i(0,\alpha) + \frac{1}{2n}\int_0^t \sum_{j\neq i} \frac{[1+\alpha\mathring{\Lambda}_{ij}(s)][\partial_\alpha x_j(s,\alpha) - \partial_\alpha x_i(s,\alpha)]}{(x_i(s,\alpha)-x_j(s,\alpha))^2}\,\mathrm{d}s \\ + \frac{1}{2n}\int_0^t \sum_{j\neq i} \frac{\mathring{\Lambda}_{ij}(s)}{x_i(s,\alpha)-x_j(s,\alpha)}\,\mathrm{d}s. \end{aligned} \tag{7.238}$$

7.B Derivation of the DBM for singular values in the real case

Let X be an $n \times n$ real random matrix, and define $Y^z := X - z$. Consider the matrix flow (7.224) defined on a probability space Ω equipped with a filtration $(\mathcal{F}_t)_{0 \le t \le T}$, and denote by H_t^z the Hermitisation of $X_t - z$. We now derive (7.133), under the assumption that the eigenvalues are all distinct. This derivation is easily made complete by the argument in the proof Proposition 7.7.6 in Section 7.7.6.

Let $\{\lambda_i^z(t), -\lambda_i^z(t)\}_{i\in[n]}$ be the eigenvalues of H_t^z, and denote by $\{\boldsymbol{w}_i^z(t), \boldsymbol{w}_{-i}^z(t)\}_{i\in[n]}$ their corresponding orthonormal eigenvectors, i.e. for any $i, j \in [n]$, omitting the t-dependence, we have that

$$H^z \boldsymbol{w}_{\pm i}^z = \pm\lambda_i^z \boldsymbol{w}_{\pm i}^z, \qquad (\boldsymbol{w}_i^z)^* \boldsymbol{w}_j^z = \delta_{ij}, \qquad (\boldsymbol{w}_i^z)^* \boldsymbol{w}_{-j}^z = 0. \tag{7.239}$$

In particular, for any $i \in [n]$, by the block structure of H^z it follows that

$$\boldsymbol{w}_{\pm i}^z = (\boldsymbol{u}_i^z, \pm \boldsymbol{v}_i^z), \qquad Y^z \boldsymbol{v}_i^z = \lambda_i^z \boldsymbol{u}_i^z, \qquad (Y^z)^* \boldsymbol{u}_i^z = \lambda_i^z \boldsymbol{v}_i^z. \tag{7.240}$$

Moreover, since $\{\boldsymbol{w}_{\pm i}^z\}_{i=1}^n$ is an orthonormal basis, we conclude that

$$(\boldsymbol{u}_i^z)^* \boldsymbol{u}_i^z = (\boldsymbol{v}_i^z)^* \boldsymbol{v}_i^z = \frac{1}{2}. \tag{7.241}$$

In the following, for any fixed entry x_{ab} of X, we denote the derivative in the x_{ab} direction by

$$\dot{f} := \frac{\partial f}{\partial x_{ab}}, \tag{7.242}$$

where $f = f(X)$ is a function of the matrix X. From now on we only consider positive indices $1 \le i \le n$. We may also drop the z and t dependence to make our notation lighter. For any $i, j \in [n]$, differentiating (7.239) we obtain

$$\dot{H}\boldsymbol{w}_i + H\dot{\boldsymbol{w}}_i = \dot{\lambda}_i \boldsymbol{w}_i + \lambda_i \dot{\boldsymbol{w}}_i, \tag{7.243}$$

$$\dot{\boldsymbol{w}}_i^* \boldsymbol{w}_j + \boldsymbol{w}_i^* \dot{\boldsymbol{w}}_j = 0, \tag{7.244}$$

$$\boldsymbol{w}_i^* \dot{\boldsymbol{w}}_i + \dot{\boldsymbol{w}}_i^* \boldsymbol{w}_i = 0. \tag{7.245}$$

Note that (7.245) implies that $\Re[\boldsymbol{w}_i^* \dot{\boldsymbol{w}}_i] = 0$. Moreover, since the eigenvectors are defined modulo a phase, we can choose eigenvectors such that $\Im[\boldsymbol{w}_i^* \dot{\boldsymbol{w}}_i] = 0$ for any $t \ge 0$ hence $\boldsymbol{w}_i^* \dot{\boldsymbol{w}}_i = 0$. Then, multiplying (7.243) by $\boldsymbol{w}_i^*$ we conclude that

$$\dot{\lambda}_i = \boldsymbol{u}_i^* \dot{Y} \boldsymbol{v}_i + \boldsymbol{v}_i^* \dot{Y}^* \boldsymbol{u}_i. \tag{7.246}$$

Moreover, multiplying (7.243) by $\boldsymbol{w}_j^*$, with $j \ne i$, and by $\boldsymbol{w}_{-j}^*$, we get

$$(\lambda_i - \lambda_j)\boldsymbol{w}_j^* \dot{\boldsymbol{w}}_i = \boldsymbol{w}_j^* \dot{H} \boldsymbol{w}_i, \qquad (\lambda_i + \lambda_j)\boldsymbol{w}_{-j}^* \dot{\boldsymbol{w}}_i = \boldsymbol{w}_{-j}^* \dot{H} \boldsymbol{w}_i, \tag{7.247}$$

respectively. By (7.245) and $\boldsymbol{w}_i^* \dot{\boldsymbol{w}}_i = 0$ it follows that

$$\dot{\boldsymbol{w}}_i = \sum_{\substack{j\in[n],\\ j\ne i}} (\boldsymbol{w}_j^* \dot{\boldsymbol{w}}_i)\boldsymbol{w}_j + \sum_{j\in[n]} (\boldsymbol{w}_{-j}^* \dot{\boldsymbol{w}}_i)\boldsymbol{w}_{-j}, \tag{7.248}$$

hence, by (7.247), we conclude

$$\dot{\boldsymbol{w}}_i = \sum_{j\neq i} \frac{\boldsymbol{v}_j^* \dot{Y}^* \boldsymbol{u}_i + \boldsymbol{u}_j^* \dot{Y} \boldsymbol{v}_i}{\lambda_i - \lambda_j} \boldsymbol{w}_j + \sum_j \frac{\boldsymbol{u}_j^* \dot{Y} \boldsymbol{v}_i - \boldsymbol{v}_j^* \dot{Y}^* \boldsymbol{u}_i}{\lambda_i + \lambda_j} \boldsymbol{w}_{-j}. \tag{7.249}$$

Throughout this appendix we use the convention that for any vectors $\boldsymbol{v} \in \mathbf{C}^n$ we denote its entries by $v(a)$, with $a \in [n]$. By (7.246)–(7.249) it follows that

$$\frac{\partial \lambda_i}{\partial x_{ab}} = 2\Re[u_i^*(a) v_i(b)], \tag{7.250}$$

and that

$$\frac{\partial w_i}{\partial x_{ab}}(k) = \sum_{j\neq i} \left[\frac{u_j^*(a)v_i(b) + v_j^*(b)u_i(a)}{\lambda_i - \lambda_j} w_j(k) + \frac{u_j^*(a)v_i(b) - v_j^*(b)u_i(a)}{\lambda_i + \lambda_j} w_{-j}(k) \right] \\ + \frac{u_i^*(a)v_i(b) - v_i^*(b)u_i(a)}{2\lambda_i} w_{-i}(k).$$

By Ito's formula we have that

$$\mathrm{d}\lambda_i = \sum_{ab} \frac{\partial \lambda_i}{\partial x_{ab}} \,\mathrm{d}x_{ab} + \frac{1}{2} \sum_{ab} \sum_{kl} \frac{\partial^2 \lambda_i}{\partial x_{ab} \partial x_{kl}} \,\mathrm{d}x_{ab} \,\mathrm{d}x_{kl}. \tag{7.251}$$

Then we compute

$$\begin{aligned} \frac{\partial^2 \lambda_i}{\partial x_{ab} \partial x_{kl}} &= 2\Re \left[\frac{\partial v_i^*}{\partial x_{ab}}(l) u_i(k) + v_i^*(l) \frac{\partial u_i}{\partial x_{ab}}(k) \right] \\ &= 2\Re \Bigg[\sum_{j\neq i} \left[\frac{u_j(a)v_i^*(b) + v_j(b)u_i^*(a)}{\lambda_i - \lambda_j} v_j^*(l)u_i(k) - \frac{u_j(a)v_i^*(b) - v_j(b)u_i^*(a)}{\lambda_i + \lambda_j} v_j^*(l) u_i(k) \right] \\ &\quad - \frac{u_i(a)v_i^*(b) - v_i(b)u_i^*(a)}{2\lambda_i} v_i^*(l)u_i(k) + \frac{u_i^*(a)v_i(b) - v_i^*(b)u_i(a)}{2\lambda_i} u_i(k)v_i^*(l) \\ &\quad + \sum_{j\neq i} \left[\frac{u_j^*(a)v_i(b) + v_j^*(b)u_i(a)}{\lambda_i - \lambda_j} u_j(k)v_i^*(l) + \frac{u_j^*(a)v_i(b) - v_j^*(b)u_i(a)}{\lambda_i + \lambda_j} u_j(k)v_i^*(l) \right] \Bigg]. \end{aligned} \tag{7.252}$$

Hence, combining (7.250)–(7.252), we finally conclude that

$$\begin{aligned} \mathrm{d}\lambda_i^z = \frac{\mathrm{d}b_i^z}{\sqrt{n}} &+ \frac{1}{2n} \sum_{j\neq i} \left[\frac{1 + 4\Re[\langle \overline{u_j^z}, u_i^z \rangle \langle v_i^z, \overline{v_j^z} \rangle]}{\lambda_i^z - \lambda_j^z} + \frac{1 + 4\Re[\langle \overline{u_j^z}, u_i^z \rangle \langle v_i^z, -\overline{v_j^z} \rangle]}{\lambda_i^z + \lambda_j^z} \right] \mathrm{d}t \\ &+ \frac{1 + 4\Re[\langle \overline{u_i^z}, u_i^z \rangle \langle v_i^z, -\overline{v_i^z} \rangle]}{4n\lambda_i^z} \,\mathrm{d}t. \end{aligned} \tag{7.253}$$

In (7.253) we used the convention that for any vector $\boldsymbol{v} \in \mathbf{C}^n$ by $\overline{\boldsymbol{v}}$ we denote the vector with entries $\overline{v}(a) = \overline{v(a)}$, for any $a \in [n]$. The driving martingales in (7.253) are defined as

$$\mathrm{d}b_i^z := \mathrm{d}B_{ii}^z + \mathrm{d}\overline{B_{ii}^z}, \quad \text{with} \quad \mathrm{d}B_{ij}^z := \sum_{ab} (u_i^z)^*(a) \,\mathrm{d}B_{ab} v_j^z(b), \tag{7.254}$$

with $B = B_t$ the matrix valued Brownian motion in (7.224), and their covariance given by

$$\mathbf{E}\Big[\mathrm{d}b_i^z\,\mathrm{d}b_j^z\,\Big|\,\mathcal{F}_t\Big] = \frac{\delta_{ij} + 4\Re\left[\langle\overline{u_j^z}, u_i^z\rangle\langle v_i^z, \overline{v_j^z}\rangle\right]}{2}\,\mathrm{d}t. \tag{7.255}$$

Note that $\{b_i^z\}_{i\in[n]}$ defined in (7.254) are not Brownian motions, as a consequence of the non deterministic quadratic variation (7.255).

Optimal lower bound on the least singular value of the shifted Ginibre ensemble 8

We consider the least singular value of a large random matrix with real or complex i.i.d. Gaussian entries shifted by a constant $z \in \mathbf{C}$. We prove an optimal lower tail estimate on this singular value in the critical regime where z is around the spectral edge thus improving the classical bound of Sankar, Spielman and Teng [168] for the particular shift-perturbation in the edge regime. Lacking Brézin-Hikami formulas in the real case, we rely on the superbosonization formula [142].

Published as G. Cipolloni et al., *Optimal lower bound on the least singular value of the shifted ginibre ensemble*, Accepted to Probability and Mathemtical Physics (2020), arXiv: 1908.01653

8.1 Introduction

The effective numerical solvability of a large system of linear equations $Ax = b$ is determined by the condition number of the matrix A. In many practical applications the norm of A is bounded and thus the condition number critically depends on the smallest singular value $\sigma_1(A)$ of A. When the matrix elements of A come from noisy measured data, then the lower tail probability of $\sigma_1(A)$ tends to exhibit a universal scaling behavior, depending on the variance of the noise. In the simplest case A can be decomposed as

$$A = A_0 + X, \tag{8.1}$$

where A_0 is a deterministic square matrix and X is drawn from the *Ginibre ensemble*, i.e. X has i.i.d. centred Gaussian matrix elements with variance $\mathbf{E}|x_{ij}|^2 = N^{-1}$, where N is the dimension.

The randomness in X smoothens out possible singular behavior of A^{-1}. In particular Sankar, Spielman and Teng [168] showed that the smallest singular value $\sigma_1(A)$, lives on a scale not smaller than N^{-1}, equivalently, the smallest eigenvalue $\lambda_1(AA^*)$ of AA^* lives on

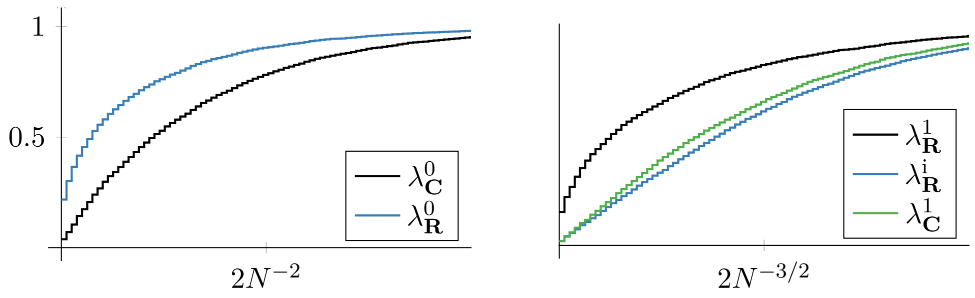


Figure 8.1: Plots of the cumulative histograms of the smallest eigenvalue $\lambda_{\mathbf{R},\mathbf{C}}^{z}$ of the matrix $(X-z)(X-z)^*$, where $\mathbf{R},\mathbf{C}$ indicates whether X is distributed according to the real or complex Ginibre ensemble. The data was generated by sampling 5000 matrices of size 200×200. The first plot confirms the difference between the x- and $\sqrt{x}$-scaling close to 0, see (8.3). The second plot shows that this difference is also observable for shifted Ginibre matrices at the edge $|z|=1$, but only for real spectral parameters $z=\pm 1$. When the complex parameter z is away from the real axis, then the real case behaves similarly to the complex case.

a scale $\le N^{-2}$, i.e.

$$\mathbf{P}(\lambda_1(AA^*)=[\sigma_1(A)]^2\le xN^{-2})\lesssim\sqrt{x},\quad \text{for any}\quad x>0, \tag{8.2}$$

up to logarithmic corrections, uniformly in A_0. If X is a complex Ginibre matrix, then the $\sqrt{x}$ bound improves to x.

The special case $A_0=0$ shows that the bound (8.2) is essentially optimal. Indeed, the tail probability of $\lambda_1(XX^*)$ of real and complex Ginibre ensembles has been explicitly computed by Edelman [75] as

$$\lim_{N\to\infty}\mathbf{P}(\lambda_1(XX^*)\le xN^{-2})=\begin{cases}1-e^{-x/2-\sqrt{x}}=\sqrt{x}+\mathcal{O}(x), & \text{in the real case}\\ 1-e^{-x}=x+\mathcal{O}(x^2), & \text{in the complex case.}\end{cases} \tag{8.3}$$

The complex Ginibre ensemble has a stronger smoothing effect in (8.3) is due to the additional degrees of freedom. This observation is analogous to the different strength of the level repulsion in real symmetric and complex Hermitian random matrices.

The support of the spectrum of such *information plus noise matrices* AA^* becomes deterministic as $N\to\infty$ and it can be computed from the solution of a certain self-consistent equation [71]. Almost surely no eigenvalues lie outside the support of the limiting measure [21]. Thus $\lambda_1(AA^*)$ has a simple N-independent positive lower bound if 0 is away from this support. However, when 0 is well inside the limiting spectrum, the smoothing mechanism becomes important yielding that $\lambda_1(AA^*)$ is of order N^{-2} with a lower tail given in (8.2). The regime where 0 is near the edge of this support is yet unexplored.

The goal of this paper is to study this transitional regime for $A=X-z$, i.e. for the important special case where $A_0=-zI$ is a constant multiple of the idenity matrix, as the spectral parameter $z\in\mathbf{C}$ is varied. The limiting density of states of $Y^z:=(X-z)(X-z)^*$ is supported in the interval $[0,\mathfrak{e}_+]$ for $|z|\le 1$ and the interval $[\mathfrak{e}_-,\mathfrak{e}_+]$ with $\mathfrak{e}_->0$ for $|z|>1$, where $\mathfrak{e}_\pm$ are explicit functions of $|z|$ given in (8.18a). As noted above, the problem is relatively simple if $|z|\ge 1+\epsilon$ with some N-independent ϵ as in this case [21] implies that almost surely $\lambda_1(Y^z) \geq C(\epsilon) > 0$ is bounded away from zero. In the opposite regime, when $|z| \leq 1 - \epsilon$, then typically $\lambda_1(Y^z) \sim N^{-2}$, and in fact (8.2) provides the correct corresponding upper bound (modulo logs).

Our main result on the tail probability of $\lambda_1(Y^z)$ is that for $|z| \leq 1 + CN^{-1/2}$

$$\mathbf{P}\Big(\lambda_1(Y^z) \leq x \cdot c(N,z)\Big) \lesssim \begin{cases} x + \sqrt{x}e^{-\frac{1}{2}N(\Im z)^2}, & \text{in the real case} \\ x, & \text{in the complex case,} \end{cases} \tag{8.4a}$$

where

$$c(N,z) := \min\Big\{\frac{1}{N^{3/2}}, \frac{1}{N^2|1-|z|^2|}\Big\}. \tag{8.5}$$

Our bound is sharp up to logarithmic corrections, see Corollary 8.2.4 for the precise statement. Notice the transition between the x and $\sqrt{x}$ behaviour in the real case of (8.4a): near the real axis, $|\Im z| \ll N^{-1/2}$, the result is analogous to the real case (8.3) at $z = 0$, otherwise the complex behaviour (8.3) dominates at the edges even for real X, see Fig. 8.1. These results reveal how the robust bound (8.2) improves near the spectral edge in the transition regime $-CN^{-1/2} \leq 1 - |z| \ll 1$ in both symmetry classes. The transition to the Tracy-Widom scaling in the regime well outside of the spectrum $|z| - 1 \gg N^{-1/2}$ is deferred to our future work.

One motivation for studying $X - z$ is the classical ODE model $\mathrm{d}u/\,\mathrm{d}t = (X-z)u$ on the stability of large biological networks by May [145]. For example, the matrix elements x_{ij} may express random connectivity rates between neurons and z is the overall decay rate of neuron activation [184]. As $\Re z$ crosses 1, there is a fine phase transition in the large time behavior of u that depends on whether X is real or complex Ginibre matrix, see [52] and [82] for the recent mathematical results, as well as for further references. Another important motivation is that an effective lower tail bound on the least singular value of $X - z$ is essential for the proof of the circular law via Girko's formula, see [34] for a detailed survey. In fact, this is the most delicate ingredient in any proof concerning eigenvalue distribution of large non-Hermitian matrices. In particular, relying on the main result of the current paper, we proved [59] that the local eigenvalue statistics for random matrices with centered i.i.d. entries near the spectral edge asymptotically coincide with those for the corresponding Ginibre ensemble as $N \to \infty$. This is the non-Hermitian analogue of the celebrated Tracy-Widom edge universality for Wigner matrices [41, 186]. Similarly, the singular value bound from the present paper is also an important ingredient for the recent CLTs for complex and real i.i.d. matrices [58, 60].

We now give a brief history of related results. In the $z = 0$ case tail estimates for $\lambda_1(XX^*)$ beyond the Gaussian distribution have been subject of intensive research [165, 199] eventually obtaining (8.3) with an additive $\mathcal{O}(e^{-cN})$ error term for any X with i.i.d. entries with subgaussian tails in [167]. The precise distribution of $\lambda_1(XX^*)$ was shown in [192] to coincide with the Gaussian case (8.3) under a bounded high moment condition and with an $\mathcal{O}(N^{-c})$ error term, see also [54, 55] for more general ensembles. In the case of general A_0 lower bounds on $\lambda_1(AA^*)$ in the non-Gaussian setting have been obtained in [196, 197], albeit not uniformly in A_0, see also [63, 201] beyond the i.i.d. case. We are not aware of any previous results improving (8.2) in the transitional regime (8.4a).

Since we consider Ginibre (i.e. purely Gaussian) ensembles, one might think that everything is explicitly computable from the well understood spectrum of X. The eigenvalue density of X converges to the uniform distribution on the unit disk and the spectral radius of X converges to 1 (these results have also been established for the general non-Gaussian case, cf. Girko's circular law [18, 20, 33, 101, 103, 191]). Also the joint probability density function of all Ginibre eigenvalues, as well as their local correlation functions are explicitly known; see [102] and [146] for the relatively simple complex case, and [35, 76, 97, 137] for the more involved real case, where the appearance of $\sim N^{1/2}$ real eigenvalues causes a singularity in the local density. However, eigenvalues of X give no direct information on the singular values of $X - z$ and the extensive literature on the Ginibre spectrum is not applicable. Notice that any intuition based upon the eigenvalues of X is misleading: the nearest eigenvalue to z is at a distance of order $N^{-1/2}$ for any $|z| \leq 1$. However, $\|(X-z)\|^{-1} \sim \max\{N^{3/4}, N|1-|z|^2|^{1/2}\}$ for $|z| \leq 1 + CN^{-1/2}$, as a consequence of our result (8.4a). This is an indication that typically X is highly non-normal (another indication is that the largest singular value of X is 2, while its spectral radius is only 1).

Regarding our strategy, in this paper we use supersymmetric methods to express the resolvent of Y^z. In particular, we use a multiple Grassmann integral formula for

$$\varrho_N^z(E) := \frac{1}{N\pi} \Im \mathbf{E} \operatorname{Tr} \frac{1}{Y^z - E + \mathrm{i}0}, \tag{8.6}$$

the averaged density of states (or one-point function) of Y^z at energy $E \in \mathbf{R}$. For $|E| \lesssim c(N,z)$ a sizeable contribution to (8.6) comes from the lowest eigenvalue $\lambda_1(Y^z)$, hence a good upper estimate on (8.6) translates into a lower tail bound on $\lambda_1(Y^z)$.

With the help of the superbosonization formula by Littelmann, Sommers and Zirnbauer [142], we can drastically reduce the number of integration variables: instead of N bosonic and N fermionic variables we will have an explicit expression for (8.6) involving merely two contour integration variables in complex case and three in the real case. The remaining integrals are still highly oscillatory, but contour deformation allows us to estimate them optimally. In fact, saddle point analysis identifies the leading term as long as $|E| \gg c(N,z)$. However, in the critical regime, $|E| \lesssim c(N,z)$, the saddle point analysis breaks down. The leading term is extracted as a specific rescaling of a universal function given by a double integral. We work out the precise answer for (8.6) in the complex case and we provide optimal bounds in the real case, deferring the precise asymptotics to further work.

Lower tail estimates require delicate knowledge about individual eigenvalues, i.e. about the density of states below the scale of eigenvalue spacing, and it is crucial to exploit the Gaussianity of X via explicit formulas. There are essentially three methods: (i) orthogonal polynomials, (ii) Brézin-Hikami contour integration formula [49] and (iii) supersymmetric formalism. We are not aware of any orthogonal polynomial approach to analyse $Y^z = (X - z)(X-z)^*$ in the real case (see [67] in the complex case and [148] for rank-1 perturbation of real X). In the complex case, the ensemble Y^z has also been extensively investigated by the Brézin-Hikami formula in [28], where even the determinantal correlation kernel was computed as a double integral involving the Bessel kernel, see also [107, 116] for a derivation via the supersymmetric version of the Itzykson-Zuber formula. Although the paper [28] did not analyse the resulting one point function, well known asymptotics for the Bessel function may be used to rederive our bounds and asymptotics on (8.6), as well as (8.4a), from [28, Theorem 7.1], see Appendix 8.C for more details. For the real case, however, there is no analogue of the Brézin-Hikami formula.

Therefore, in this paper we explore the last option, the supersymmetric approach, that is available for both symmetry classes, albeit the real case is considerably more involved.

Our main tool is the powerful superbosonization formula [142] followed by a delicate multivariable contour integral analysis. We remark that, alternatively, one may also use the Hubbard-Stratonovich transformation, e.g. [3, Proposition 1] where correlation functions, i.e. expectations of *products* of characteristic polynomials of X were computed in this way. Note, however, that the density of states (8.6) requires to analyse *ratios* of determinants, a technically much more demanding task. While explicit formulas can be obtained with both methods (see [160] and especially [125] for an explicit comparison), the subsequent analysis seems to be more feasible with the formula obtained from the superbosonisation approach, as our work demonstrates.

Supersymmetry is a compelling method originated in physics [79, 106, 205] to produce surprising identities related to random matrices whose potential has not yet been fully exploited in mathematics. It has been especially successful in deriving rigorous result on Gaussian random band matrices [23, 68–70, 171–173, 175, 177], sometimes even beyond the Gaussian case [174, 176, 188], as well as on overlaps of non-Hermitian Ginibre eigenvectors [99]. We also mention the recent results in [99] and [100] as examples of a remarkable interplay between supersymmetric and orthogonal polynomial techniques in the theory of Ginibre and related matrices.

The main object of our work, the Hermitian block random matrix

$$H = H^z := \begin{pmatrix} 0 & X - z \\ X^* - \bar{z} & 0 \end{pmatrix} \tag{8.7}$$

arose in the physics literature as a chiral random matrix model for massless Dirac operator, introduced by Stephanov in [190]. Typically, instead of z and $\bar{z}$, both shift parameters are chosen equal z (interpreted as i-times the chemical potential) so that the corresponding H is not self-adjoint; this model has been extensively investigated by both supersymmetric and orthogonal polynomial techniques, see e.g. [8, 108, 153, 206, 211]. However, in the special case when z is real, our H^z as given in (8.7) coincides with Stephanov's model where z can be interpreted as temperature (or Matsubara frequency), see [207, Section 6.1].

Acknowledgement

The authors are grateful to Nicholas Cook and Patrick Lopatto for pointing out missing references, and to Ievgenii Afanasiev for useful remarks. We would also like to thank the anonymous referees for drawing our attention to additional references in the physics literature.

8.2 Model and main results

We consider the model $Y = Y^z = (X - z)(X^* - \bar{z})$ with a fixed complex parameter $z \in \mathbf{C}$ and with a random matrix $X \in \mathbf{C}^{N\times N}$ having independent real or complex Gaussian entries $x_{ab} \sim \mathcal{N}(0, N^{-1})$, where in the complex case we additionally assume $\mathbf{E}\, x_{ab}^2 = 0$. Note that Y is related to the block matrix (8.7) through its resolvent via

$$\frac{\operatorname{Tr}(H - \sqrt{w})^{-1}}{2\sqrt{w}} = \operatorname{Tr}(Y - w)^{-1}, \quad \Re w > 0,\ \Im w > 0, \tag{8.8}$$

where the branch of $\sqrt{w}$ is chosen such that $\Im\sqrt{w} > 0$. It is well known that in the large N limit the normalized trace of the resolvent of many random matrix ensembles becomes

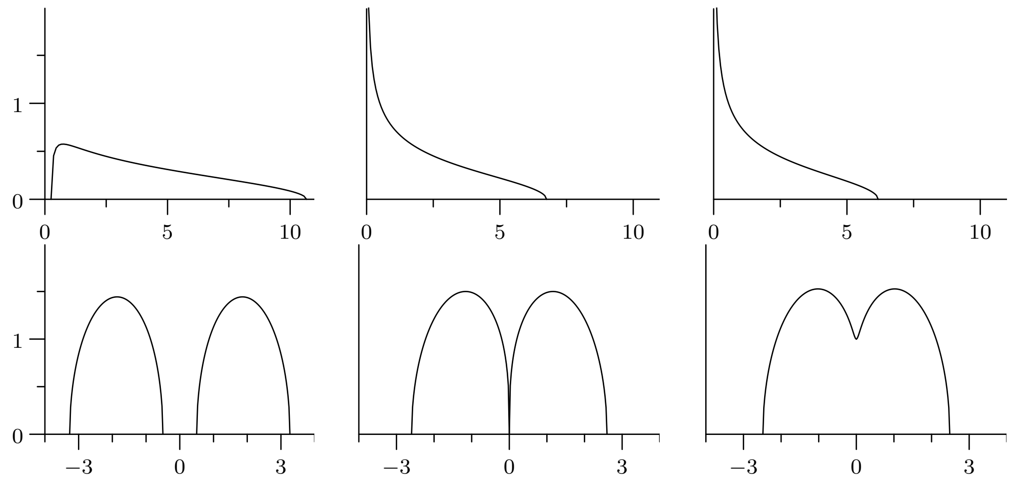


Figure 8.2: Density of states of Y^z and H^z around the cusp formation. The top and bottom figures show a plot of the boundary value of $\Im m^z = \Im m_{Y^z}$ and $\Im m_{H^z}$, respectively on the real line.

deterministic and it satisfies an algebraic equation, the matrix Dyson equation (MDE) [5]. In the current case of i.i.d. entries the MDE reduces to a simple cubic scalar equation

$$\frac{1}{m_{H^z}} + (w + m_{H^z}) - \frac{|z|^2}{w + m_{H^z}} = 0, \quad \Im m_{H^z}(w) > 0, \quad \Im w > 0 \tag{8.9}$$

that has a unique solution, denoted by m_{H^z}. The local law from [13] asserts that

$$\frac{1}{2N} \operatorname{Tr}(H^z - w)^{-1} = m_{H^z}(w) + \mathcal{O}_\prec((N\Im w)^{-1}), \tag{8.10}$$

where $\mathcal{O}_\prec$ denotes a suitable concept of high-probability error term. Together with (8.8) it follows that the normalized trace of the resolvent $(Y^z - w)^{-1}$ of Y^z is well approximated

$$\frac{1}{N} \operatorname{Tr}(Y^z - w)^{-1} \approx m^z(w)$$

by the unique solution $m = m^z = m_{Y^z}$ to the equation

$$\frac{1}{m^z} + w(1 + m^z) - \frac{|z|^2}{1 + m^z} = 0, \quad \Im m^z(w) > 0, \quad \Re w > 0, \ \Im w > 0, \tag{8.11}$$

which is given by $m^z(w) = m_{H^z}(\sqrt{w})/\sqrt{w}$. Since m approximates the trace of the resolvent, the density of states is obtained as the imaginary part of the continuous extension of m to the real line, i.e. $\varrho_\#(E) = \pi^{-1} \lim_{\epsilon \to 0+} \Im m_\#(E + \mathrm{i}\epsilon)$ for both choices $\# = H^z, Y^z$. For $\delta := 1 - |z|^2 \approx 0$ the Stieltjes transform m_{H^z} and its density of states exhibit a cusp formation at $w = 0$ as δ crosses the value 0. This cusp formation in H^z implies an analogous transition for m^z; the corresponding density of states are depicted in Figure 8.2.

Complex case

Our main result of the present paper in the complex case is an asymptotic double-integral formula for $\mathbf{E}\operatorname{Tr}(Y-w)^{-1}$ at $w=E+\mathrm{i}0$, $E\ge 0$. In the transitional regime it is convenient to introduce the rescaled variables

$$\lambda := E/c(N), \qquad \widetilde{\delta} := N^{1/2}\delta, \qquad \text{where} \quad \delta := 1-|z|^2, \tag{8.12}$$

recalling that $c(N)=c(N,z)$ was defined in (8.5). For $r\ge 0$ let $\Psi=\Psi(r)$ be the unique solution to the cubic equation $1+r\Psi+\Psi^3=0$ with $\Re\Psi,\Im\Psi>0$. It is easy to see that $\Psi(r)$ satisfies $\Psi(0)=e^{\mathrm{i}\pi/3}$ and $\Psi(r)\sim \mathrm{i}\sqrt{r}$ for $r\gg 1$. We also introduce the notations $a\wedge b := \min\{a,b\}$ and $a\vee b:=\max\{a,b\}$ for real numbers a,b.

Theorem 8.2.1 (Asymptotic 1-point function in the complex case). *Uniformly in $\widetilde{\delta}\ge -C$ and $0\le\lambda\le C$ for some fixed large constant $C>0$ we have*

$$\begin{aligned}\mathbf{E}\operatorname{Tr}(Y-\lambda\cdot c(N,\widetilde{\delta})-\mathrm{i}0)^{-1} &= \frac{1}{2\pi\mathrm{i}}\frac{N^{3/2}}{\widetilde{z}_*}\int \mathrm{d}x \oint \mathrm{d}y\, e^{h(y)-h(x)}\widetilde{H}(x,y)\\ &\quad + \mathcal{O}(N(1\vee\widetilde{\delta})(1+|\log\lambda|)),\end{aligned} \tag{8.13a}$$

where

$$\begin{aligned}\widetilde{H}(x,y) &:= \frac{1}{x^3}+\frac{1}{x^2y}+\frac{1}{xy^2}+\frac{\widetilde{\delta}\widetilde{z}_*}{xy}+\frac{\widetilde{\delta}\widetilde{z}_*}{x^2}, \quad h(x) := -(1\wedge\widetilde{\delta}^{-1})\lambda\widetilde{z}_* x+\frac{\widetilde{\delta}}{x\widetilde{z}_*}+\frac{1}{2x^2\widetilde{z}_*^2},\\ \widetilde{z}_* &:= \lambda^{-1/3}(1\vee\widetilde{\delta}^{1/3})|\Psi(\widetilde{\delta}\lambda^{-1/3}(1\vee\widetilde{\delta}^{1/3}))|, \qquad c(N,\widetilde{\delta}) = N^{-3/2}\cdot(1\wedge\widetilde{\delta}^{-1}),\end{aligned} \tag{8.13b}$$

and where the x-integration is over any contour from 0 to $e^{3\mathrm{i}\pi/4}\infty$, going out from 0 in the direction of the positive real axis, and the y-integration is over any contour around 0 in a counter-clockwise direction. Moreover, in the regime $\lambda\ll 1$ we have the bound

$$\left|\frac{1\wedge\widetilde{\delta}^{-1}}{\widetilde{z}_*}\int\mathrm{d}x\oint\mathrm{d}y\, e^{h(y)-h(x)}\widetilde{H}(x,y)\right| \lesssim \begin{cases}|\log\lambda|, & \lambda\ge\widetilde{\delta}^3,\\ |\log\lambda\widetilde{\delta}|, & \lambda<\widetilde{\delta}^3.\end{cases} \tag{8.13c}$$

In the regime above the eigenvalue scaling, i.e. for $\lambda\gg 1$, the analogue of Theorem 8.2.1 reduces essentially to the local law asymptotics (8.10), albeit with a better error term due to the presence of the expectation.

Proposition 8.2.2. *Let $Y^z=(X-z)(X-z)^*$ where X is a complex Ginibre ensemble. Then, uniformly in $\delta:=1-|z|^2$ and $E\in\mathbf{R}$, we have the asymptotic expansion in $E\pm := E-\mathfrak{e}_\pm$,*

$$\begin{aligned}&\mathbf{E}\operatorname{Tr}(Y^z-E-\mathrm{i}0)^{-1} = Nm^z(E+\mathrm{i}0)\\ &\quad\times\left(1+\mathcal{O}\Big(\frac{1}{N|E_+|^{3/2}}+\frac{1}{NE^{2/3}}\wedge\Big(\frac{\mathbf{1}_{\delta\ge 0}}{NE^{1/2}\delta^{1/2}}+\frac{\mathbf{1}_{\delta<0}}{N|E_-|^{3/2}|\delta|^{5/2}}\Big)\Big)\right).\end{aligned} \tag{8.14}$$

where the edges $\mathfrak{e}_\pm$ of $\Im m^z$ are explicit functions of δ given in (8.18a).

Real case

In the real case our main result is the following optimal bound on $\mathbf{E}\,\mathrm{Tr}(Y+E)^{-1}$ for $E>0$. Recall the notation $\delta := 1-|z|^2$.

Theorem 8.2.3 (Optimal bound on the resolvent trace in the real case)**.** *Let $\rho>0$ be any small constant. Then uniformly in $E\ge 0$ and $\delta \ge -CN^{-1/2}$ for some fixed large constant $C>0$ we have that*

$$|\mathbf{E}\,\mathrm{Tr}(Y+E)^{-1}| \lesssim \frac{e^{-\frac{1}{2}N(\Im z)^2}[N^{3/4}\vee N\sqrt{|\delta|}]}{\sqrt{E}} + (N^{3/2}\vee N^2|\delta|)[1+|\log(NE^{2/3})|] \tag{8.15}$$

Finally, we present our bound on the tail asymptotics for both real and complex cases; for most applications, this can be viewed as the main result of this paper. Since a sizeable contribution to $\Im\,\mathrm{Tr}(Y-E+\mathrm{i}0)^{-1}$ and $\Im\,\mathrm{Tr}(Y+E)^{-1}$ comes from the smallest eigenvalue $\lambda_1(Y^z)$, by a straightforward Markov inequality we immediately obtain the following corollary on the tail asymptotics of $\lambda_1(Y^z)$ as an easy consequence of Theorems 8.2.1 and 8.2.3.

Corollary 8.2.4 (Tail asymptotics of $\lambda_1(Y^z)$)**.** *For any $C>0$, uniformly in $x\in(0,C]$ and $1-|z|^2>-CN^{-1/2}$ we have the bound*

$$\mathbf{P}\Big(\lambda_1(Y^z)\le c(N,z)x\Big) \lesssim (1+|\log x|)x \tag{8.16}$$

in the complex case, and

$$\mathbf{P}\Big(\lambda_1(Y^z)\le c(N,z)x\Big) \lesssim e^{-\frac{1}{2}N(\Im z)^2}\sqrt{x} + (1+|\log x|)x \tag{8.17}$$

in the real case, where we recall the definition of the scaling factor $c(N,z)$ from (8.5).

Properties of the asymptotic Stieltjes transform m^z

We now record some information on the deterministic Stieltjes transform m^z which will be useful later. The endpoints of the support of the density of states $\pi^{-1}\Im m^z$ are the zeros of the discriminant of the cubic equation (8.11) since passing through these points with the real parameter $E=\Re w$ creates solutions with nonzero imaginary part. Elementary calculations show that the support of $\Im m_{Y^z}$ is $[0,\mathfrak{e}_+]$ if $0\le\delta\le 1$ and it is $[\mathfrak{e}_-,\mathfrak{e}_+]$ if $\delta<0$, where

$$\mathfrak{e}_\pm := \frac{8\delta^2 \pm (9-8\delta)^{3/2} - 36\delta + 27}{8(1-\delta)}, \tag{8.18a}$$

and $\mathfrak{e}_-$ is only considered if $\delta<0$. Note that while $\mathfrak{e}_+\sim 1$, the edge $\mathfrak{e}_-$ may be close to 0; more precisely $0<\mathfrak{e}_-=-4\delta^3/27(1+\mathcal{O}(|\delta|))$. The slope coefficient of the square-root density at the edge in $\mathfrak{e}_\pm$ is given by

$$\Im m(\mathfrak{e}_\pm \mp \lambda) = \begin{cases} \gamma_\pm\sqrt{\lambda}\Big(1+\mathcal{O}(\sqrt{\lambda})\Big), & \lambda\ge 0, \\ 0 & \lambda \le 0, \end{cases}, \quad \gamma_\pm := \frac{2\sqrt{2}\left(\sqrt{9-8\delta}\pm 1\right)^{3/2}}{\left(\sqrt{9-8\delta}\pm 3\right)^{5/2}\sqrt[4]{9-8\delta}}. \tag{8.18b}$$

Note that while the square-root edge at $\mathfrak{e}_+$ is non-singular in the sense $\gamma_+ \sim 1$, the square-root edge in $\mathfrak{e}_-$ becomes singular for small $|\delta|$ as

$$\gamma_- = \frac{9}{4|\delta|^{5/2}}\Big(1 + \mathcal{O}(|\delta|)\Big).$$

8.3 Supersymmetric method

Let $\chi_1, \overline{\chi_1}, \ldots, \chi_N, \overline{\chi_N}$ denote Grassmannian variables satisfying the commutation rules

$$\chi_i\chi_j = -\chi_j\chi_i, \quad \chi_i\overline{\chi_j} = -\overline{\chi_j}\chi_i, \quad \overline{\chi_i\chi_j} = -\overline{\chi_j\chi_i},$$

from which it follows that $\chi_i^2 = \overline{\chi_i}^2 = 0$. As a convention we set $\overline{\overline{\chi_i}} := -\chi_i$. The power series of any function of Grassmannian variables is multilinear and it suffices to define the integral in the sense of Berezin [29] over Grassmannian variables as the derivatives

$$\partial_{\chi_k}\chi_k = \partial_{\overline{\chi_k}}\overline{\chi_k} = 1, \quad \partial_{\chi_k}1 = \partial_{\overline{\chi_k}}1 = 0, \quad \partial_\chi := \partial_{\chi_1}\partial_{\overline{\chi_1}}\ldots\partial_{\chi_N}\partial_{\overline{\chi_N}}$$

and extend them multilinearly to all finite combinations of monomials in Grassmannians. We denote the column vectors with entries $\chi_1, \ldots, \chi_N$ and $\overline{\chi_1}, \ldots, \overline{\chi_N}$ by χ and $\overline{\chi}$, respectively. The conjugate transposes of those vectors, i.e. the row vectors with entries $\overline{\chi_1}, \ldots, \overline{\chi_N}$ and $-\chi_1, \ldots, -\chi_N$ will be denoted by χ^* and $\overline{\chi}^*$, respectively. Note that $(\chi^*)^* = -\chi$, $[\overline{\chi}^*]^* = -\overline{\chi}$. We now define the inner product of Grassmannian vectors χ, ϕ by

$$\langle \chi, \phi\rangle := \sum_i \overline{\chi_i}\phi_i,$$

so that the quadratic form $\sum_{i,j}\overline{\chi_i}A_{ij}\chi_j$ can be written as

$$\langle \chi, A\chi\rangle = \sum_{i,j}\overline{\chi_i}A_{ij}\chi_j,$$

where the matrix-vector product is understood in its usual sense. Similarly, s and $\overline{s}$ denote the column vectors with complex entries $s_1, \ldots, s_N$ and their complex conjugates $\overline{s_1}, \ldots, \overline{s_N}$, respectively, and for the conjugate transpose we have $(s^*)^* = s$ as usual. We have

$$\langle s, \phi\rangle := \sum_i \overline{s_i}\phi_i, \qquad \langle \chi, s\rangle := \sum_i \overline{\chi_i}s_i,$$

and similarly for quadratic forms. The commutation rules naturally also apply to linear functions of the Grassmannians, and therefore also, for example, $\langle s, \chi\rangle^2 = \langle \chi, s\rangle^2 = 0$ for any vector s of complex numbers. The complex numbers s_i and often called *bosonic* variables, while Grassmannians are called *fermions*, motivated by the basic (anti)commutativity of the bosonic/fermionic field operators in physics.

8.3.1 Determinant identities

The backbone of the supersymmetric method are the determinant identities

$$\frac{1}{\mathrm{i}^N}\frac{\operatorname{sgn}(\Im w)^N}{\det(H-w)} = \int_{\mathbf{C}^N} \exp\big(-\mathrm{i}\operatorname{sgn}(\Im w)\langle s, (H-w)s\rangle\big)\,\mathrm{d}s, \quad \mathrm{d}s := \prod_{i=j}^N \frac{\mathrm{d}\Re s_j\,\mathrm{d}\Im s_j}{\pi}$$

$$\mathrm{i}^N\det(H-w) = \partial_\chi \exp(\mathrm{i}\langle\chi, (H-w)\chi\rangle), \quad \partial_\chi := \partial_{\chi_1}\partial_{\overline{\chi_1}}\ldots\partial_{\chi_N}\partial_{\overline{\chi_N}},$$

where the exponential is defined by its (terminating) Taylor series. Consequently we can conveniently express the generating function as

$$Z(w,w_1) := \mathbf{E}\frac{\det(H-w_1)}{\det(H-w)} = \mathbf{E}\int_{\mathbf{C}^N} \mathrm{d}s\,\partial_\chi \exp\Big(\mathrm{i}\langle\chi,(H-w_1)\chi\rangle - \mathrm{i}\langle s,(H-w)s\rangle\Big),$$

for $w \in \mathbf{H}$ and $w_1 \in \mathbf{C}$, where choice of w with $\Im w > 0$ guarantees the convergence of the integral. By taking the w_1 derivative and setting $w = w_1$ it follows that

$$\begin{aligned} \mathrm{Tr}(H-w)^{-1} &= -\frac{\partial}{\partial w_1}\frac{\det(H-w_1)}{\det(H-w)}\bigg|_{w_1=w} = \mathrm{i}\int \langle\chi,\chi\rangle e^{-\mathrm{i}\,\mathrm{Tr}(H-w)[ss^*+\chi\chi^*]}, \\ \int &:= \int_{\mathbf{C}^N}\mathrm{d}s\,\partial_\chi. \end{aligned} \tag{8.19}$$

8.3.2 Superbosonization identity

After taking expectations, i.e. performing the Gaussian integration for the entries of $Y = Y^z = (X-z)(X-z)^*$, the resolvent identity (8.19) will depend on the complex vector s and the Grassmannian vector χ only via certain inner products. More specifically, after defining the $N\times 2$ and $N\times 4$ matrices $\Phi := (s,\chi)$ and $\Psi := (s,\overline{s},\chi,\overline{\chi})$, the expectation of the resolvent can be expressed as an integral over the 2×2 or 4×4 supermatrices $\Phi^*\Phi$ or $\Psi^*\Psi$ in the complex and real case, respectively. *Supermatrices* are 2×2 block matrices whose diagonal blocks are commonly referred to as the boson-boson and the fermion-fermion block, while the off-diagonal blocks are the boson-fermion and fermion-boson block. For supermatrices Q the *supertrace* and *superdeterminant*, the natural generalizations of trace and determinant, are given by

$$\mathrm{STr}\begin{pmatrix} x & \sigma \\ \tau & y\end{pmatrix} := \mathrm{Tr}(x) - \mathrm{Tr}(y), \quad \mathrm{SDet}\begin{pmatrix} x & \sigma \\ \tau & y\end{pmatrix} := \frac{\det(x)}{\det(y-\tau x^{-1}\sigma)}, \tag{8.20}$$

and the inverse of a supermatrix is

$$\begin{pmatrix} x & \sigma \\ \tau & y\end{pmatrix}^{-1} = \begin{pmatrix} (x-\tau y^{-1}\sigma)^{-1} & -x^{-1}\sigma(y-\sigma x^{-1}\tau) \\ -y^{-1}\tau(x-\tau y^{-1}\sigma)^{-1} & (y-\sigma x^{-1}\tau)^{-1}\end{pmatrix}. \tag{8.21}$$

The integral over the remaining degrees of freedom in Φ, Ψ other than the inner products in $\Phi^*\Phi$, $\Psi^*\Psi$ can conveniently be performed using the well known *superbosonization formula* which we now recall. It basically identifies the integration volume of the irrelevant degrees of freedom with the high power of the superdeterminant of the supermatrix containing the relevant inner products (collected in a 2×2 supermatrix Q in the complex case and a 4×4 supermatrix Q in the real case).

8.3.2.1 Complex superbosonization

For any analytic function F with sufficiently fast decay at $+\infty$ in the boson-boson sector (in the variable x) the complex superbosonization identity from [142, Eq. (1.10)] implies

$$\int F(\Phi^*\Phi) = \int_Q \mathrm{SDet}^N(Q)F(Q), \quad \int_Q := \frac{1}{2\pi\mathrm{i}}\int \mathrm{d}x \oint \mathrm{d}y\,\partial_\sigma\partial_\tau, \quad Q := \begin{pmatrix} x & \sigma \\ \tau & y\end{pmatrix}, \tag{8.22}$$

where $\int \mathrm{d}x$ denotes the Lebesgue integral on $[0,\infty)$, $\oint \mathrm{d}y$ denotes the counterclockwise complex line integral on $\{z \in \mathbf{C} | |z| = 1\}$ and σ, τ denote independent scalar Grassmannian variables. The key point is that while the integral on the left hand side is performed over N complex numbers and $2N$ Grassmannians, the integral on the right is simply over a 2×2 supermatrix, i.e. two complex variables and two Grassmannians. Note that the identity in [142] is more general than (8.22) in the sense that it allows for bosonic and fermionic sectors of unequal sizes. For the case of equal sizes, which concerns us, the formula gets simplified, the measure DQ in [142, Eq. (1.8)] becomes the flat Lebesgue measure since two determinants cancel each other as

$$\begin{aligned}\det(1 - x^{-1}\sigma y^{-1}\tau)\det(1 - y^{-1}\tau x^{-1}\sigma) &= e^{\mathrm{Tr}(\log(1-x^{-1}\sigma y^{-1}\tau)+\log(1-y^{-1}\tau x^{-1}\sigma))}\\ &= e^{-\sum_{k\geq 1}\frac{1}{k}\mathrm{Tr}\left((x^{-1}\sigma y^{-1}\tau)^k+(y^{-1}\tau x^{-1}\sigma)^k\right)} = 1,\end{aligned} \tag{8.23}$$

where the sum is finite and the last equality followed using the commutation rules.

8.3.2.2 Real superbosonization

In the real case we similarly have the real superbosonization identity from [142, Eq. (1.13)]

$$\begin{aligned}\int F(\Psi^*\Psi) &= \int_Q \mathrm{SDet}^{N/2}(Q)F(Q),\\ \int_Q &:= \frac{1}{(2\pi)^2\mathrm{i}}\int \mathrm{d}x \oint \mathrm{d}y\, \partial_\sigma\Big(\frac{\det(y)}{\det(x)}\Big)^{1/2}\det\Big(1 - \frac{x^{-1}}{y}\sigma\tau\Big)^{-1/2}\end{aligned} \tag{8.24}$$

The supermatrix Q has 2×2 blocks: x is non-negative Hermitian, y is a scalar multiple of the identity matrix. The off-diagonal blocks σ, τ are related by

$$\tau := -t_a\sigma^t t_s^{-1}, \quad t_s := \begin{pmatrix} 0 & 1 \\ bm1 & 0 \end{pmatrix}, \quad t_a := \begin{pmatrix} 0 & -1 \\ bm1 & 0 \end{pmatrix}.$$

Here the $\int \mathrm{d}x$ integral is the Lebesgue measure on non-negative Hermitian 2×2 matrices x satisfying $x_{11} = x_{22}$, i.e.

$$\int \mathrm{d}x := \int_0^\infty \mathrm{d}_{x_{11}} \int_{|x_{12}|\leq x_{11}} \mathrm{d}\Re x_{12}\, \mathrm{d}\Im x_{12},$$

and the fermionic integral is defined as $\partial_\sigma := \partial_{\sigma_{11}}\partial_{\sigma_{22}}\partial_{\sigma_{21}}\partial_{\sigma_{12}}$. Furthermore, under the slight abuse of notation of identifying the 2×2 matrix y, which is a scalar multiple of the identity matrix, with the corresponding scalar, $\oint \mathrm{d}y$ is the complex line integral over $|y| = 1$ in a counter-clockwise direction. Unlike in the complex case, the matrix elements of the 4×4 supermatrix Q are not independent; there are only 4 (real) bosonic and 4 fermionic degrees of freedom. These identities among the elements of Q stem from natural relations between the scalar product of the column vectors of Ψ. For example the identity $\langle s, s\rangle = \langle \overline{s}, \overline{s}\rangle$ from the first two diagonal elements of $(\Psi^*\Psi)$ corresponds to $x_{11} = x_{22}$, while $\langle \overline{\chi}, \overline{\chi}\rangle = 0$ is responsible for $y_{12} = 0$. The relation

$$\tau = \begin{pmatrix} \sigma_{22} & \sigma_{12} \\ -\sigma_{21} & -\sigma_{11} \end{pmatrix}$$

encoded in the last line of (8.24) corresponds to relations between scalar products of bosonic and fermionic vectors and their complex conjugates; for example $\tau_{21} = -\sigma_{21}$ comes from the identity $(\Psi^*\Psi)_{41} = \langle -\bar{\chi}, s\rangle = -\langle \bar{s}, \chi\rangle = -(\Psi^*\Psi)_{23}$, etc.

8.3.3 Application to Y^z in the complex case

Our goal is to evaluate $\mathbf{E}\,\mathrm{Tr}(H-w)^{-1}$ asymptotically on the scale where E is comparable with the eigenvalue spacing. We now use the identity

$$\begin{aligned}\mathrm{Tr}(Y-w)^{-1} &= \mathrm{i}\int \langle \chi,\chi\rangle e^{-\mathrm{i}\,\mathrm{Tr}[(X-z)(X^*-\overline{z})-w](ss^*+\chi\chi^*)}\\ &= \mathrm{i}\int \langle \chi,\chi\rangle e^{\mathrm{i}w\langle s,s\rangle-\mathrm{i}w\langle\chi,\chi\rangle-\mathrm{i}\,\mathrm{Tr}(X-z)(X^*-\overline{z})\Phi\Phi^*}\end{aligned}\tag{8.25}$$

for $w = E+\mathrm{i}\epsilon$ with $|E| \gg \epsilon > 0$. We now compute the ensemble average as

$$\begin{aligned}&\mathbf{E}\, e^{-\mathrm{i}\,\mathrm{Tr}(X-z)(X^*-\overline{z})\Phi\Phi^*}\\ &= \Big(\frac{N}{\pi}\Big)^{N^2}\int \exp\Big[-N\,\mathrm{Tr}\, X^*\Big(1+\mathrm{i}\frac{\Phi\Phi^*}{N}\Big)X + \mathrm{i}\overline{z}\,\mathrm{Tr}\,\Phi\Phi^* X + \mathrm{i}z\,\mathrm{Tr}\, X^*\Phi\Phi^* - \mathrm{i}|z|^2\,\mathrm{Tr}\,\Phi\Phi^*\Big]\\ &= \mathrm{SDet}\Big(1+\mathrm{i}\frac{\Phi^*\Phi}{N}\Big)^{-N}\exp\Big(-\mathrm{i}|z|^2\Big[\mathrm{Tr}\,\Phi\Phi^* - \frac{\mathrm{i}}{N}\,\mathrm{Tr}\,\Phi\Phi^*\Big(1+\mathrm{i}\frac{\Phi\Phi^*}{N}\Big)^{-1}\Phi\Phi^*\Big]\Big),\\ &= \mathrm{SDet}\Big(1+\mathrm{i}\frac{\Phi^*\Phi}{N}\Big)^{-N}\exp\Big(-\mathrm{i}|z|^2\,\mathrm{Tr}\,\Phi\Big(1+\frac{\mathrm{i}}{N}\Phi^*\Phi\Big)^{-1}\Phi^*\Big),\\ &= \mathrm{SDet}\Big(1+\mathrm{i}\frac{\Phi^*\Phi}{N}\Big)^{-N}\exp\Big(-N|z|^2\,\mathrm{STr}\Big(1+\frac{\mathrm{i}}{N}\Phi^*\Phi\Big)^{-1}\frac{\mathrm{i}}{N}\Phi^*\Phi\Big).\end{aligned}\tag{8.26}$$

To perform the $\int = \int_{\mathbf{C}^N}\mathrm{d}s\,\partial_\chi$ integration in (8.25) we use the superbosonization formula (8.22) for the function

$$\begin{aligned}F(\Phi^*\Phi) := \langle\chi,\chi\rangle\,\mathrm{SDet}\Big(1+\mathrm{i}\frac{\Phi^*\Phi}{N}\Big)^{-N}\exp\Big(&-N|z|^2\,\mathrm{STr}\Big(1+\frac{\mathrm{i}}{N}\Phi^*\Phi\Big)^{-1}\frac{\mathrm{i}}{N}\Phi^*\Phi\\ &+\mathrm{i}w\,\mathrm{STr}\,\Phi^*\Phi\Big),\end{aligned}\tag{8.27}$$

We view F as a function of the four independent variables collected in the entries of the 2×2 matrix $\Phi^*\Phi$. Strictly speaking the function F is only meromorphic but not entire in these four variables, but since the integration regimes on both sides of the superbosonisation formula are well separated away from the poles of F, a simple approximation argument outlined in Appendix 8.A justifies its usage. Together with the change of variables $\frac{\mathrm{i}}{N}Q \mapsto Q$ it now implies that

$$\mathbf{E}\,\mathrm{Tr}(Y-w)^{-1} = N\int_{Q'} y e^{Nw\,\mathrm{STr}(Q)+N\log\mathrm{SDet}(Q)-N\log\mathrm{SDet}(1+Q)-N|z|^2\,\mathrm{STr}(1+Q)^{-1}Q}$$

where $\int_{Q'}$ indicates the changed integration regime due to the change of variables, more specifically under, Q' the x-integration is over $[0,\mathrm{i}\infty)$ and the y-integration is over a small circle $\{u\in\mathbf{C}||u| = N^{-1}\}$. Note that the change of variables through scaling does not contribute an additional factor, since superdeterminants are scale invariant. It remains to perform the Berezinian integral. To do so we split

$$Q = q + \mu, \quad q = \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}, \quad \mu = \begin{pmatrix} 0 & \sigma \\ \tau & 0 \end{pmatrix}$$

and compute

$$\begin{aligned} \exp\Big(-N \log \mathrm{SDet}(1+Q)\Big) &= \exp\Big(-N \,\mathrm{STr} \log(1+q+\mu)\Big) \\ &= \exp\Big(-N \,\mathrm{STr}\log(1+q) + \frac{N}{2}\,\mathrm{STr}(1+q)^{-1}\mu(1+q)^{-1}\mu\Big) \\ &= \exp\Big(-N\log(1+x) + N\log(1+y)\Big)\Big(1 + N\frac{\sigma\tau}{(1+x)(1+y)}\Big) \\ \exp\Big(N\log \mathrm{SDet}(Q)\Big) &= \exp\Big(N\log(x) - N\log(y)\Big)\Big(1 - N\frac{\sigma\tau}{xy}\Big) \end{aligned}$$

and

$$\begin{aligned} &\exp\Big(-N|z|^2\,\mathrm{STr}(1+Q)^{-1}Q\Big) \\ &\quad = \exp\Big(-N|z|^2\,\mathrm{STr}\Big[(1+q)^{-1}q - (1+q)^{-1}\mu(1+q)^{-1}\mu(1+q)^{-1}\Big]\Big) \\ &\quad = \exp\Big(-N|z|^2\frac{x}{1+x} + N|z|^2\frac{y}{1+y}\Big)\Big(1 + N|z|^2\frac{\sigma\tau}{(1+x)(1+y)}\Big(\frac{1}{1+x} + \frac{1}{1+y}\Big)\Big). \end{aligned}$$

By combining these identities we arrive at the final result[1]

$$\begin{aligned} \mathbf{E}\,\mathrm{Tr}(Y-w)^{-1} &= \frac{N^2}{2\pi \mathrm{i}}\int_0^\infty \mathrm{d}x \oint \mathrm{d}y e^{-Nf(x)+Nf(y)} y \cdot G(x,y), \\ G(x,y) &:= \frac{1}{xy} - \frac{1}{(1+x)(1+y)}\Big[1 + \frac{|z|^2}{1+x} + \frac{|z|^2}{1+y}\Big], \\ f(x) &:= \log\frac{1+x}{x} - \frac{|z|^2}{1+x} - wx, \end{aligned} \tag{8.28}$$

where the x-integration is over $(0, \mathrm{i}\infty)$ and the y-integration is over a circle of radius N^{-1} around the origin.

8.3.4 Application to Y^z in the real case

We now consider the real case and introduce the $N \times 4$ matrix $\Psi := (s, \overline{s}, \chi, \overline{\chi})$, the 2×2 matrix $Z := \left(\begin{smallmatrix} z & 0 \\ 0 & \overline{z} \end{smallmatrix}\right)$ and the 4×4 matrix $Z_2 := \left(\begin{smallmatrix} Z & 0 \\ 0 & Z \end{smallmatrix}\right)$, and use that

$$\mathrm{i}\overline{z}\,\mathrm{Tr}\,\Phi\Phi^* X + \mathrm{i}z\,\mathrm{Tr}\,X^t\Phi\Phi^* = \frac{\mathrm{i}}{2}\,\mathrm{Tr}\,\Psi Z_2^*\Psi^* X + \frac{\mathrm{i}}{2}\,\mathrm{Tr}\,X^t\Psi Z_2\Psi^*, \quad \Big(\Psi Z_2 \Psi^*\Big)^t = \Psi Z_2^* \Psi^*$$

[1]Essentially the same formula, obtained by direct computations, was presented by M. Shcherbina in her seminar talk on Jan 11, 2016 in Bonn. Our derivation of the same formula via superbosonization is merely a pedagogical preparation to the much more involved real case in Section 8.3.4.

to compute

$$\begin{aligned}
&\mathbf{E}\, e^{-\mathrm{i}\operatorname{Tr}(X-z)(X^t-\overline{z})\Phi\Phi^*}\\
&= \mathbf{E}\Big(\frac{N}{2\pi}\Big)^{N^2/2}\int \exp\Big(-\frac{N}{2}\operatorname{Tr} X^t\Big(1+\frac{2\mathrm{i}}{N}\Phi\Phi^*\Big)X + \mathrm{i}\overline{z}\operatorname{Tr}\Phi\Phi^* X + \mathrm{i}z\operatorname{Tr} X^t\Phi\Phi^*\\
&\qquad\qquad - \mathrm{i}|z|^2\operatorname{Tr}\Phi\Phi^*\Big)\\
&= \mathbf{E}\Big(\frac{N}{2\pi}\Big)^{N^2/2}\int \exp\Big(-\frac{N}{2}\operatorname{Tr} X^t\Big(1+\frac{\mathrm{i}}{N}\Psi\Psi^*\Big)X + \frac{\mathrm{i}}{2}\operatorname{Tr}\Psi Z_2^*\Psi^* X\\
&\qquad\qquad + \frac{\mathrm{i}}{2}\operatorname{Tr} X^t\Psi Z_2\Psi^* - \frac{\mathrm{i}|z|^2}{2}\operatorname{Tr}\Psi\Psi^*\Big)\\
&= \operatorname{SDet}\Big(1+\frac{\mathrm{i}}{N}\Psi^*\Psi\Big)^{-N/2}\exp\Big(-\frac{\mathrm{i}}{2}\operatorname{Tr}\Psi Z_2^*\Big(1-\frac{\mathrm{i}}{N}\Psi^*\Big(1+\frac{\mathrm{i}}{N}\Psi\Psi^*\Big)^{-1}\Psi\Big)Z_2\Psi^*\Big)\\
&= \operatorname{SDet}\Big(1+\frac{\mathrm{i}}{N}\Psi^*\Psi\Big)^{-N/2}\exp\Big(-\frac{\mathrm{i}}{2}\operatorname{Tr}\Psi Z_2^*\Big(1+\frac{\mathrm{i}}{N}\Psi^*\Psi\Big)^{-1}Z_2\Psi^*\Big)\\
&= \operatorname{SDet}\Big(1+\frac{\mathrm{i}}{N}\Psi^*\Psi\Big)^{-N/2}\exp\Big(-\frac{N}{2}\operatorname{STr}\Big(1+\frac{\mathrm{i}}{N}\Psi^*\Psi\Big)^{-1}Z_2\frac{\mathrm{i}}{N}\Psi^*\Psi Z_2^*\Big),
\end{aligned} \tag{8.29}$$

where we used that X is real and $\Psi\Psi^*$ is symmetric. The superbosonization formula thus implies

$$\begin{aligned}
\mathbf{E}\operatorname{Tr}(H-w)^{-1} &= \frac{N}{2(2\pi)^2\mathrm{i}}\int\frac{\operatorname{Tr}(y)\det(y)^{1/2}}{\det(x)^{1/2}}\exp\Big(-\frac{1}{2}\log\det(1-x^{-1}\sigma y^{-1}\tau)\Big)\\
&\times\exp\Big(\frac{N}{2}\Big[\operatorname{STr}(wQ)-\log\operatorname{SDet}(1+Q)+\log\operatorname{SDet}(Q)-\operatorname{STr}(1+Q)^{-1}Z_2QZ_2^*\Big]\Big)\\
&= \frac{N}{2(2\pi)^2}\int\frac{\operatorname{Tr}(y)\det(y)^{1/2}}{\det(x)^{1/2}}\exp\Big(-\frac{1}{2}\operatorname{Tr}\log(1-x^{-1}\sigma y^{-1}\tau)\Big)\\
&\times\exp\Big(\frac{N}{2}\Big[\operatorname{STr}(wQ)-\operatorname{STr}\log(1+Q)+\operatorname{STr}\log(Q)-\operatorname{STr}(1+Q)^{-1}Z_2QZ_2^*\Big]\Big)
\end{aligned}$$

where

$$Q=\begin{pmatrix}x & \sigma\\ \tau & y\end{pmatrix}\equiv\frac{\mathrm{i}}{N}\Psi^*\Psi,\qquad \tau=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\sigma^t\begin{pmatrix}0 & 1\\ bm1 & 0\end{pmatrix}.$$

In order to expand the exponential terms to fourth order in σ we introduce the short-hand notations

$$Q=q+\mu,\quad q=\begin{pmatrix}x & 0\\ 0 & y\end{pmatrix},\quad \mu=\begin{pmatrix}0 & \sigma\\ \tau & 0\end{pmatrix}. \tag{8.30}$$

We compute

$$\begin{aligned}
\operatorname{STr}(1+Q)^{-1}Z_2QZ_2^* &= \operatorname{STr}(1+q)^{-1}Z_2qZ_2^*\\
&\quad-\operatorname{STr}(1+q)^{-1}\mu(1+q)^{-1}\Big(Z_2\mu Z_2^*-\mu(1+q)^{-1}Z_2qZ_2^*\Big)\\
&\quad-\operatorname{STr}\big((1+q)^{-1}\mu\big)^3(1+q)^{-1}\Big(Z_2\mu Z_2^*-\mu(1+q)^{-1}Z_2qZ_2^*\Big),\\
&= \operatorname{Tr}(1+x)^{-1}ZxZ^*-|z|^2\operatorname{Tr}\frac{y}{1+y}-\operatorname{Tr}(\sigma Z\tau Z^*A+Z\sigma Z^*\tau A)\\
&\quad+\operatorname{Tr}\sigma\tau AC'-\operatorname{Tr}\sigma\tau A(\sigma Z\tau Z^*A+Z\sigma Z^*\tau A)+\operatorname{Tr}\sigma\tau A\sigma\tau AC',
\end{aligned}$$

where we introduced matrices A, C' as in

$$A := \frac{(1+x)^{-1}}{1+y}, \quad B := \frac{x^{-1}}{y}, \quad C' := ZxZ^*(1+x)^{-1} + |z|^2 \frac{y}{1+y},$$

as well as B, which will be used in the sequel. In deriving these formulas we used that q and Z_2 have zero off-diagonal blocks and μ has zero diagonal blocks, to eliminate terms with odd powers of μ after taking the supertrace, and that y is a scalar multiple of the identity. Similarly we find for the logarithmic terms

$$\begin{aligned}
&\operatorname{STr}\Big(\log(q+\mu) - \log(1+q+\mu)\Big) \\
&= \operatorname{STr}\log(q(1+q)^{-1}) - \frac{1}{2}\operatorname{STr}(q^{-1}\mu)^2 - \frac{1}{4}\operatorname{STr}(q^{-1}\mu)^4 \\
&\quad + \frac{1}{2}\operatorname{STr}((1+q)^{-1}\mu)^2 + \frac{1}{4}\operatorname{STr}((1+q)^{-1}\mu)^4 \\
&= \log\frac{\det(x)}{\det(1+x)} - \log\frac{\det(y)}{\det(1+y)} + \operatorname{Tr}\sigma\tau(A-B) + \frac{1}{2}\operatorname{Tr}(\sigma\tau A)^2 - \frac{1}{2}\operatorname{Tr}(\sigma\tau B)^2
\end{aligned}$$

and

$$-\frac{1}{2}\operatorname{Tr}\log(1 - x^{-1}\sigma y^{-1}\tau) = \frac{1}{2}\operatorname{Tr}\sigma\tau B + \frac{1}{4}\operatorname{Tr}(\sigma\tau B)^2.$$

Whence

$$\begin{aligned}
\mathbf{E}\operatorname{Tr}(H-w)^{-1} &= \frac{N}{2(2\pi)^2\mathrm{i}}\int \mathrm{d}x \oint \mathrm{d}y \frac{\operatorname{Tr}(y)\det(y)^{1/2}}{\det(x)^{1/2}} \exp\Big(-\frac{N}{2}f(x) + \frac{N}{2}f(y)\Big) G(x,y,z) \\
f(x) &:= -w\operatorname{Tr}x + \log\frac{\det(1+x)}{\det(x)} + \operatorname{Tr}ZxZ^*(1+x)^{-1} - 2|z|^2, \\
G(x,y,z) &:= \partial_\sigma \exp\Big[\frac{N}{2}\Big(\operatorname{Tr}\sigma\tau(A(1-C') - (1-\frac{1}{N})B) + \operatorname{Tr}(\sigma Z\tau Z^* A + Z\sigma Z^*\tau A) \\
&\quad + \operatorname{Tr}\sigma\tau A\Big(\sigma\tau A(\frac{1}{2} - C') + \sigma Z\tau Z^* A + Z\sigma Z^*\tau A\Big) - \frac{1}{2}(1-\frac{1}{N})\operatorname{Tr}(\sigma\tau B)^2\Big)\Big],
\end{aligned} \tag{8.31}$$

where $\int \mathrm{d}x = \int \mathrm{d}x_{11}\,\mathrm{d}\Re x_{12}\,\mathrm{d}\Im x_{12}$ is the integral over matrices of the form

$$x = \begin{pmatrix} \mathrm{i}x_{11} & \mathrm{i}x_{12} \\ \mathrm{i}\overline{x_{12}} & \mathrm{i}x_{11} \end{pmatrix}$$

with $x_{11} \in [0,\infty)$ and $x_{12} \in \mathbf{C}$ with $|x_{12}| \le x_{11}$. The integral $\oint \mathrm{d}y = \oint \mathrm{d}y_{11}$ is the integral over scalar matrices $y = y_{11}I$ with $\mathrm{d}y_{11}$ being the complex line integral over $|y_{11}| = N^{-1}$ in a counter-clockwise direction.

To integrate out the Grassmannians we expand the exponential to second order, use the relation (8.30) between σ and τ, and use that for 2×2 matrices X, Y which are constant on the diagonal ($x_{11} = x_{22}, y_{11} = y_{22}$) we have the identities

$$\begin{aligned}
\frac{N^2}{8}\partial_\sigma \operatorname{Tr}^2(\sigma Z\tau Z^* X + Z\sigma Z^*\tau X) &= -4N^2|z|^2(\Re z)^2 \det(X) \\
&\quad + N^2|z|^2(\Im z)^2 \operatorname{Tr}^2(X) \\
\frac{N^2}{8}\partial_\sigma \operatorname{Tr}^2(\sigma\tau X) &= -N^2 \det(X) \\
\frac{N^2}{4}\partial_\sigma \operatorname{Tr}(\sigma\tau X) \operatorname{Tr}\Big(\sigma Z\tau Z^* Y + Z\sigma Z^*\tau Y\Big) &= 2N^2(\Re z)^2(\operatorname{Tr}(XY) - \operatorname{Tr}(X)\operatorname{Tr}(Y)) \\
&\quad + 2\mathrm{i}N^2(\Im z)(\Re z)(X_{12}Y_{21} - X_{21}Y_{12}). \\
\frac{N}{2}\partial_\sigma \operatorname{Tr}(\sigma\tau X\sigma\tau Y) &= N(\operatorname{Tr}(X)\operatorname{Tr}(Y) - \operatorname{Tr}(XY)) \\
\frac{N}{2}\partial_\sigma \operatorname{Tr} \sigma\tau X\Big(\sigma Z\tau Z^* X + Z\sigma Z^*\tau X\Big) &= 4N(\Re z)^2 \det(X).
\end{aligned}$$

Whence we finally have the expression

$$\begin{aligned}
G = -N^2\Big[&\det\Big(A(1-C') - (1-\frac{1}{N})B\Big) + (4|z|^2(\Re z)^2 + 2(\Re z)^2(2 - \operatorname{Tr} C'))\det A \\
&- |z|^2(\Im z)^2 \operatorname{Tr}^2 A - 2(\Re z)^2(1-\frac{1}{n})\Big(\operatorname{Tr} A \operatorname{Tr} B - \operatorname{Tr} AB\Big) \\
&- 2(\Re z)^2(\Im z)^2 \det A^2 \det(1+y)(4\det(x) - \operatorname{Tr}^2 x)\Big] \\
+ N\Big(&\det(A)\Big(1 + 4(\Re z)^2 - \operatorname{Tr} C'\Big) - (1-\frac{1}{N})\det(B)\Big).
\end{aligned} \tag{8.32}$$

We now rewrite (8.31) by using the parametrizations

$$x = \begin{pmatrix} a & a\sqrt{1-\tau}e^{\mathrm{i}\varphi} \\ a\sqrt{1-\tau}e^{-\mathrm{i}\varphi} & a \end{pmatrix}, \qquad y = \begin{pmatrix} \xi & 0 \\ 0 & \xi \end{pmatrix}, \tag{8.33}$$

with $a \in \mathrm{i}\mathbf{R}_+$, $\tau \in [0,1]$, $\varphi \in [0,2\pi]$ and $|\xi| = N^{-1}$. Since the integral over $\varphi \in [0,2\pi]$ is equal to 2π as a consequence of the fact that the functions f, g, G_N defined below do not depend on φ, we have that

$$\mathbf{E}\operatorname{Tr}[Y-w]^{-1} = \frac{N}{4\pi\mathrm{i}}\oint \mathrm{d}\xi \int_0^{\mathrm{i}\infty} \mathrm{d}a \int_0^1 \mathrm{d}\tau \frac{\xi^2 a}{\tau^{1/2}} e^{N[f(\xi) - g(a,\tau,\eta)]} G_N(a,\tau,\xi,z), \tag{8.34}$$

where, using the notation $\eta := \Im z$, the functions f and g are defined by

$$f(\xi) := -w\xi + \log(1+\xi) - \log\xi - \frac{|z|^2}{1+\xi}, \tag{8.35}$$

$$\begin{aligned}
g(a,\tau,\eta) := &-wa + \frac{1}{2}\log[1 + 2a + a^2\tau] - \log a - \frac{1}{2}\log\tau \\
&- \frac{|z|^2(1+a) - 2\eta^2 a^2(1-\tau)}{1 + 2a + a^2\tau}.
\end{aligned} \tag{8.36}$$

Note that $g(a,1,\eta) = f(a)$; in particular, we remark that $g(a,1,\eta)$ is independent of η for any $a \in \mathbf{C}$. Furthermore, using the parameterizations in (8.33) the function $G_N := G_{1,N} + G_{2,N}$ is given by

$$\begin{aligned} G_{1,N} &= \Big(N^2 \frac{p_{2,0,0}}{a^2\xi^2(\xi+1)^2\tau} - N\frac{p_{1,0,0}}{a^2\xi^2(\xi+1)\tau} + \delta N^2 \frac{p_{2,0,1}}{a\xi(\xi+1)^2\tau} - N\delta\frac{p_{1,0,1}}{a\xi(\xi+1)\tau} \\ &\quad + N^2\delta^2\frac{p_{2,0,2}}{(\xi+1)^2}\Big) \times \Big((a^2\tau+2a+1)^2(\xi+1)^2\Big)^{-1}, \\ G_{2,N} &= \Big(N^2\eta^2\frac{p_{2,2,0}}{a\xi(\xi+1)^3\tau} - N\eta^2\frac{p_{1,2,0}}{a\xi\tau} + N^2\eta^2\delta\frac{p_{2,2,1}}{(\xi+1)}\Big) \\ &\quad \times \Big((a^2\tau+2a+1)^2(\xi+1)^2\Big)^{-1}, \end{aligned} \tag{8.37}$$

where $p_{i,j,k} = p_{i,j,k}(a,\tau,\xi)$ are explicit polynomials in a,τ,ξ which we defer to Appendix 8.B, $\eta := \Im z$ and $\delta := 1-|z|^2$. The indices i,j,k in the definition of $p_{i,j,k}$ denote the N, η and δ power, respectively. We split G_N as the sum of $G_{1,N}$ and $G_{2,N}$ since $G_{1,N}$ depends only on $|z|$, whilst $G_{2,N}$ depends explicitly by $\eta = \Im[z]$, hence $G_{2,N} = 0$ if $z \in \mathbf{R}$.

8.4 Asymptotic analysis in the complex case for the saddle point regime

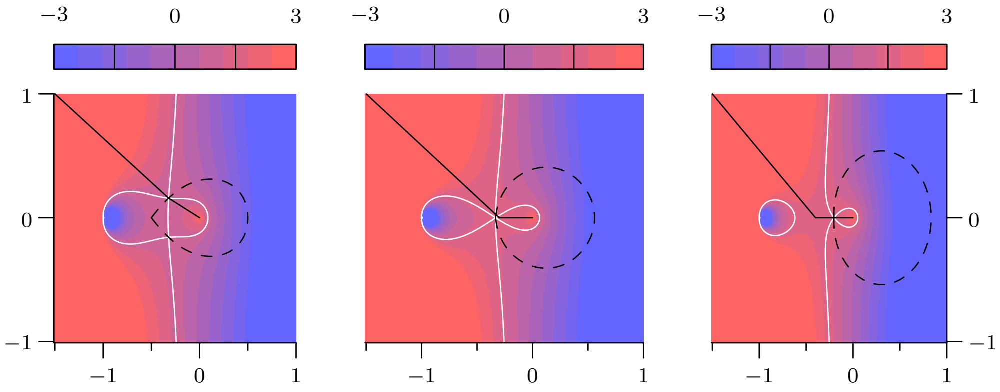
For the density of states ϱ_{Y^z} on the positive semi-axis $E > 0$ we expect a singular behaviour for $E \approx 0$ and a square-root edge for $E \approx \mathfrak{e}_+$. The singularity at $E = 0$ exhibits a phase transition in δ at 0; for $\delta > 0$ the transition is between an $E^{-1/3}$–singularity for $\delta = 0$ and a $\delta^{1/2}E^{-1/2}$–singularity for $0 < \delta \leq 1$, while for $\delta < 0$ the transition is between the $E^{-1/3}$–singularity for $\delta = 0$ and square-root edge in $\mathfrak{e}_- \sim |\delta|^3$ of slope $|\delta|^{-5/2}$. We now analyse the location of the critical point(s) x_*, i.e. the solutions to $f'(x_*) = 0$, as well as the asymptotics of the phase function f around them precisely in all of the above regimes. For the saddle-point approximation the second derivative $f''(x_*)$ is of particular importance and we find that it can only vanish in the vicinity of $E \approx \mathfrak{e}_+$ and $E \approx \mathfrak{e}_- \vee 0$, and otherwise satisfies $|f''(x_*)| \gtrsim 1$.

The saddle point equation $f'(x_*) = 0$ leads to the simple cubic equation

$$wx_*^3 + 2wx_*^2 + wx_* + \delta x_* + 1 = 0,$$

which is precisely the MDE equation from (8.11), whose explicit solution via Cardano's formula reveals that for $E \in (\mathfrak{e}_-, \mathfrak{e}_+)$ there are two relevant critical points $x_*, \overline{x_*}$ with $\Re f(x_*) = \Re f(\overline{x_*})$, while for $E \geq \mathfrak{e}_+$ or $0 \leq E \leq \mathfrak{e}_-$ there is one relevant critical point x_*, where x_* is given by

$$\begin{aligned} x_* &= \begin{cases} e^{-2i\pi/3}\sqrt[3]{q+\sqrt{q^2+p^3}} + e^{2i\pi/3}\sqrt[3]{q-\sqrt{q^2+p^3}} - \frac{2}{3}, & E \leq \mathfrak{e}_- \\ e^{2i\pi/3}\sqrt[3]{q+\sqrt{q^2+p^3}} + e^{-2i\pi/3}\sqrt[3]{q-\sqrt{q^2+p^3}} - \frac{2}{3}, & \mathfrak{e}_- \leq E \leq \mathfrak{e}_+ \\ \sqrt[3]{q+\sqrt{q^2+p^3}} + \sqrt[3]{q-\sqrt{q^2+p^3}} - \frac{2}{3}, & E \geq \mathfrak{e}_+ \end{cases} \\ q &:= \frac{\delta}{3E} + \frac{1}{27} - \frac{1}{2E}, \quad p := \frac{\delta}{3E} - \frac{1}{9}, \end{aligned} \tag{8.38}$$



((a)) $E_+ < 0$ ((b)) $|E_+| = 0$ ((c)) $E_+ > 0$

Figure 8.3: Contour plot of $\Re f(x)$ in the regime $E \approx \mathfrak{e}_+$. The solid white lines represent the level set $\Re f(x) = \Re f(x_*)$, while the solid and dashed black lines represent the chosen contours for the x- and y-integrations, respectively.

where $q^2 + p^3 > 0$ as long as $E \in (\mathfrak{e}_-, \mathfrak{e}_+)$ and $q^2 + p^3 < 0$ for $E > \mathfrak{e}_+$ or $E < \mathfrak{e}_-$. Here we chose the branch of the cubic root such that $\sqrt[3]{\mathbf{R}} = \mathbf{R}$ and that $\sqrt[3]{z}$ for $z \in \mathbf{C} \setminus \mathbf{R}$ is the cubic root with the maximal real part. Note that the choice of the cubic root implies $\Im x_* \geq 0$ and $x_* = x_*(E) = m^z(E + \mathrm{i}0)$, where m^z has been defined in (8.11).

Before concluding this section with the proof of Proposition 8.2.2, we collect certain asymptotics of the critical point x_* and the phase function f in its vicinity, which will also be used in the main estimates of the present paper in Sections 8.5–8.6. In the edges $\mathfrak{e}_\pm$ the critical points have the simple expressions

$$x_*(\mathfrak{e}_\pm) = -\frac{2}{3 \pm \sqrt{9 - 8\delta}}$$

and satisfy $x_*(\mathfrak{e}_+) \sim -1$ and $x_*(\mathfrak{e}_-) = -3/(2\delta)[1 + \mathcal{O}(|\delta|)]$. Elementary expansions of (8.38) for E near the edges reveal the following asymptotics of x_* in the various regimes.

Regime $E \approx \mathfrak{e}_+$

Close to the spectral edge $E \approx \mathfrak{e}_+$ we have the asymptotic expansion

$$x_* = x_*(\mathfrak{e}_+) + \gamma_+ \sqrt{E_+}\Big(1 + \mathcal{O}(|E_+^{1/2}|)\Big) \tag{8.39a}$$

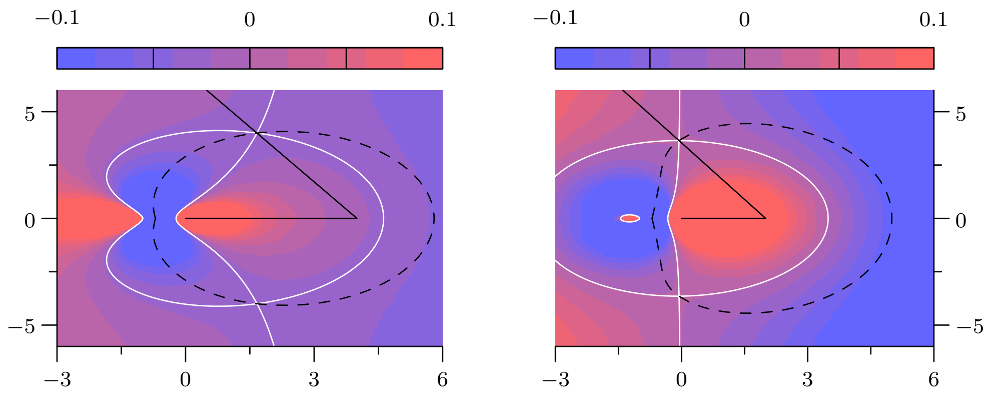
in $E_+ := E - \mathfrak{e}_+$, where γ_+ was defined in (8.18b). The location of saddle point(s) in the regime $E \approx \mathfrak{e}_+$ is depicted in Figure 8.3. The second derivative of f is asymptotically given by

$$f''(x_*) = \frac{2\sqrt{E_+}}{\gamma_+}\Big(1 + \mathcal{O}(|E_+^{1/2}|)\Big). \tag{8.39b}$$

Regime $E \approx 0$ in the case $\delta \geq 0$

For $E \approx 0$ we have the asymptotic expansions

$$x_* = E^{-1/3}\Psi\Big(\frac{\delta}{E^{1/3}}\Big)\Big[1 + \mathcal{O}(\delta + E^{1/3})\Big], \tag{8.40a}$$



((a)) $0 < \delta \ll E^{1/3} \ll 1$ ((b)) $0 < E^{1/3} \ll \delta$

FIGURE 8.4: Contour plot of $\Re f(x)$ for $\delta > 0$ in the regime $E \approx 0$. The solid white lines represent the level set $\Re f(x) = \Re f(x_*)$, while the solid and dashed black lines represent the chosen contours for the x- and y-integrations, respectively.

where $\Psi(\lambda)$ is the unique solution to the cubic equation

$$1 + \lambda\Psi(\lambda) + \Psi(\lambda)^3 = 0, \quad \Re\Psi(\lambda) > 0, \quad \Im\Psi(\lambda) > 0, \quad \lambda \geq 0.$$

The explicit function $\Psi(\lambda)$ has the asymptotics

$$\lim_{\lambda \searrow 0} \Psi(\lambda) = \Psi(0) = e^{\mathrm{i}\pi/3}, \qquad \lim_{\lambda\to\infty} \frac{\Psi(\lambda)}{\sqrt{\lambda}} = \mathrm{i}.$$

Thus it follows that

$$x_* = \Big(\mathrm{i}\sqrt{\frac{\delta}{E}} - 1 + \frac{1}{2\delta}\Big)\Big(1 + \mathcal{O}(\frac{E}{\delta^3})\Big) = \Big(\frac{e^{\mathrm{i}\pi/3}}{E^{1/3}} - \frac{2}{3}\Big)\Big(1 + \mathcal{O}(E^{2/3} + \frac{\delta}{E^{1/3}})\Big), \tag{8.40b}$$

where the first expansion is informative in the $E \ll \delta^3$, and the second one in the $E \gg \delta^3$ regime. The location of saddle point(s) in the regime $E \approx 0$ is depicted in Figure 8.4. For the second derivative we have the expansions

$$f''(x_*) = 3e^{2\mathrm{i}\pi/3}E^{4/3}\Big(1 + \mathcal{O}(E + \frac{\delta}{E^{1/3}})\Big) = 2\mathrm{i}\frac{E^{3/2}}{\delta^{1/2}}\Big(1 + \mathcal{O}(\frac{E}{\delta^3})\Big) \tag{8.40c}$$

and similarly for higher derivatives, $|f^{(k)}(x_*)| \sim E^{(2+k)/3} \wedge E^{(k+1)/2}\delta^{-(k-1)/2}$ for $k \geq 3$.

Regime $E \approx \mathfrak{e}_-$ in the case $\delta < 0$

Around the spectral edge $\mathfrak{e}_-$ the critical point admits the asymptotic expansion

$$\begin{aligned} x_* &= x_*(\mathfrak{e}_-) + \gamma_-\Big(1 + \mathcal{O}(\frac{|E|^{1/2}}{\delta^{3/2}})\Big)\begin{cases} \mathrm{i}\sqrt{|E_-|}, & E_- \geq 0 \\ -\sqrt{|E_-|}, & E_- \leq 0 \end{cases} \\ &= \frac{1}{E^{1/3}}\Big(e^{\mathrm{i}\pi/3} + \frac{\mathrm{i}}{3}e^{\mathrm{i}\pi/3}\frac{\delta}{E^{1/3}}(1 + \mathcal{O}(|E|^{1/3})) + \mathcal{O}(\frac{|\delta|^2}{E^{2/3}})\Big), \end{aligned} \tag{8.41a}$$

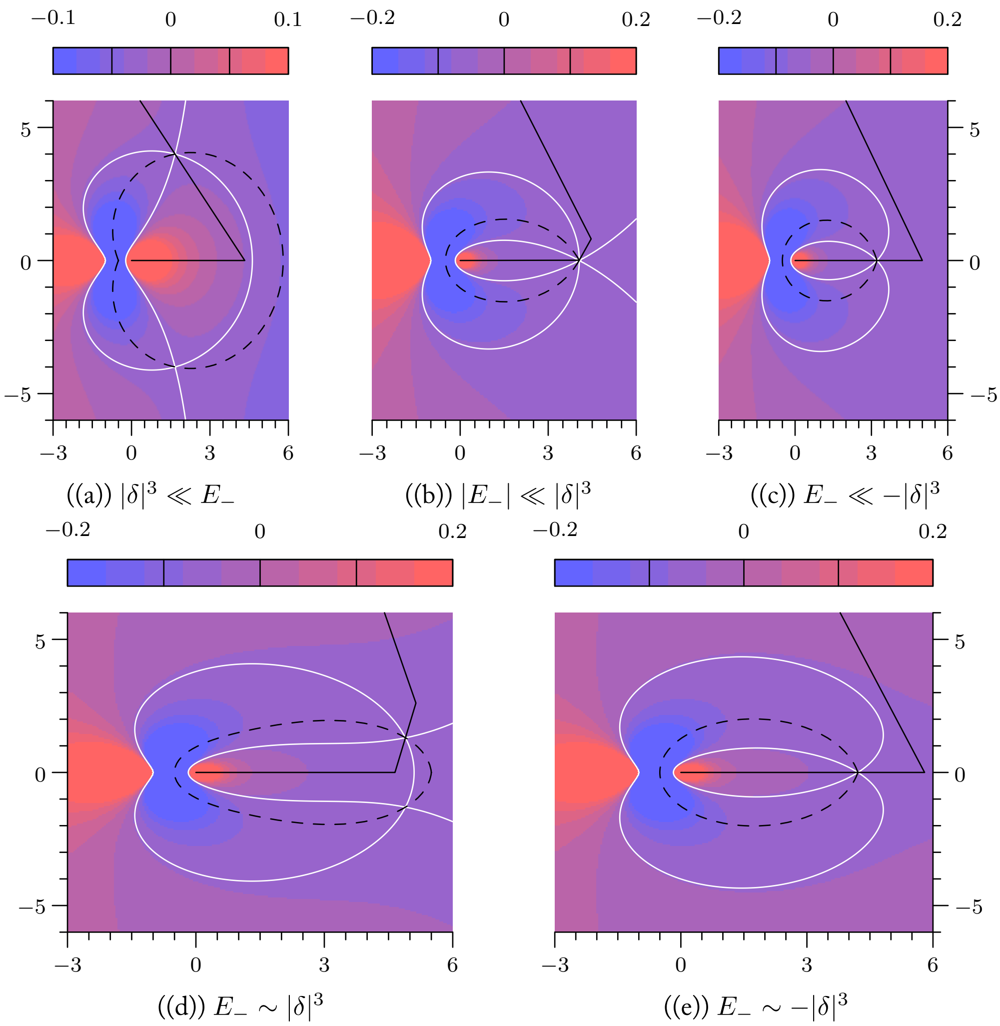


Figure 8.5: Contour plot of $\Re f(x)$ in the regime $E \approx \mathfrak{e}_-$. The solid white lines represent the level set $\Re f(x) = \Re f(x_*)$, while the solid and dashed black lines represent the chosen contours for the x- and y-integrations, respectively.

where $E_- := E - \mathfrak{e}_-$, and these separate expansions are relevant in the $|E| \ll |\delta|^3$ and $|E| \gg |\delta|^3$ regimes, respectively. The location of saddle point(s) in the regime $E \approx \mathfrak{e}_-$ is depicted in Figure 8.5. The second derivative around x_* is given by

$$\begin{aligned} f''(x_*) &= \frac{2}{\gamma_-}\Big(1 + \mathcal{O}(|E_-|^{1/2}|\delta|^{-3/2})\Big) \times \begin{cases} \sqrt{|E_-|}, & E_- \le 0 \\ -\mathrm{i}\sqrt{|E_-|}, & E_- \ge 0 \end{cases} \\ &= 3e^{2\mathrm{i}\pi/3}E^{4/3}\Big(1 + \mathcal{O}(E + \frac{|\delta|}{E^{1/3}})\Big), \end{aligned} \tag{8.41b}$$

with $\gamma_- \sim |\delta|^{-5/2}$.

Proof of Proposition 8.2.2. As the functions f and G in (8.28) are meromorphic we are free to deform the contours for the x- and y-integrals as long as we are not crossing 0 or -1 and the x-contour goes out from 0 in the ”right” direction (in the region $\Re[x] < 0, \Im[x] > 0$

in Figure 8.3, and in the region $\Re[x] > |\Im[x]|$ in Figures 8.4–8.5). It is easy to see that the contours can always be deformed in such a way that $\Re f(x) > \Re f(x_*) = \Re f(\overline{x_*})$ and $\Re f(y) < \Re f(x_*)$ for all $x, y \neq x_*, \overline{x_*}$, see Figures 8.3–8.5 for an illustration of the chosen contours.

We now compute the integral (8.28) in the large N limit when E is near the edges. In certain regimes of the parameters N, E and δ a saddle point analysis is applicable after a suitable contour deformation. In most cases, the result is a point evaluation of the integrand at the saddle points. In some transition regimes of the parameters the saddle point analysis only allows us to explicitly scale out some combination of the parameters and leaving an integral depending only on a reduced set of rescaled parameters.

We recall the classical quadratic saddle point approximation for holomorphic functions $f(z), g(z)$ such that $f(z)$ has a unique critical point in some z_*, and that γ can be deformed to go through z_* in such a way that $\Re f(z) < \Re f(z_*)$ for all $\gamma \ni z \neq z_*$. Then for large $\lambda \gg 1$ the saddle point approximation is given by

$$\int_\gamma g(z) e^{\lambda f(z)} \, \mathrm{d}z = \pm g(z_*) e^{\lambda f(z_*)} \sqrt{\frac{2\pi}{\lambda |f''(z_*)|}} \mathrm{i} e^{-\frac{\mathrm{i}}{2} \arg f''(z_*)} \Big(1 + \mathcal{O}\big(\frac{1}{\lambda}\big) \Big), \tag{8.42}$$

where $\pm$ is determined by the direction of γ through z_* with $+$ corresponding to the direction parallel to $\mathrm{i} \exp(-\frac{\mathrm{i}}{2} \arg f''(z_*))$. This formula is applicable, i.e. we can use point evaluation in the *saddle point regime*, whenever the lengthscale $\ell_f \sim (N|f''(x_*)|)^{-1/2}$ of the exponential decay from the quadratic approximation of the phase function is much smaller than the scale $\ell_g \sim |g(x_*)|/|\nabla g(x_*)|$ on which g is essentially unchanged. For our integral (8.28) we thus need to check the condition

$$\frac{1}{\sqrt{N|f''(x_*)|}} \ll \left| \frac{\nabla(yG(x,y))}{yG(x,y)} \Big|_{(x,y)=(x_*,x_*)} \right|^{-1} \tag{8.43}$$

in all regimes separately.

In the regime $E \approx \mathfrak{e}_+$, using the asymptotics for x_* from (8.39a) and (8.39b), the quadratic saddle point approximation is valid if

$$\frac{1}{\sqrt{N|E_+|^{1/2}}} \ll |E_+|^{1/2},$$

i.e. if $|E_+| \gg N^{-2/3}$. Here the length-scale $|E_+|^{1/2}$ represents the length-scale on which $(x,y) \mapsto yG(x,y)$ is essentially constant which can be obtained by explicitly computing the log-derivative

$$\left| \frac{\nabla(yG(x,y))}{yG(x,y)} \Big|_{(x,y)=(x_*,x_*)} \right| \sim |E_+|^{-1/2}.$$

Similar calculations yield that for $E \approx 0$ and $\delta \geq 0$ the quadratic saddle point approximation is valid if

$$\frac{1}{\sqrt{N(E^{4/3} \wedge E^{3/2}\delta^{-1/2})}} \ll E^{-1/3} \vee \delta^{1/2} E^{-1/2}, \quad \text{i.e.} \quad E \gg N^{-3/2} \wedge N^{-2}\delta^{-1},$$

while for $E \approx \mathfrak{e}_-$ and $\delta < 0$ the condition (8.43) reads

$$\frac{1}{\sqrt{N(E^{4/3} \vee |E_-|^{1/2}|\delta|^{5/2})}} \ll E^{-1/3} \wedge |E_-|^{1/2}|\delta|^{-5/2}, \quad \text{i.e.} \quad |E_-| \gg N^{-2/3}|\delta|^{5/3},$$

recalling that $E = E_- + \mathfrak{e}_-$ and $\mathfrak{e}_- \sim \delta^3$ from (8.18b).

In these regimes we can thus apply (8.42) to (8.28) and using that

$$G(x_*, x_*) = f''(x_*), \qquad G(x_*, \overline{x_*}) = 0,$$

as follows from explicit computations, we thus finally conclude (8.14). Here the error terms in (8.14) follow from (8.42) by choosing $\lambda \sim (\ell_g/\ell_f)^2$ according to asymptotics of the second derivatives and log-derivatives above. More precisely, for example in the second case $E \approx 0$ and $E^{1/3} \gg \delta > 0$, the phase function f is approximately given by

$$f(x) \approx E^{2/3}\Big[\frac{1}{2(E^{1/3}x)} - E^{1/3}x\Big],$$

while G can asymptotically be written as

$$yG(x,y) \approx x_*G(x_*,x_*)+3E^{4/3}e^{2\mathrm{i}\pi/3}\Big(2(x-x_*)+(y-x_*)\Big)+\mathcal{O}(E^{5/3}\Big(|x-x_*|^2+|y-x_*|^2\Big)).$$

Thus we make the change of variables $x = x_* + E^{-1/3}x'$, $y = x_* + E^{-1/3}y'$ to find

$$\begin{aligned}&\frac{N^2}{2\pi\mathrm{i}}E^{-2/3}\int \mathrm{d}x' \int \mathrm{d}y' e^{-NE^{-2/3}f''(x_*)\Big(\frac{x'^2}{2}-\frac{y'^2}{2}\Big)-NE^{-1}f'''(x_*)\Big(\frac{x'^2}{6}-\frac{y'^2}{6}\Big)+\mathcal{O}(NE^{2/3}(|x'|^4+|y'|^4))}\\ &\qquad\qquad \times \Big(x_*G(x_*,x_*) + 3Ee^{2\mathrm{i}\pi/3}(y'+2x') + \mathcal{O}(E(|x'|^2+|y'|^2))\Big)\\ &\quad = x_*\Big(1+\mathcal{O}(\frac{1}{NE^{2/3}})\Big),\end{aligned}$$

where we used that $G(x_*, x_*) \sim E^{4/3}$. The other cases in (8.14) can be checked similarly. □

8.5 Derivation of the 1-point function in the critical regime for the complex case

In this section we prove Theorem 8.2.1, i.e. we study $\mathbf{E}\,\mathrm{Tr}[Y-w]^{-1}$, with $w = E + \mathrm{i}\epsilon$, $1 \gg |E| \gg \epsilon > 0$, for E so close to 0 such that $|E|$ is smaller or comparable with the eigenvalues scaling around 0. We will first consider the case $\delta \geq 0$ and afterwards explain the necessary changes in the regime $-CN^{-1/2} \leq \delta < 0$.

8.5.1 Case $0 \leq \delta \leq 1$.

In the following of this section we assume that $E > 0$, since we are interested in the computations of (8.28) for $E = \Re[w]$ inside the spectrum of Y. In order to study the transition between the local law regime, that is considered in Section 8.4, and the regime when the main contribution to (8.28) comes from the smallest eigenvalue of Y, we define the parameter

$$c(N) = c(N,\delta) := \frac{1}{N^{3/2}} \wedge \frac{1}{\delta N^2}. \tag{8.44}$$

In particular, in the regime $E \gg c(N)$ the double integral in (8.28) is computed by saddle point analysis in (8.14), i.e. the main contribution comes from the regime around the stationary point x_* of f, with x_* defined in (8.40b), whilst for $E \lesssim c(N)$ the main contribution to (8.28) comes from a larger regime around the stationary point x_*. From now on we assume that $E \lesssim c(N)$. In the following we denote the leading order of the stationary point x_* by

$$z_* = z_*(E, \delta) := E^{-1/3}\Psi(\delta E^{-1/3}), \tag{8.45}$$

where $\Psi(\lambda)$ was defined in (8.40a) and has the asymptotics $\Psi(0) = e^{\mathrm{i}\pi/3}$ and $\Psi(\lambda)/\sqrt{\lambda} \to \mathrm{i}$ as $\lambda \to \infty$. Note that $|z_*| \gg 1$ for any $E \ll 1$, $0 \le \delta \le 1$. For this reason, we expect that the main contribution to the double integral in (8.28) comes from the regime when $|x|$ and $|y|$ are both large, say $|x|, |y| \ge N^\rho$, for some small fixed $0 < \rho < 1/2$. Later on in this section, see Lemma 8.5.4, we prove that the contribution to (8.28) in the regime when either $|x|$ or $|y|$ are smaller than N^ρ is exponentially small. In order to get the asymptotics in (8.13a), is not affordable to estimate the error terms in the Taylor expansion by absolute value. In particular, it is not affordable to estimate the integral of $e^{Nf(y)}$ over Γ by absolute value, hence the improved bound in (8.54) is needed. To make our writing easier, for any $R \in \mathbf{N}$, $R \ge 2$, we introduce the notation

$$\mathcal{O}^\#(x^{-R}) := \{g \in \mathcal{P}_R\}, \quad \mathcal{O}^\#((x,y)^{-R}) := \{g \in \mathcal{Q}_R\}, \tag{8.46}$$

where $\mathcal{P}_R$ and $\mathcal{Q}_R$ are defined as Laurent series of order at least R around infinity, i.e.

$$\mathcal{P}_R := \{g \colon \mathbf{C} \to \mathbf{C} | g(x) = \sum_{\alpha \ge R} \frac{\tilde{c}_\alpha}{x^\alpha},\ |\tilde{c}_\alpha| \le C^\alpha,\ \text{if } |x| \ge 2C\},$$

$$\mathcal{Q}_R := \{\tilde{g} \colon \mathbf{C} \times \mathbf{C} \to \mathbf{C} | \tilde{g}(x,y) = \sum_{\alpha,\beta \ge 1, \alpha+\beta \ge R} \frac{c_{\alpha,\beta}}{x^\alpha y^\beta},\ \text{with } |c_{\alpha,\beta}| \le C^{\alpha+\beta},\ \text{if } |x|,|y| \ge 2C\},$$

for some constant $C > 0$ that is implicit in the $\mathcal{O}^\#$ notation. Here α, β are integer exponents. Note that $\mathcal{O}^\#(|x|^{-R}) = \mathcal{O}(|x|^{-R})$ for any $x \in \mathbf{C}$. Then, we expand the phase function f for large argument as follows

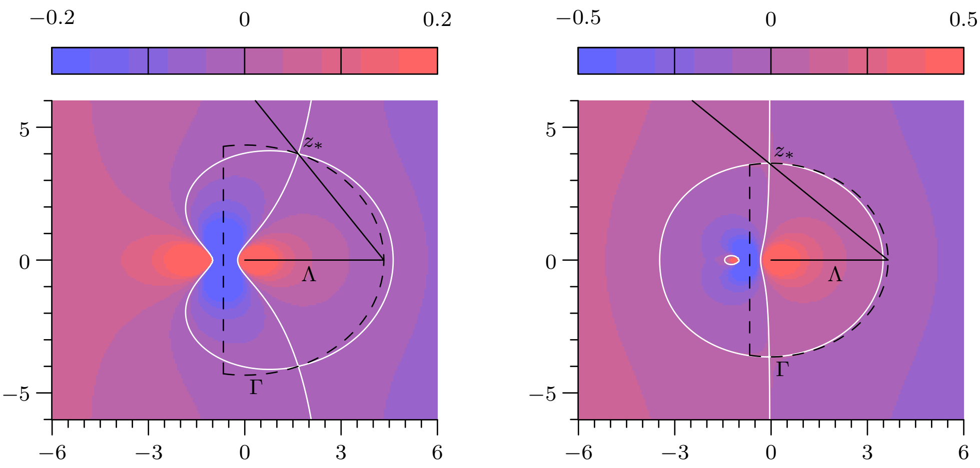
$$f(x) = g(x) + \mathcal{O}^\#(x^{-3} + \delta x^{-2}), \quad g(x) := -(E + \mathrm{i}\epsilon)x + \frac{\delta}{x} + \frac{1}{2x^2}. \tag{8.47a}$$

and for large x and y we expand G as

$$\begin{aligned} G(x,y) &= H(x,y) + \mathcal{O}^\#\left((x,y)^{-5} + \delta(x,y)^{-4}\right), \\ H(x,y) &:= \frac{1}{x^3 y} + \frac{1}{x^2 y^2} + \frac{1}{x y^3} + \frac{\delta}{x y^2} + \frac{\delta}{x^2 y}. \end{aligned} \tag{8.47b}$$

In order to compute the integral in (8.28) we deform the contours Λ and Γ through z_*, with z_* defined in (8.45). In particular, we are allowed to deform the contours as long as the x-contour goes out from zero in region $\Re[x] > |\Im[x]|$, it ends in the region $\Re[x] < 0, \Im[x] > 0$, and it does not cross 0 and -1 along the deformation; the y-contour, instead, can be freely deformed as long as it does not cross 0 and -1. Hence, we can deform the y-contour as $\Gamma = \Gamma_{z_*} := \Gamma_{1,z_*} \cup \Gamma_{2,z_*}$, where

$$\Gamma_{1,z_*} := \left\{-\frac{2}{3} + \mathrm{i}t : 0 \le |t| \le \sqrt{|z_*|^2 - \frac{4}{9}}\right\}, \quad \Gamma_{2,z_*} := \left\{|z^*|e^{\mathrm{i}\psi} : \psi \in [-\psi_{z_*}, \psi_{z_*}]\right\}, \tag{8.48a}$$



((a)) $0 < \delta \ll E^{1/3} \ll 1$ ((b)) $0 < E^{1/3} \ll \delta$

Figure 8.6: Illustration of the contours (8.48a)–(8.48b) together with the phase diagram of $\Re f$, where the white line represents the level set $\Re f(x) = \Re f(z_*)$. Note that the precise choice of the contours is only important close to 0 and for very large $|x|$ as otherwise the phase function is small.

with $\psi_{z_*} = \arccos[-2/(3|z_*|)]$, and the x-contour as $\Lambda = \Lambda_{z_*} := \Lambda_{1,z_*} \cup \Lambda_{2,z_*}$, with

$$\Lambda_{1,z_*} := [0, |z_*|), \quad \Lambda_{2,z_*} := \{|z_*| - qs + \mathrm{i}s : \ s \in [0, +\infty)\}, \tag{8.48b}$$

where

$$q = q_{z_*} := \Im[z_*]^{-1}(|z_*| - \Re[z_*]). \tag{8.48c}$$

Note that $q \sim 1$ uniformly in N, E and δ, since $\Re[z_*] \lesssim \Im[z_*]$ for any $E \ll 1$, $0 \leq \delta \leq 1$. We assume the convention that the orientation of Γ is counter clockwise. See Figure 8.6 for an illustration of Γ and Λ.

Before proceeding with the computation of the leading term of (8.28), in the following lemma we state some properties of the function f on the contours Γ, Λ. Using that $\epsilon \ll E$, the proof of the lemma below follows by easy computations.

Lemma 8.5.1. *Let f be the phase function defined in* (8.28)*, then the following properties hold true:*

(i) For any $y = -2/3 + \mathrm{i}t \in \Gamma_{1,z_*}$, we have that

$$\Re[f(-2/3 + \mathrm{i}t)] = \frac{2E}{3} + \epsilon t - \frac{1}{2t^2} + \mathcal{O}(|t|^{-3} + \delta|t|^{-2}), \tag{8.49}$$

and

$$\Im[f(-2/3 + \mathrm{i}t)] = \frac{2\epsilon}{3} - Et - \frac{\delta}{t} + \mathcal{O}(|t|^{-3}). \tag{8.50}$$

(ii) For $\epsilon = 0$, the function $t \mapsto \Re[f(-2/3 + \mathrm{i}t)]$ on Γ_{1,z_*} is strictly increasing if $t > 0$ and strictly decreasing if $t < 0$.

(iii) The function $x \mapsto \Re[f(x)]$ is strictly decreasing on Λ_{1,z_*}.

(iv) Let $x \in \Lambda_{2,z_*}$ be parametrized as $x = |z_*| - qs + \mathrm{i}s$, for $s \in [0, +\infty)$, with q defined in (8.48c)*, then*

$$\Re[f(x)] = -E(|z_*| - qs) + \epsilon s + \frac{\delta(|z_*| - qs)}{s^2 + (qs - |z_*|)^2} + \frac{(1-2\delta)[(|z_*| - qs)^2 - s^2]}{(s^2 + (qs - |z_*|)^2)^2} + \mathcal{O}\left([s^2 + |z_*|^2]^{-3/2}\right). \tag{8.51}$$

Despite the fact that saddle point analysis is not useful anymore in this regime, we expect that the main contribution to (8.28) comes from the regime in the double integral when both x and y are large, i.e. $|x|, |y| \geq N^\rho$. For this purpose we define

$$\widetilde{\Lambda} := \{x \in \Lambda_{1,z_*} : |x| \leq N^\rho\}, \quad \widetilde{\Gamma} := \{y \in \Gamma_{1,z_*} : |y| \leq N^\rho\}. \tag{8.52}$$

In the following part of this section we will firstly prove that the contribution to (8.28) in the regime when either $x \in \widetilde{\Lambda}$ or $y \in \widetilde{\Gamma}$ is exponentially small and then we explicitly compute the leading term of (8.28) in the regime $(x, y) \in (\Lambda \setminus \widetilde{\Lambda}) \times (\Gamma \setminus \widetilde{\Gamma})$. For this purpose, we first prove a bound for the double integral in the regime $y \in \Gamma \setminus \widetilde{\Gamma}$ or $x \in \Lambda \setminus \widetilde{\Lambda}$ in Lemma 8.5.2 and Lemma 8.5.3, respectively, and then we conclude the estimate for $x \in \widetilde{\Lambda}$ or $y \in \widetilde{\Gamma}$ in Lemma 8.5.4. Finally, in Theorem 8.2.1 we consider the regime $(x, y) \in (\Lambda \setminus \widetilde{\Lambda}) \times (\Gamma \setminus \widetilde{\Gamma})$ and compute the leading term of (8.28).

Lemma 8.5.2. *Let $c(N)$ be defined in* (8.44)*, $E \lesssim c(N)$, $b \in \mathbf{N}$, and let f be defined in* (8.28)*, then*

$$\left| \int_{\Gamma \setminus \widetilde{\Gamma}} \frac{e^{Nf(y)}}{y^b} \, \mathrm{d}y \right| \lesssim |z_*|^{1-b} + \begin{cases} |z_*|, & b = 0, \\ 1 + |\log(N|z_*|^{-2})|, & b = 1,\ \delta < |z_*|^{-1}, \\ 1 + |\log(N\delta|z_*|^{-1})|, & b = 1,\ \delta \geq |z_*|^{-1}, \\ N^{\frac{1-b}{2}} \wedge (N\delta)^{1-b}, & b \geq 2, \end{cases} \tag{8.53}$$

where Γ, $\widetilde{\Gamma}$ are defined in (8.48a) *and* (8.52) *respectively. Furthermore, we have that*

$$\int_{\Gamma \setminus \widetilde{\Gamma}} e^{Nf(y)} \, \mathrm{d}y = \mathcal{O}(N^{1/2} \vee (N\delta)), \qquad \int_{\Gamma \setminus \widetilde{\Gamma}} \frac{e^{Nf(y)}}{y} \, \mathrm{d}y = \mathcal{O}(1). \tag{8.54}$$

Proof. Firstly, we notice that if $y \in \Gamma \setminus \widetilde{\Gamma}$ then $|y| \geq N^\rho$, hence we expand f as in (8.47a), i.e.

$$f(y) = -(E + \mathrm{i}\epsilon)y + \frac{\delta}{y} + \frac{1}{2y^2} + \mathcal{O}(|y|^{-3} + \delta|y|^{-2}). \tag{8.55}$$

Moreover, by (8.45) it follows that $|z_*| \sim E^{-1/3} \vee \sqrt{\delta E^{-1}}$, and so that

$$|Nf(y)| \lesssim NE^{2/3} + N\sqrt{\delta E} \tag{8.56}$$

for any $|y| \sim |z_*|$, $E \lesssim c(N)$. Note that $\Gamma \setminus \widetilde{\Gamma} = (\Gamma_{1,z_*} \setminus \widetilde{\Gamma}) \cup \Gamma_{2,z_*}$, with Γ_{1,z_*}, Γ_{2,z_*} defined in (8.48a). By (8.56) it easily follows that $|Nf(y)| \lesssim 1$ for any $y \in \Gamma_{2,z_*}$, which clearly implies that

$$\int_{\Gamma_{2,z_*}} \left| \frac{e^{Nf(y)}}{y^b} \right| |\,\mathrm{d}y| \lesssim |z_*|^{1-b}. \tag{8.57}$$

To conclude the proof of (8.53) we bound the integral on $\Gamma_{1,z_*} \setminus \widetilde{\Gamma}$. Let $w = E + \mathrm{i}\epsilon$, then in this regime, by (8.49)–(8.50), we have

$$\int_{\Gamma_{1,z_*}\setminus\widetilde{\Gamma}} \frac{e^{Nf(y)}}{y^b}\,\mathrm{d}y = -\mathrm{i}\int_{N^\rho}^{\sqrt{|z_*|^2-4/9}} e^{-N\left[\frac{1}{2t^2}+\mathcal{O}(|t|^{-3}+\delta|t|^{-2})\right]} \times\left(\frac{e^{-N\left[\mathrm{i}wt+\mathrm{i}\frac{\delta}{t}\right]}}{(-2/3+\mathrm{i}t)^b}+\frac{e^{N\left[\mathrm{i}wt+\mathrm{i}\frac{\delta}{t}\right]}}{(-2/3-\mathrm{i}t)^b}\right)(1+\mathcal{O}(NE))\,\mathrm{d}t. \tag{8.58}$$

For any $b\in\mathbf{N}$, we estimate the integral above as follows

$$\left|\int_{\Gamma_{1,z_*}\setminus\widetilde{\Gamma}} \frac{e^{Nf(y)}}{y^b}\,\mathrm{d}y\right| \lesssim \left|\int_{N^\rho}^{|z_*|} \frac{e^{-\frac{N}{2t^2}+\frac{\mathrm{i}\delta N}{t}}}{t^b}\,\mathrm{d}t\right| \lesssim \begin{cases} |z_*|, & b=0,\\ 1+|\log(N|z_*|^{-2})|, & b=1,\ \delta<|z_*|^{-1},\\ 1+|\log(N\delta|z_*|^{-1})|, & b=1,\ \delta\ge|z_*|^{-1},\\ N^{\frac{1-b}{2}}\wedge(N\delta)^{1-b}, & b\ge 2,\end{cases} \tag{8.59}$$

Note that in (8.59) for $b=0,1$ we get the bound in the r.h.s. bringing the absolute value inside the integral, whilst this is not affordable to get the bound for $b\ge 2$, since the term $e^{\mathrm{i}\delta N/t}$ has to be used. Indeed, we would get a bound $N^{(1-b)/2}$ for $b\ge 2$ if we estimate the integral in (8.59) moving the absolute value inside. In the following part of the proof we compute the integral (8.58) for $b=0,1$ without estimating it by absolute value.

In particular, for $b=1$, we prove that the leading term of the r.h.s. of (8.58) is $\mathcal{O}(1)$, instead of the overestimate $1+|\log(N|z_*|^{-2})|$ in (8.59), as a consequence of the symmetry of Γ_{1,z_*} respect to 0. For this computation we have to distinguish the cases $\delta\gg N^{-1/2}$ and $\delta\lesssim N^{-1/2}$. If $E\sim c(N)$ and $\delta\lesssim N^{-1/2}$, then $N|z_*|^{-2}\sim 1$, hence the bound in (8.54) directly follows by (8.57) and (8.59). We are left with the cases $E\ll c(N)$ and $E\sim c(N)$, $\delta\gg N^{-1/2}$. For $\delta\gg N^{-1/2}$, we have $|z_*|\sim\sqrt{\delta/E}$ and using $|Nwt|\le NE|z_*|\ll 1$, if $E\ll c(N)$, and $|Nwt|\sim 1$, if $E\sim c(N)$, we conclude

$$\int_{N^\rho}^{\sqrt{|z_*|^2-4/9}} \frac{e^{-N\left[\frac{1}{2t^2}\pm\mathrm{i}\frac{\delta}{t}\pm\mathrm{i}tw\right]}}{-2/3+\mathrm{i}t}\,\mathrm{d}t = \int_{N^\rho}^{|z_*|}\frac{e^{-\frac{N}{2t^2}\pm\mathrm{i}\frac{\delta}{t}}}{t}\,\mathrm{d}t+\mathcal{O}(1) = |\log(N\delta|z_*|^{-1})|+\mathcal{O}(1).$$

Using similar computations to above, we prove that the integral in the l.h.s. of the above equalities is equal to $|\log(N|z_*|^{-2})|+\mathcal{O}(1)$ if $\delta\ll N^{-1/2}$. Similar calculation holds if the denominator is $(-2/3-\mathrm{i}t)$ instead of $(-2/3+\mathrm{i}t)$, just an overall sign changes. Thus the leading terms from the two parts of the integral in (8.58) cancel each other. We thus conclude the second bound in (8.54) combining the above computations with (8.57) and (8.58).

Next, we compute the integral of $e^{Nf(y)}$ on $\Gamma\setminus\widetilde{\Gamma}$, i.e. we prove the first bound in (8.54). We consider only the regime $E\ll c(N)$, since in the regime $E\sim c(N)$ the bound in (8.54) follows directly by (8.57),(8.59), and the definition of $c(N)$ in (8.44), since $|z_*|\sim N^{1/2}\vee N\delta$. On Γ_{2,z_*}, using the parametrization $y=|z_*|e^{\mathrm{i}\psi}$, and that by (8.56) we have $|Nf(y)|\ll 1$ for $E\ll c(N)$, we Taylor expand $e^{Nf(y)}$ and conclude that

$$\int_{\Gamma_{2,z_*}} e^{Nf(y)}\,\mathrm{d}y = 2|z_*|\mathrm{i}\cdot\left[1+\mathcal{O}\left(NE|z_*|+\delta N|_*|^{-1}+N|z_*|^{-2}\right)\right]. \tag{8.60}$$

Furthermore, by (8.58) for $b = 0$, using that $E \ll c(N)$ and so that $|Nwt| \ll 1$ on Γ_{1,z_*}, we have

$$\int_{\Gamma_{1,z_*}\setminus\widetilde{\Gamma}} e^{Nf(y)}\,\mathrm{d}y = -2\mathrm{i}|z_*| + \mathcal{O}(N^{1/2} \vee N\delta).$$

The minus sign is due to the counter clockwise orientation of Γ, i.e. the vertical line Γ_{1,z_*} is parametrized from the top to the bottom. Combining this computation with (8.60) and using that $NE|z_*|^2+N\delta+N|z_*|^{-1} \ll N^{1/2}+N\delta$, since $|z_*| \sim E^{-1/3}+\sqrt{\delta E^{-1}}$ by (8.45), we conclude the proof of this lemma. □

Lemma 8.5.3. *Let* $c(N)$ *be defined in* (8.44), $E \lesssim c(N)$, *let* f *be defined in* (8.28) *and* $a \in \mathbf{R}$, *then the following bound holds true*

$$\int_{\Lambda\setminus\widetilde{\Lambda}} \left|\frac{e^{-Nf(x)}}{x^a}\right| |\,\mathrm{d}x| \lesssim \begin{cases} |z_*|^{1-a} + (NE)^{a-1}, & a<1, \\ 1+|\log(N|z_*|^{-2})|, & a=1, \delta<|z_*|^{-1}, \\ 1+|\log(N\delta|z_*|^{-1})|, & a=1, \delta \geq |z_*|^{-1}, \\ N^{\frac{1-a}{2}} \wedge (N\delta)^{1-a}, & a>1, \end{cases} \tag{8.61}$$

where Λ, $\widetilde{\Lambda}$ *are defined in* (8.48b) *and* (8.52).

Proof. We split the computation of the integral of $e^{Nf(x)}x^{-a}$ as the sum of the integral over $\Lambda_{1,z_*} \setminus \widetilde{\Lambda}$ and Λ_{2,z_*}. Using the parametrization $x = |z_*| - qs + \mathrm{i}s$, with $s \in [0,+\infty)$ and q defined in (8.48c), by (8.51), we estimate the integral over Λ_{2,z_*} as follows

$$\int_{\Lambda_{2,z_*}} \left|\frac{e^{-Nf(x)}}{x^a}\right| |\,\mathrm{d}x| \lesssim \int_0^{+\infty} \frac{e^{-N\left[-E(|z_*|-qs)+\frac{\delta(|z_*|-qs)}{(|z_*|-qs)^2+s^2}+\frac{(|z_*|-qs)^2-s^2}{2[(|z_*|-qs)^2+s^2]^2}\right]}}{[(|z_*|-qs)^2+s^2]^{a/2}}\,\mathrm{d}s. \tag{8.62}$$

We split the computation of the integral in the r.h.s. of (8.62) into two parts: $|s| \in [0,|z_*|)$ and $s \in [|z_*|,+\infty)$. Since $q \sim 1$ and $NE|z_*| \lesssim 1$, in the regime $|s| \in [0,|z_*|)$ we estimate the integral in the r.h.s. of (8.62) as

$$e^{NE|z_*|}\int_0^{|z_*|} \frac{e^{-\frac{N\delta}{|z_*|}-\frac{N}{|z_*|^2}}}{|z_*|^a}\,\mathrm{d}s \lesssim |z_*|^{1-a}. \tag{8.63}$$

In the regime $s \in [|z_*|,+\infty)$, instead, we have

$$e^{NE|z_*|}\int_{|z_*|}^{+\infty} \frac{e^{-NEqs}}{s^a}\,\mathrm{d}s \lesssim \begin{cases} (NE)^{a-1}, & a<1, \\ 1+|\log(NE|z_*|)|, & a=1, \\ |z_*|^{1-a}, & a>1. \end{cases} \tag{8.64}$$

We are left with the estimate of the integral over $\Lambda_{1,z_*} \setminus \widetilde{\Lambda}$. Similarly to the bound in (8.59), using that $\Lambda_{1,z_*} \setminus \widetilde{\Lambda} = [N^\rho, |z_*|)$ and that $NE|z_*| \lesssim 1$, we have that

$$\begin{aligned} \int_{\Lambda_{1,z_*}\setminus\widetilde{\Lambda}} \left|\frac{e^{-Nf(x)}}{x^a}\right| |\,\mathrm{d}x| &\lesssim \int_{N^\rho}^{|z_*|} \frac{e^{-\frac{N\delta}{s}-\frac{N}{2s^2}}}{s^a}\,\mathrm{d}s \\ &\lesssim \begin{cases} |z_*|^{1-a}, & a<1, \\ 1+|\log(N|z_*|^{-2})|, & a=1, \delta<|z_*|^{-1}, \\ 1+|\log(N\delta|z_*|^{-1})|, & a=1, \delta \geq |z_*|^{-1}, \\ N^{\frac{1-a}{2}} \wedge (N\delta)^{1-a}, & a>1, \end{cases} \end{aligned} \tag{8.65}$$

Combining (8.62)–(8.65) we conclude the proof of (8.61). □

Using Lemma 8.5.2, Lemma 8.5.3 in the following lemma we prove that the contribution to (8.28) in the regime where either $x \in \widetilde{\Lambda}$ or $y \in \widetilde{\Gamma}$ is exponentially small.

Lemma 8.5.4. *Let $c(N)$ be defined in (8.44), and let f, G be defined in (8.28), then, as $\epsilon \to 0^+$, for any $E \lesssim c(N)$ we have that*

$$\left| \left(\int_\Lambda \mathrm{d}x \int_\Gamma \mathrm{d}y - \int_{\Lambda \setminus \widetilde{\Lambda}} \mathrm{d}x \int_{\Gamma \setminus \widetilde{\Gamma}} \mathrm{d}y \right) \left[e^{N[f(y)-f(x)]} y G(x,y) \right] \right| \\ \lesssim N^\rho (N^{1/2} + N\delta + |\log(NE^{2/3})|) e^{-\frac{1}{2}N^{1-2\rho}}. \tag{8.66}$$

Proof. We split the estimate of the integral over $(\Lambda \times \Gamma) \setminus [(\Lambda \setminus \widetilde{\Lambda}) \times (\Gamma \setminus \widetilde{\Gamma})]$ into three regimes: $(x,y) \in \widetilde{\Lambda} \times \widetilde{\Gamma}$, $(x,y) \in \widetilde{\Lambda} \times (\Gamma \setminus \widetilde{\Gamma})$, $(x,y) \in (\Lambda \setminus \widetilde{\Lambda}) \times \widetilde{\Gamma}$. By ii, iii of Lemma 8.5.1, in the regime $y \in \widetilde{\Gamma}$ and $x \in \widetilde{\Lambda}$, respectively, it follows that the function f attains its maximum at $y = -2/3 \pm \mathrm{i}N^\rho$ on $\widetilde{\Gamma}$ and f attains its minimum at $x = N^\rho$ on $\widetilde{\Lambda}$. Hence, by the expansion in (8.47a) it follows that

$$\sup_{y \in \widetilde{\Gamma}} |e^{Nf(y)}| + \sup_{x \in \widetilde{\Gamma}} |e^{-Nf(x)}| \lesssim e^{-Nf(N^\rho)}, \tag{8.67}$$

with

$$f(N^\rho) = \frac{\delta}{N^\rho} + \frac{1}{2N^{2\rho}} + \mathcal{O}(N^{-3\rho} + \delta N^{-2\rho}). \tag{8.68}$$

Then, by (8.67) and (8.68), it follows that the integral over $(x,y) \in \widetilde{\Lambda} \times \widetilde{\Gamma}$ is bounded by $N^{2\rho} e^{-N^{1-2\rho}}$. Note that in the regimes $(x,y) \in \widetilde{\Lambda} \times (\Gamma \setminus \widetilde{\Gamma})$ and $(x,y) \in (\Lambda \setminus \widetilde{\Lambda}) \times \widetilde{\Gamma}$ one among $|x|$ and $|y|$ is bigger than N^ρ. Hence, expanding (8.28) for large x or y argument, using Lemma 8.5.2, Lemma 8.5.3 to estimate the regime $x \in \widetilde{\Lambda}$ and $y \in \widetilde{\Gamma}$, respectively, by (8.67)–(8.68), we conclude that the integral over $(x,y) \in \widetilde{\Lambda} \times (\Gamma \setminus \widetilde{\Gamma})$ is bounded by $N^\rho(N^{1/2} + (N\delta))e^{-N^{1-2\rho}/2}$, and that the one over $(x,y) \in (\Lambda \setminus \widetilde{\Lambda}) \times \widetilde{\Gamma}$ is bounded by $N^\rho(1 + |\log(NE^{2/3})|)e^{-N^{1-2\rho}/2}$. □

Next, we compute the leading term of (8.28). We define $\widetilde{z}_*$ as

$$\widetilde{z}_*(\lambda, \widetilde{\delta}) := N^{-1/2} |z_*(\lambda c(N), N^{-1/2}\widetilde{\delta})|, \qquad \lambda = \frac{E}{c(N)}, \quad \widetilde{\delta} = N^{1/2}\delta, \tag{8.69}$$

where we also recalled the rescaled parameters λ and $\widetilde{\delta}$ from (8.12). Note that $\widetilde{z}_*(\lambda, \widetilde{\delta})$ is N-independent, indeed all N factors scale out by using the definition of $z_*(E, \delta)$ from (8.45). Since $E \ll 1$ and $\delta \in [0,1]$, by (8.69) it follows that the range of the new parameters is $\widetilde{\delta} \le N^{1/2}$ and $\lambda \ll c(N)^{-1}$.

We are now ready to prove our main result on the leading term of (8.28), denoted by $q_{\widetilde{\delta}}(\lambda)$, in the complex case, Theorem 8.2.1. Then, the one point function of Y is asymptotically given by $p_{\widetilde{\delta}}(\lambda) := \Im[q_{\widetilde{\delta}}(\lambda)]$. The main inputs for the proof are the bounds in (8.53) and (8.61) that will be used to estimate the error terms in the expansions for large arguments of f and G in (8.47a) and (8.47b).

Proof of Theorem 8.2.1 in the case $\delta \geq 0$. By (8.28) and Lemma 8.5.4 it follows that

$$\begin{aligned} \mathbf{E}\,\mathrm{Tr}[Y-w]^{-1} &= \frac{N^2}{2\pi \mathrm{i}} \int_{\Lambda\setminus\widetilde{\Lambda}} \mathrm{d}x \int_{\Gamma\setminus\widetilde{\Gamma}} \mathrm{d}y e^{-Nf(x)+Nf(y)} y \cdot G(x,y) \\ &\quad + \mathcal{O}\left(N^\rho (N^{1/2} + N\delta + |\log(NE^{2/3})|) e^{-\frac{1}{2}N^{1-2\rho}}\right). \end{aligned} \tag{8.70}$$

Note that $|x|, |y| \geq N^\rho$ for any $(x,y) \in (\Lambda \setminus \widetilde{\Lambda}) \times (\Gamma \setminus \widetilde{\Gamma})$. In order to prove (8.13a) we first estimate the error terms in the expansions of f and G in (8.47a)–(8.47b) and then in order to get an N-independent double integral we rescale the phase function by $|z_*|$. By Lemma 8.5.2 and Lemma 8.5.3, using that $|\log(N|z_*|^{-2})| + |\log(N\delta|z_*|^{-1})| \lesssim |\log(NE^{2/3})|$ by the definition of z_* in (8.45), it follows that

$$\begin{aligned} &\left| \int_{\Lambda\setminus\widetilde{\Lambda}} \mathrm{d}x \int_{\Gamma\setminus\widetilde{\Gamma}} \mathrm{d}y \frac{e^{-Nf(x)+Nf(y)}}{x^a y^b} \right| \\ &\quad \lesssim \begin{cases} N^{\frac{2-d}{2}} (1 \wedge \widetilde{\delta}^{1-d})(1 \vee \widetilde{\delta}), & b = 0, \\ N^{\frac{1-b}{2}} (1 \wedge \widetilde{\delta}^{1-b})(1 + |\log(NE^{2/3})|), & a = 1,\ b \geq 1, \\ N^{\frac{2-d}{2}} (1 \wedge \widetilde{\delta}^{2-d}), & a > 1,\ b \geq 1, \end{cases} \end{aligned} \tag{8.71}$$

for any $a \geq 1$, $b \in \mathbf{N}$, where $d := a + b$. In order to get the bound in the r.h.s. of (8.71) we estimated the terms with $b = 0$ and $b = 1$ using the improved bound in (8.54), all the other terms are estimated by absolute value. Note that for $\lambda \ll 1$ the bound in (8.71) and the definition of λ in (8.69) imply that

$$\lim_{\epsilon \to 0^+} |\mathbf{E}\,\mathrm{Tr}[Y - (E + \mathrm{i}\epsilon)]^{-1}| \lesssim N^{3/2}(1 \vee \widetilde{\delta}) \begin{cases} |\log \lambda|, & \lambda \geq \widetilde{\delta}^3, \\ |\log \lambda \widetilde{\delta}|, & \lambda < \widetilde{\delta}^3. \end{cases}$$

if $\lambda \ll 1$, since the leading term in the expansion of $yG(x,y)$ in (8.47b) consists of monomials of the form $x^{-a}y^{-b}$, with $a + b = 3$, and $\delta x^{-a} y^{-b}$, with $a + b = 2$. This concludes the proof of (8.13c).

Now we prove the more precise asymptotics (8.13a). We will replace the functions f and G in (8.70) by their leading order approximations, denoted by g and H from (8.47a) and (8.47b). The error of this replacement in the phase function f is estimated by the Taylor expanding the exponent $e^{\mathcal{O}^\#(x^{-3}+\delta x^{-2})}$, with $\mathcal{O}^\#(x^{-3} + \delta x^{-2})$ defined in (8.46). Hence, by (8.47a) and (8.47b) and the bound in (8.71), as $\epsilon \to 0^+$, we conclude that

$$\begin{aligned} \mathbf{E}\,\mathrm{Tr}[Y-w]^{-1} &= \frac{N^2}{2\pi \mathrm{i}} \int_{\Lambda\setminus\widetilde{\Lambda}} \mathrm{d}x \int_{\Gamma\setminus\widetilde{\Gamma}} \mathrm{d}y e^{-Ng(x)+Ng(y)} H(x,y) \\ &\quad + \mathcal{O}\left((N + N^{3/2}\delta)[1 + |\log(NE^{2/3})|]\right). \end{aligned} \tag{8.72}$$

The error estimates in (8.72) come from terms with $d \geq 4$ or terms with $d \geq 3$ multiplied by δ in (8.71).

We recall that $\lambda = Ec(N)^{-1}$, $\widetilde{\delta} = \delta N^{1/2}$, and that $|z_*| = N^{1/2}\widetilde{z}_*(\lambda, \widetilde{\delta})$. Then, defining the contours

$$\widehat{\Gamma} := |z_*|^{-1}\Gamma, \quad \widehat{\Lambda} := |z_*|^{-1}\Lambda, \tag{8.73}$$

and using the change of variables $x \to x|z_*|$, $y \to y|z_*|$ in the leading term of (8.72) we conclude that

$$\begin{aligned}\mathbf{E}\,\mathrm{Tr}[Y-w]^{-1} = \frac{N^{3/2}\widetilde{z}_*(\lambda,\widetilde{\delta})^{-1}}{2\pi\mathrm{i}} \int_{\widehat{\Gamma}} \mathrm{d}y \int_{\widehat{\Lambda}} \mathrm{d}x e^{h_{\lambda,\widetilde{\delta}}(y)-h_{\lambda,\widetilde{\delta}}(x)} \widetilde{H}_{\lambda,\widetilde{\delta}}(x,y) \\ + \mathcal{O}(N(1\vee\widetilde{\delta})[1+|\log\lambda|])\end{aligned} \tag{8.74}$$

with $h_{\lambda,\widetilde{\delta}}(x)$ and $\widetilde{H}_{\lambda,\widetilde{\delta}}(x,y)$ defined in (8.13b). Note that in order to get (8.74) we used that the integral in the regime when either $x \in [0, N^\rho|z_*|^{-1}]$ or $y \in [-2|z_*|^{-1}/3, -2|z_*|^{-1}/3 + \mathrm{i}N^\rho|z_*|^{-1}]$ is exponentially small. Moreover, since by holomorphicity we can deform the contour $\widehat{\Lambda}$ to any contour, which does not cross -1, from 0 to $e^{\frac{3\mathrm{i}\pi}{4}}\infty$ and we can deform the contour $\widehat{\Gamma}$ as long as it does not cross 0, (8.74) concludes the proof of Theorem 8.2.1. □

8.5.2 Case $\delta < 0$, $|\delta| \lesssim N^{-1/2}$.

We now explain the necessary changes in the case $\delta < 0$. All along this section we assume that $E \lesssim N^{-3/2}$. Let x_* be the stationary point of f defined in (8.41a), that is the point around where the main contribution to (8.28) comes from in the saddle point regime for $\delta < 0$. Then, at leading order, x_* is given by $3|\delta|^{-1}/2$ if $E \ll |\delta|^3$, by $e^{\frac{\pi\mathrm{i}}{3}}E^{-1/3}$ if $E \gg |\delta|^3$, and by $\mu(c)e^{\frac{\pi\mathrm{i}}{3}}E^{-1/3}$ if $E = c|\delta|^3$, for some function $\mu(c) > 0$ for any fixed constant $c > 0$ independent of N, E and δ.

This regime can be treated similarly to the regime $0 \le \delta \le 1$, since for $|\delta| \lesssim N^{-1/2}$ the term δx^{-1} in the expansion of f, for $|x| \gg 1$, does not play any role in the bounds of Lemma 8.5.2, Lemma 8.5.3. Indeed, instead of the deforming the contours Γ and Λ through the leading term of the stationary point x_*, we deform Γ and Λ through $z_* := e^{\frac{\pi\mathrm{i}}{3}}E^{-1/3}$ as $\Gamma = \Gamma_{z_*} := \Gamma_{1,z_*} \cup \Gamma_{2,z_*}$ and $\Lambda = \Lambda_{z_*} := \Lambda_{1,z_*} \cup \Lambda_{2,z_*}$, where $\Gamma_{1,z_*}, \Gamma_{2,z_*}$ and $\Lambda_{1,z_*}, \Lambda_{2,z_*}$ are defined in (8.48a) and (8.48b), respectively. We could have done the same choice in the case $0 \le \delta \lesssim N^{-1/2}$, but not for the regime $N^{-1/2} \ll \delta \le 1$, hence, to treat both the regimes in the same way, in Section 8.5.1 we deformed the contours trough (8.45). Note that z_* defined here is not the analogue of (8.45), since in all cases $z_* = e^{\frac{\pi\mathrm{i}}{3}}E^{-1/3}$. The fact that $E \ll 1$ implies that $|z_*| \gg 1$, hence, similarly to the case $0 \le \delta \le 1$, we expect that the main contribution to (8.28) comes from the regime when $|x|, |y| \ge N^\rho$, for some small $0 < \rho < 1/2$. Hence, in order to compute the leading term of (8.28) we expand f and G for large arguments as in (8.47a) and (8.47b).

The phase function f defined in (8.28) satisfies the properties i, ii and iv of Lemma 8.5.1, but iii does not hold true for $\delta < 0$ if $E \ll |\delta|^3$. Instead, it is easy to see that the following lemma holds true.

Lemma 8.5.5. *Let f be the phase function defined in* (8.28)*, then, as $\epsilon \to 0^+$, the function $x \mapsto \Re[f(x)]$ has a unique global minimum on Λ_{1,z_*} at $x = 3|\delta|^{-1}/2$ if $E \ll |\delta|^3$.*

Note that, since $|\delta| \lesssim N^{-1/2}$, by Lemma 8.5.5 it follows that the function $x \mapsto \Re[f(x)]$ is strictly decreasing for $0 \le x \ll N^{-1/2}$.

Proof of Theorem 8.2.1 for $-CN^{-1/2} \le \delta < 0$. Let $\widetilde{\Gamma}, \widetilde{\Lambda}$ be defined in (8.52), then using that

$$e^{-N\left[\frac{\delta}{s}+\frac{1}{2s^2}\right]} \lesssim e^{-\frac{N}{4s^2}},$$

for $s \in [N^\rho, |z_*|]$ and $|\delta| \lesssim N^{-1/2}$, Lemma 8.5.3 and the improved bounds in (8.54) of Lemma 8.5.2, for $b = 0, 1$, we conclude the bound in the following lemma exactly as in (8.71) without the improvement involving $\widetilde{\delta}$.

Lemma 8.5.6. *Let $E \lesssim N^{-3/2}$, and let f be defined in* (8.28)*. Then, the following bound holds true*

$$\left| \int_{\Lambda \setminus \widetilde{\Lambda}} \mathrm{d}x \int_{\Gamma \setminus \widetilde{\Gamma}} \mathrm{d}y \frac{e^{-Nf(x)+Nf(y)}}{x^a y^b} \right| \lesssim \begin{cases} N^{\frac{2-d}{2}}, & b = 0, \\ N^{\frac{1-b}{2}}(1 + |\log(NE^{2/3})|), & a = 1,\, b \geq 1, \\ N^{\frac{2-d}{2}}, & a > 1,\, b \geq 1, \end{cases}$$

for any $a \geq 1$ and $b \in \mathbf{N}$, where $d := a + b$.

Then, similarly to the case $0 \leq \delta \leq 1$, by Lemma 8.5.6 we conclude that the contribution to (8.28) of the regime when either $x \in \widetilde{\Lambda}$ or $y \in \widetilde{\Gamma}$ is exponentially small, i.e. Lemma 8.5.4 holds true. Hence, by (8.28), Lemma 8.5.6 and the expansion of G in (8.47b), we easily conclude Theorem 8.2.1 also in the regime $-CN^{-1/2} \leq \delta < 0$. □

8.6 The real case below the saddle point regime

In this section we prove Theorem 8.2.3. Throughout this section we always assume that $\Re w < 0$, hence to make our notation easier we define $w = -E + \mathrm{i}\epsilon$ with some $E > 0$ and $\epsilon > 0$. Moreover, we always assume that $E \lesssim c(N)$, with $c(N)$ defined in (8.44). We are interested in estimating (8.34) in the transitional regime of $|z|$ around one. For this purpose we introduce the parameter $\delta = \delta_z := 1 - |z|^2$. In order to have an optimal estimate of the leading order term of (8.34) it is not affordable to estimate the error terms in the expansions, for large a and ξ, of f, $g(\cdot, 1, \eta)$ and $G_{1,N}, G_{2,N}$ by absolute value. For this reason, to compute the error terms in the expansions of f, $g(\cdot, 0, \eta), G_{1,N}, G_{2,N}$ we use a notation $\mathcal{O}^\#(\cdot)$ similar to the one introduced in (8.46). In order to keep track of the power of τ in the expansion of $G_{1,N}$ and $G_{2,N}$, we define the set of functions

$$\mathcal{O}^\#((a,\tau,\xi)^{-1}) := \Bigg\{ h : \mathbf{C} \times [0,1] \times \mathbf{C} \to \mathbf{C} :$$
$$h(a,\tau,\xi) = \sum_{\substack{\alpha+\beta \geq 1,\\ 0 \leq \gamma \leq \alpha}} \frac{c_{\alpha,\beta,\gamma}}{a^\alpha \tau^\gamma \xi^\beta} \text{ with, } |c_{\alpha,\beta,\gamma}| \leq C^{\alpha+\beta}, \text{ if} |a|, |a\tau|, |\xi| \geq 2C \Bigg\},$$

for some constant $C > 0$ implicit in the $\mathcal{O}^\#(\cdot)$ notation. The exponents α, β are non negative integers.

We expand the functions $f, g(\cdot, 1, \eta), G_{1,N}, G_{2,N}$ for large a and ξ arguments as

$$\begin{aligned} f(\xi) &= (E - \mathrm{i}\epsilon)\xi + \frac{\delta}{\xi} + \frac{1}{2\xi^2} + \mathcal{O}^\#(\xi^{-3} + \delta\xi^{-2}), \\ g(a,1,\eta) &= (E - \mathrm{i}\epsilon)a + \frac{\delta}{a} + \frac{1}{2a^2} + \mathcal{O}(a^{-3} + \delta a^{-2}), \end{aligned} \tag{8.75}$$

with $\mathcal{O}^{\#}(\cdot)$ defined (8.46), and

$$G_{1,N}(a,\tau,\xi,|z|) = \Bigg[\sum_{\substack{a,\beta\geq 2,\, \alpha+\beta=8,\\ \gamma=\min\{\alpha-1,3\}}} \frac{c_{\alpha,\beta,\gamma}N^2}{a^\alpha\tau^\gamma\xi^\beta} + \sum_{\substack{\alpha,\beta\geq 2,\, \alpha+\beta=7,\\ \gamma=\min\{\alpha-1,3\}}} \frac{c_{\alpha,\beta,\gamma}N^2\delta}{a^\alpha\tau^\gamma\xi^\beta} \tag{8.76}$$

$$+ \sum_{\substack{a,\beta\geq 2,\, \alpha+\beta=6,\\ \gamma=\min\{\alpha-1,2\}}} \frac{c_{\alpha,\beta,\gamma}N}{a^\alpha\tau^\gamma\xi^\beta} + \sum_{\substack{\alpha,\beta\geq 2,\alpha+\beta=6\\ \gamma=\min\{\alpha-1,2\}}} \frac{c_{\alpha,\beta,\gamma}N^2\delta^2}{a^\alpha\tau^\gamma\xi^\beta}\Bigg]$$

$$\times\left[1+\mathcal{O}^{\#}\left((a,\tau,\xi)^{-1}\right)\right], \tag{8.77}$$

$$G_{2,N}(a,\tau,\xi,z) = \Bigg[\sum_{\substack{\alpha,\beta\geq 2,\, \alpha+\beta=6,\\ \gamma=\max\{\alpha-1,2\}}} \frac{c_{\alpha,\beta,\gamma}N^2\eta^2}{a^\alpha\tau^\gamma\xi^\beta} + \sum_{\substack{\alpha,\beta\geq 2,\alpha+\beta=5,\\ \gamma=\max\{\alpha-1,2\}}} \frac{c_{\alpha,\beta,\gamma}N^2\eta^2\delta}{a^\alpha\tau^\gamma\xi^\beta}$$

$$+ \sum_{\substack{\alpha,\beta=2,3,\, \alpha+\beta=5\\ \gamma=\max\{\alpha-1,2\}}} \frac{c_{\alpha,\beta,\gamma}N\eta^2}{a^\alpha\tau^\gamma\xi^\beta}\Bigg] \times \left[1+\mathcal{O}^{\#}\left((a,\tau,\xi)^{-1}\right)\right], \tag{8.78}$$

where $c_{\alpha,\beta,\gamma}\in\mathbf{R}$ is a constant that may change term by term. To make our notation easier in (8.76)-(8.78) we used the convention to write a common multiplicative error for all the terms, even if in principle the constants in the series expansion of the error terms differ term by term.

In the following we deform the integration contours in (8.34) with the following constraints: the ξ contour can be freely deformed as long as it does not cross 0 and -1, the a-contour can be deformed as long as it goes out from zero in the region $\Re[a]>|\Im[a]|$, it ends in the region $\Re[a]>0$, and it does not cross 0 and -1 along the deformation. The τ-contour will not be deformed.

Specifically, we deform the a-contour in (8.34) to $\Lambda=[0,+\infty)$, and we can deform the ξ-contour to any contour around 0 not encircling -1 (this contour is denoted by Γ in (8.79)). Moreover, since for $a\in\mathbf{R}_+$ we have $|e^{-N(E-\mathrm{i}\epsilon)a}|=e^{-NEa}$ and since the factor e^{-NEa} makes the integral convergent, we may pass to the limit $\epsilon\to 0^+$. Hence, for any $E>0$ we conclude that

$$\mathbf{E}\,\mathrm{Tr}[Y+E]^{-1} = \frac{N}{4\pi\mathrm{i}}\int_\Gamma \mathrm{d}\xi\int_0^{+\infty}\mathrm{d}a\int_0^1\mathrm{d}s\,\frac{\xi^2 a}{\tau^{1/2}}e^{N[f(\xi)-g(a,\tau,\eta)]}G_N(a,\tau,\xi,z). \tag{8.79}$$

We split the computation of the leading order term of (8.79) into the cases $-CN^{-1/2}\leq\delta<0$ and $\delta\geq 0$.

8.6.1 Case $0\leq\delta\leq 1$.

In order to estimate the leading term of (8.79) we compute the ξ-integral and the (a,τ)-integral separately. In particular, we compute the (a,τ)-integral firstly performing the τ-integral for any fixed a and then we compute the a-integral. Note that, since $E'=-E$ is negative, the relevant stationary point of $f(\xi)$ is real and its leading order ξ_* is given by

$$\xi_* := \begin{cases}\sqrt{\frac{\delta}{E}}, & E\ll\delta^3,\\ \mu(c)E^{-1/3}, & E=c\delta^3,\\ E^{-1/3}, & E\gg\delta^3,\end{cases} \tag{8.80}$$

for some $\mu(c) > 0$ and any fixed constant $c > 0$ independent of N, E and δ. Note that $\xi_* \gg 1$ for any $E \lesssim c(N)$, $0 \leq \delta \leq 1$. We will show that in the regime $a \in [N^\rho, +\infty)$ the τ-integral is concentrated around 1 as long as $|e^{-N[g(a,\tau,\eta)-g(a,1,\eta)]}|$ is effective, i.e. as long as $N|g(a,\tau,\eta) - g(a,1,\eta)| \gg 1$, and that it is concentrated around 0 if $N|g(a,\tau,\eta) - g(a,1,\eta)| \lesssim 1$. For this purpose, using that $g(a,1,\eta) = f(a)$ for any $a \in \mathbf{C}$, we rewrite (8.79) as

$$\mathbf{E}\,\mathrm{Tr}[Y+E]^{-1} = \frac{N}{4\pi \mathrm{i}} \int_\Gamma \mathrm{d}\xi e^{Nf(\xi)} \xi^2 \int_0^{+\infty} \mathrm{d}a e^{-Nf(a)} a \int_0^1 \mathrm{d}\tau \frac{e^{-N[g(a,\tau,\eta)-g(a,1,\eta)]}}{\tau^{1/2}} G_N, \tag{8.81}$$

where we used that by holomorphicity, we can deform the contour Γ as $\Gamma = \Gamma_{\xi_*} := \Gamma_{1,\xi_*} \cup \Gamma_{2,\xi_*}$, with $\Gamma_{1,\xi_*}, \Gamma_{2,\xi_*}$ defined in (8.48a) replacing z_* by ξ_*. We will show that the contribution to (8.81) of the integrals in the regime when either $|a| \leq N^\rho$ or $|\xi| \leq N^\rho$, for some small fixed $0 < \rho < 1/2$, is exponentially small. Moreover, we will show that also the τ-integral is exponentially small for τ very close to 0 because of the term $\log \tau$ in the phase function $g(a,\tau,\eta)$. Hence, we define $\widetilde{\Gamma}$, $\widetilde{\Lambda}$ as in (8.52), and $I \subset [0,1]$ as $I = I_a := [0, N^{\rho/2} a^{-1}]$, for any $a \in [0,+\infty)$.

In order to compute the leading term of (8.81) we first bound the integral in the regime $(a,\tau,\xi) \in (\Lambda \setminus \widetilde{\Lambda}) \times ([0,1] \setminus I) \times (\Gamma \setminus \widetilde{\Gamma})$, with $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ defined in (8.52), that is the regime where we expect that the main contribution comes from, and then we use these bounds to firstly prove that the integral in the regime when either $|\xi| \leq N^\rho$ or $|a| \leq N^\rho$ is exponentially small for any $\tau \in [0,1]$, and then prove that also the τ-integral on I is exponentially small if $|a| \geq N^\rho$. The bounds for the ξ-integral over $\Gamma \setminus \widetilde{\Gamma}$ are exactly the same as Lemma 8.5.2, since the phase function $f(\xi)$ and the Γ-contour are exactly the same as the complex case. In order to estimate the integral over $(a,\tau) \in (\Lambda \setminus \widetilde{\Lambda}) \times ([0,1] \setminus I)$, we start with the estimate of the τ-integral over $[0,1] \setminus I$ in Lemma 8.6.2 and then we will conclude the computation of the a-integral over $[N^\rho, +\infty)$ in Lemma 8.6.3.

Before proceeding with the bounds for large $|a|, |\xi|$, in the following lemma we state some properties of the functions f and g. The proof of this lemma follows by elementary computations. From now on, for simplicity, we assume that $\eta \geq 0$; the case $\eta < 0$ is completely analogous since the functions g and G_N in (8.79) depends only on η^2 and $|z|^2$.

Lemma 8.6.1. *Let f and g be the phase functions defined in* (8.35) *and* (8.36)*, respectively, then the following properties hold true:*

(i) For any $\xi = -2/3 + \mathrm{i}t \in \Gamma_{1,\xi_*}$, we have that

$$\Re[f(-2/3+\mathrm{i}t)] = \frac{2E}{3} + \epsilon t - \frac{1}{2t^2} + \mathcal{O}(|t|^{-3} + \delta |t|^{-2}), \tag{8.82}$$

and

$$\Im[f(-2/3+\mathrm{i}t)] = \frac{2\epsilon}{3} - Et - \frac{\delta}{t} + \mathcal{O}(|t|^{-3}). \tag{8.83}$$

(ii) For $\epsilon = 0$, the function $t \mapsto \Re[f(-2/3+\mathrm{i}t)]$ on Γ_{1,ξ_*} is strictly increasing if $t > 0$ and strictly decreasing if $t < 0$.

(iii) For any $a \in [0,+\infty)$, we have that $g(a,\tau,\eta) \geq g(a,\tau,0)$ and the function $\tau \mapsto g(a,\tau,0)$ is strictly decreasing on $[0,1]$.

(iv) The function $a \mapsto g(a,1,\eta)$ is strictly decreasing on $[0, \xi_*/2]$.

Lemma 8.6.2. *Let $\rho > 0$ be sufficiently small, $I = I_a = [0, N^{\rho/2}a^{-1}]$, $\gamma \in \mathbf{N}$, $\gamma \geq 1$, $c(N)$ be defined in (8.44), $E \lesssim c(N)$, $0 \leq \delta \leq 1$ and let g be defined in (8.36). Then, for any $a \in [N^\rho, +\infty)$, we have*

$$\int_{[0,1]\setminus I_a} \frac{\left|e^{-N[g(a,\tau,\eta)-g(a,1,\eta)]}\right|}{\tau^{\gamma+1/2}}\, \mathrm{d}\tau \lesssim F(a) := \begin{cases} a^2N^{-1} \wedge 1, & N^\rho \leq a \leq \delta^{-1} \wedge \eta^{-1}, \\ a(N\delta)^{-1} \wedge 1, & \delta^{-1} \vee N^\rho \leq a \leq \delta\eta^{-2}, \\ (N\eta^2)^{-1} \wedge 1, & a \geq (\delta\eta^{-1} \vee 1)\eta^{-1}, \end{cases}$$
$$+ e^{-\frac{1}{2}N\eta^2} \times \begin{cases} e^{-\frac{N(\delta \vee N^{-1/2})}{a}}, & N^\rho \leq a \leq N(\delta \vee N^{-1/2}), \\ (a(N\delta \vee \sqrt{N})^{-1})^{\gamma-1/2}, & N(\delta \vee N^{-1/2}) \leq a \leq (\delta \vee N^{-1/2})\eta^{-2}, \\ (N\eta^2)^{1/2-\gamma}, & a \geq (\delta \vee N^{-1/2})\eta^{-2}, \end{cases} \tag{8.84}$$

where some regimes in (8.84) might be empty for certain values of δ and η.

Proof. In order to estimate the integral in the l.h.s. of (8.84) we first compute the expansion

$$\begin{aligned} g(a,\tau,\eta) - g(a,1,\eta) &= \frac{(1-2\delta)(1-\tau)-(1-\tau)^2}{a^2\tau^2} + \frac{2\eta^2(1-\tau)}{\tau} + \frac{(1-\tau)\delta}{a\tau} \\ &\quad + \mathcal{O}\left(\frac{1-\tau}{a^2\tau} + \frac{(1-\tau)(\delta+a^{-1})}{a^2\tau^2} + \frac{\eta^2(1-\tau)}{a\tau^2},\right), \end{aligned} \tag{8.85}$$

which holds true for any $\tau \in [0,1] \setminus I = [N^{\rho/2}a^{-1}, 1]$. Note that by (8.85) it follows that for any $(a,\tau) \in [N^\rho, +\infty) \times [N^{\rho/2}a^{-1}, 1]$ it holds

$$g(a,\tau,\eta) - g(a,1,\eta) \geq \frac{1-\tau}{2}\left[\frac{1}{a^2\tau^2} + \frac{\delta}{a\tau} + \frac{\eta^2}{\tau}\right].$$

Then, by (8.85) it follows that

$$\int_{[0,1]\setminus I} \frac{\left|e^{-N[g(a,\tau,\eta)-g(a,1,\eta)]}\right|}{\tau^{\gamma+1/2}}\, \mathrm{d}\tau \lesssim \int_{N^{\rho/2}a^{-1}}^1 \frac{e^{-(1-\tau)\frac{N}{2}\left[\frac{1}{a^2\tau^2}+\frac{\delta}{a\tau}+\frac{\eta^2}{\tau}\right]}}{\tau^{\gamma+1/2}}\, \mathrm{d}\tau. \tag{8.86}$$

In order to bound the r.h.s. of (8.86) we split the computations into two cases: $\delta \leq N^{-1/2}$ and $\delta > N^{-1/2}$. We firstly consider the case $\delta > N^{-1/2}$. In order to prove the bound in the r.h.s. of (8.84) we further split the computation of the τ-integral into the regimes $\tau \in [N^{\rho/2}a^{-1}, 1/2]$ and $\tau \in [1/2, 1]$. We start estimating the integral over $[1/2, 1]$ as follows

$$\begin{aligned} \int_{1/2}^1 \frac{e^{-(1-\tau)\frac{N}{2}\left[\frac{1}{a^2\tau^2}+\frac{\delta}{a\tau}+\frac{\eta^2}{\tau}\right]}}{\tau^{\gamma+1/2}}\, \mathrm{d}\tau &\lesssim \int_{1/2}^1 e^{-(1-\tau)\frac{N}{2}\left[\frac{1}{a^2}+\frac{\delta}{a}+\eta^2\right]}\, \mathrm{d}\tau \\ &\lesssim \begin{cases} a^2N^{-1} \wedge 1, & N^\rho \leq a \leq \delta^{-1} \wedge \eta^{-1}, \\ a(N\delta)^{-1} \wedge 1, & N^\rho \vee \delta^{-1} \leq a \leq \delta\eta^{-2}, \\ (N\eta^2)^{-1} \wedge 1, & a \geq \delta\eta^{-2} \vee \eta^{-1}. \end{cases} \end{aligned} \tag{8.87}$$

For the integral over $\tau \in [N^{\rho/2}a^{-1}, 1/2]$, instead, we bound the r.h.s. of (8.86) as

$$\int_{N^{\rho/2}a^{-1}}^{1/2} \frac{e^{-\frac{N}{2}\left[\frac{\delta}{a\tau}+\frac{\eta^2}{\tau}\right]}}{\tau^{\gamma+1/2}}\,\mathrm{d}\tau \lesssim e^{-\frac{1}{2}N\eta^2} \times \begin{cases} e^{-\frac{1}{2}N\delta a^{-1}}(a(N\delta)^{-1})^{\gamma-1/2}, & N^\rho \le a \le \delta\eta^{-2}, \\ (N\eta^2)^{1/2-\gamma}, & a \ge \delta\eta^{-2}. \end{cases} \tag{8.88}$$

Then, combining (8.87)–(8.88) we conclude the bound in (8.84) for $\delta > N^{-1/2}$. Using similar computations for $\delta \le N^{-1/2}$, we conclude the bound in (8.84). □

In the following lemma we conclude the bound for the double integral (8.81) in the regime $(a, \tau) \in ([N^\rho, +\infty) \times ([0, 1] \setminus I)$ using the bound in (8.84) as an input.

Lemma 8.6.3. *Let $\rho > 0$ be sufficiently small, $I = [0, N^{\rho/2}a^{-1}]$, $0 \le \delta \le 1$, let $c(N)$ be defined in* (8.44)*, $E \lesssim c(N)$, and let g be defined in* (8.36)*. Then, for any integers $\alpha \ge 2$, $1 \le \gamma \le \alpha$, we have*

$$\int_{\Lambda\setminus\widetilde{\Lambda}} \int_{[0,1]\setminus I}^{1} \left| \frac{e^{-Ng(a,\tau,\eta)}}{a^{\alpha-1}\tau^{\gamma+1/2}} \right| \mathrm{d}\tau\,\mathrm{d}a \lesssim C_1 + e^{-\frac{1}{2}N\eta^2}(N\eta^2)^{1/2-\gamma}C_2 + +e^{-\frac{1}{2}N\eta^2}(N\delta \vee \sqrt{N})^{1/2-\gamma}C_3, \tag{8.89}$$

where

$$C_1 := \begin{cases} 1 + |\log[N(\delta \vee N^{-1/2})\xi_*^{-1}]|, & \alpha = 2,\ [\delta\eta^{-1} \vee 1]\eta^{-1} > \xi_*, \\ ((N\eta^2)^{-1} \wedge 1)(1 + |\log[N(\delta \vee N^{-1/2})\xi_*^{-1}]|), & \alpha = 2,\ [\delta\eta^{-1} \vee 1]\eta^{-1} \le \xi_*, \\ [N(\delta \vee N^{-1/2})]^{2-\alpha}, & \alpha \ge 3,\ [\delta\eta^{-1} \vee 1]\eta^{-1} > \xi_* \\ ((N\eta^2)^{-1} \wedge 1)[N(\delta \vee N^{-1/2})]^{2-\alpha}, & \alpha \ge 3,\ [\delta\eta^{-1} \vee 1]\eta^{-1} \le \xi_* \end{cases}$$

$$C_2 := \begin{cases} (1 + |\log[N(\delta \vee N^{-1/2})\xi_*^{-1}]|)e^{-\frac{NE}{2\eta^2}(\delta \vee N^{-1/2})}, & \alpha = 2, \\ [(\delta^{-1} \wedge \sqrt{N})\eta^2]^{\alpha-2}e^{-\frac{NE}{2\eta^2}(\delta \vee N^{-1/2})}, & \alpha \ge 3,\ (\delta^{-1} \wedge \sqrt{N})\eta^2 \le NE, \\ [(\delta^{-1} \wedge \sqrt{N})\eta^2]^{\alpha-2}, & \alpha \ge 3,\ (\delta^{-1} \wedge \sqrt{N})\eta^2 \ge NE, \end{cases}$$

$$C_3 := \begin{cases} (NE)^{\alpha-\gamma-3/2}, & \gamma = \alpha, \alpha - 1,\ (\delta^{-1} \wedge \sqrt{N})\eta^2 \le NE, \\ [(\delta^{-1} \wedge N^{1/2})\eta^2]^{\alpha-\gamma-3/2}, & \gamma = \alpha, \alpha - 1,\ (\delta^{-1} \wedge \sqrt{N})\eta^2 \ge NE, \\ (N\delta \vee \sqrt{N})^{3/2+\gamma-\alpha}, & \gamma \le \alpha - 2. \end{cases}$$

Proof. Firstly, we add and subtract $Ng(a, 1, \eta) = Nf(a)$ to the phase function in the exponent and conclude, by Lemma 8.6.2, that

$$\int_{N^\rho}^{+\infty} \int_{[0,1]\setminus I}^{1} \left| \frac{e^{-Ng(a,\tau,\eta)}}{a^{\alpha-1}\tau^{\gamma+1/2}} \right| \mathrm{d}\tau\,\mathrm{d}a \lesssim \int_{N^\rho}^{+\infty} \frac{\left|e^{-Ng(a,1,\eta)}\right| \cdot F(a)}{a^{\alpha-1}}\,\mathrm{d}a, \tag{8.90}$$

with $F(a)$ defined in (8.84). In the following of the proof we often use that $NE\xi_* \lesssim 1$ by the definition of ξ_* in (8.80), that implies $e^{NE\xi_*} \lesssim 1$. We split the computation of the integral in the r.h.s. of (8.90) as the sum of the integrals over $[N^\rho, \xi_*]$ and $[\xi_*, +\infty)$. From now on we consider only the case $\delta > N^{-1/2}$, since the case $\delta \le N^{-1/2}$ is completely analogous. In the regime $\delta > N^{-1/2}$ we have $E \lesssim c(N) = \delta^{-1}N^{-2} \lesssim \delta^3$, therefore $\xi_* \sim \sqrt{\delta E^{-1}}$ from (8.80).

Then, using the expansion for large a-argument of $g(a,1,\eta)=f(a)$ in (8.75), we start estimating the integral over $[N^\rho,+\infty)$ as follows

$$
\begin{aligned}
&e^{NE\xi_*}\int_{N^\rho}^{\xi_*}\frac{e^{-N\left[\frac{\delta}{a}+\frac{1}{2a^2}\right]}F(a)}{a^{\alpha-1}}\,\mathrm{d}a \\
&\lesssim \chi(\delta\eta^{-2}\le\xi_*)e^{-\frac{1}{2}N\eta^2}(N\eta^2)^{1/2-\gamma}\times\begin{cases}1+|\log(N\delta\xi_*^{-1})|, & \alpha=2,\\ (\delta^{-1}\eta^2)^{\alpha-2}, & \alpha\ge 3.\end{cases} \\
&\qquad + e^{-\frac{1}{2}N\eta^2}\times\begin{cases}N^{1/2-\gamma}\delta^{2-\alpha}\eta^{2\alpha-2\gamma-3}, & \gamma=\alpha,\alpha-1,\delta\eta^{-2}\le\xi_*\\ (N\delta)^{1/2-\gamma}\xi_*^{3/2+\gamma-\alpha}, & \gamma=\alpha,\alpha-1,\delta\eta^{-2}\ge\xi_*\\ (N\delta)^{2-\alpha}, & \gamma\le\alpha-2,\end{cases} \\
&\qquad + \begin{cases}1+|\log(N\delta\xi_*^{-1})|, & \alpha=2,\ [\delta\eta^{-1}\vee 1]\eta^{-1}>\xi_*,\\ ((N\eta^2)^{-1}\wedge 1)(1+|\log(N\delta\xi_*^{-1})|), & \alpha=2,\ [\delta\eta^{-1}\vee 1]\eta^{-1}\le\xi_*,\\ (N\delta)^{2-\alpha}, & \alpha\ge 3\,[\delta\eta^{-1}\vee 1]\eta^{-1}>\xi_*,\\ ((N\eta^2)^{-1}\wedge 1)(N\delta)^{2-\alpha} & \alpha\ge 3\,[\delta\eta^{-1}\vee 1]\eta^{-1}\le\xi_*,\end{cases}
\end{aligned}
\tag{8.91}
$$

To conclude the proof we are left with the estimate of the a-integral on $[\xi_*,+\infty)$. In this regime we bound the r.h.s. of (8.90) as follows

$$
\begin{aligned}
&e^{NE\xi_*}\int_{\xi_*}^{+\infty}\frac{e^{-NEa}F(a)}{a^{\alpha-1}}\,\mathrm{d}a \\
&\lesssim e^{-\frac{1}{2}N\eta^2}(N\eta^2)^{1/2-\gamma}\times\begin{cases}(1+|\log(NE\xi_*)|)\,e^{-NE\delta\eta^{-2}/2}, & \alpha=2,\\ e^{-NE\delta\eta^{-2}}(\delta^{-1}\eta^2)^{\alpha-2}, & \alpha\ge 3,\ \delta^{-1}\eta^2\le NE\wedge\xi_*^{-1},\\ (\delta^{-1}\eta^2)^{\alpha-2}, & \alpha\ge 3,\ NE\le\delta^{-1}\eta^2\le\xi_*^{-1},\\ \xi_*^{2-\alpha}, & \alpha\ge 3,\ \delta^{-1}\eta^2\ge\xi_*^{-1},\end{cases} \\
&\quad +\chi(\xi_*\le\delta\eta^{-2})e^{-\frac{1}{2}N\eta^2}(N\delta)^{1/2-\gamma}\times\begin{cases}(NE)^{\alpha-\gamma-3/2}, & \gamma=\alpha,\alpha-1,\ \delta^{-1}\eta^2\le NE,\\ (\delta^{-1}\eta^2)^{\alpha-\gamma-3/2}, & \gamma=\alpha,\alpha-1,\ \delta^{-1}\eta^2\ge NE,\\ \xi_*^{3/2+\gamma-\alpha}, & \gamma\le\alpha-2,\end{cases} \\
&\quad +\begin{cases}1+|\log(N\delta\xi_*^{-1})|, & \alpha=2,\ [\delta\eta^{-1}\vee 1]\eta^{-1}>\xi_*,\\ ((N\eta^2)^{-1}\wedge 1)(1+|\log(N\delta\xi_*^{-1})|), & \alpha=2,\ [\delta\eta^{-1}\vee 1]\eta^{-1}\le\xi_*,\\ (\xi_*)^{2-\alpha}, & \alpha\ge 3\,[\delta\eta^{-1}\vee 1]\eta^{-1}>\xi_*,\\ ((N\eta^2)^{-1}\wedge 1)(\xi_*)^{2-\alpha}, & \alpha\ge 3\,[\delta\eta^{-1}\vee 1]\eta^{-1}>\xi_*.\end{cases}
\end{aligned}
\tag{8.92}
$$

Finally, combining (8.91) and (8.92) we conclude the bound in (8.89). □

In order to conclude the estimate of the leading order term of (8.81), in the following lemma, using the bounds in Lemma 8.5.2 for the ξ-integral and the ones in Lemma 8.6.3 for the (a,τ)-integral, we prove that the contribution to (8.81) in the regime when either $a\in[0,N^\rho]$ or $\xi\in\widetilde{\Gamma}$ and in the regime $\tau\in I$ is exponentially small.

Lemma 8.6.4. *Let* $c(N)$ *be defined in* (8.44), $0 \le \delta \le 1$, $I = I_a = [0, N^{\rho/2}a^{-1}]$, *and let* f, g *and* G_N *be defined in* (8.35)–(8.37), *then, for any* $E \lesssim c(N)$, *we have that*

$$\left| \left(\int_\Gamma \mathrm{d}\xi \int_\Lambda \mathrm{d}a \int_0^1 \mathrm{d}\tau - \int_{\Gamma\setminus\widetilde{\Gamma}} \mathrm{d}\xi \int_{\Lambda\setminus\widetilde{\Lambda}} \mathrm{d}a \int_{[0,1]\setminus I} \mathrm{d}\tau \right) \left[e^{N[f(\xi)-g(a,\tau,\eta)]} \frac{a\xi^2}{\tau^{1/2}} G_N(a,\tau,\xi,z) \right] \right|$$
$$\lesssim \frac{N^{5/2+5\rho}(N^{1/2}+N\delta)}{E^{1/2}} e^{-\frac{1}{2}N^{1-2\rho}}. \tag{8.93}$$

Proof. We split the proof into three parts, we first prove that the contribution to (8.81) in the regime $a \in \widetilde{\Lambda} = [0, N^\rho]$ is exponentially small uniformly in $\tau \in [0,1]$ and $\xi \in \Gamma$, then we prove that for $a \ge N^\rho$ the contribution to (8.81) in the regime $\tau \in I$ is exponentially small uniformly in $\xi \in \Gamma$, and finally we conclude that also the contribution for $\xi \in \widetilde{\Gamma}$ is negligible.

Note that for any $a \in [0, +\infty)$, $\tau \in [0,1]$ we have that the map $\tau \mapsto g(a,\tau,0)$ is strictly decreasing by iii of Lemma 8.6.1, hence, using that $g(a,\tau,\eta) \ge g(a,\tau,0)$ and ii-iv of Lemma 8.6.1, it follows that

$$\sup_{\xi\in\widetilde{\Gamma}} |e^{Nf(\xi)}| + \sup_{a\in\widetilde{\Lambda}} |e^{-Ng(a,\tau,\eta)}| \le \sup_{\xi\in\widetilde{\Gamma}} |e^{Nf(\xi)}| + \sup_{a\in\widetilde{\Lambda}} |e^{-Ng(a,1,0)}| \lesssim e^{-Nf(N^\rho)}, \tag{8.94}$$

with

$$f(N^\rho) = \frac{\delta}{N^\rho} + \frac{1}{2N^{2\rho}} + \mathcal{O}(N^{-3\rho} + \delta N^{-2\rho}). \tag{8.95}$$

In order to estimate the regime $a \in \widetilde{\Lambda}$, we split the computation into two cases: $(a,\xi) \in \widetilde{\Lambda} \times \widetilde{\Gamma}$ and $(a,\xi) \in \widetilde{\Lambda} \times (\Gamma \setminus \widetilde{\Gamma})$. Then, by (8.94)–(8.95) it follows that the integral in the regime $(a,\tau,\xi) \in \widetilde{\Lambda} \times [0,1] \times \widetilde{\Gamma}$ is bounded by $N^{2\rho}e^{-N^{1-2\rho}/2}$. Note that in the regime $(a,\tau,\xi) \in \widetilde{\Lambda} \times [0,1] \times (\Gamma \setminus \widetilde{\Gamma})$ we have $|\xi| \ge N^\rho$. Hence, by the explicit form of $G_{1,N}, G_{2,N}$ in (8.37), using the bound in (8.94) for $e^{-Ng(a,\tau,\eta)}$, that $|\xi| \ge N^\rho$ and so Lemma 8.5.2 to bound teh regime $\Gamma \setminus \widetilde{\Gamma}$, we conclude that the integral over $(a,\tau,\xi) \in \widetilde{\Lambda} \times [0,1] \times (\Gamma \setminus \widetilde{\Gamma})$ is bounded by $N^{3+\rho}(N^{1/2} + (N\delta))e^{-N^{1-2\rho}/2}$.

Next, we consider the integral over $(a,\tau,\xi) \in (\Lambda \setminus \widetilde{\Lambda}) \times [0, N^{\rho/2}a^{-1}] \times \Gamma$. Note that in this regime $a \ge N^\rho$. Since $g(a,\tau,\eta) \ge g(a,\tau,0)$ and $\tau \mapsto g(a,\tau,0)$ is strictly decreasing by iii of Lemma 8.6.1, we have that

$$e^{-Ng(a,\tau,\eta)} \le e^{-Ng(a,N^{\rho/2}a^{-1},0)}, \tag{8.96}$$

where

$$g(a, N^{\rho/2}a^{-1}, \eta) = Ea + \frac{\delta}{N^{\rho/2}} + \frac{1}{N^\rho} + \mathcal{O}(N^{-3\rho/2} + \delta N^{-\rho}). \tag{8.97}$$

Additionally, using the explicit expression of G_N in (8.37), the bound (8.94) on $e^{Nf(\xi)}$ for the regime $\xi \in \widetilde{\Gamma}$, and Lemma 8.5.2 for $\xi \in \Gamma \setminus \widetilde{\Gamma}$, we get

$$\left| \int_\Gamma \mathrm{d}\xi G_N(a,\tau,\xi,z)\xi^2 e^{Nf(\xi)} \right| \lesssim C(N,\eta,a,\tau), \tag{8.98}$$

where

$$C(a,\tau) = C(N,\eta,a,\tau) := (N^{1/2} + N\delta) \left(N^2\eta^2 + N^2\tau + \frac{N^2}{a} + \frac{N^2}{a^2\tau} \right).$$

Thus, using (8.98) and the explicit form of $g(a,\tau,\eta)$ in (8.36), for $\tau\in[0,N^{\rho/2}a^{-1}]$ we have

$$\begin{aligned}\left|\frac{a}{\tau^{1/2}}e^{-Ng(a,\tau,\eta)}\int_\Gamma \mathrm{d}\xi G_N(a,\tau,\xi,z)\xi^2e^{Nf(\xi)}\right| &\\ &\lesssim C(a,\tau)e^{-(N-2)g(a,\tau,\eta)}\frac{a^3\tau}{\tau^{1/2}(1+2a+a^2\tau)} \\ &\lesssim C(a,\tau)a^2\tau^{1/2}e^{-(N-2)g(a,\tau,0)}e^{-\frac{(N-2)\eta^2a^2}{1+2a+a^2\tau}} \\ &\lesssim C(a,\tau)a^2\tau^{1/2}e^{-\frac{1}{2}N^{1-2\rho}}e^{-NEa-\frac{(N-2)\eta^2a^2}{1+2a+a^2\tau}}.\end{aligned} \tag{8.99}$$

Hence, integrating (8.99) with respect to (a,τ), and using that in the regime $\tau\in[0,N^{\rho/2}a^{-1}]$ it holds

$$N^{-\rho/2}a\lesssim\frac{a^2}{1+2a+a^2\tau}\lesssim a,$$

we conclude that the integral over $(a,\tau,\xi)\in(\Lambda\setminus\widetilde{\Lambda})\times[0,N^{\rho/2}a^{-1}]\times\Gamma$ is bounded by $N^{5/2+5\rho}(N^{1/2}+N\delta)E^{-1/2}e^{-N^{1-2\rho}}$.

Finally, in order to conclude the bound in (8.93), we are left with the estimate of the integral over $(a,\tau,\xi)\in(\Lambda\setminus\widetilde{\Lambda})\times[N^{\rho/2}a^{-1},1]\times\widetilde{\Gamma}$. In this regime, using the bound in (8.94) on $e^{Nf(\xi)}$ for $\xi\in\widetilde{\Gamma}$, and Lemma 8.6.3 to estimate the integral over $(a,\tau)\in(\Lambda\setminus\widetilde{\Lambda})\times[N^{\rho/2}a^{-1},1]$, we get the bound $N^{5/2+\rho}E^{-1/2}e^{-N^{1-2\rho}/2}$. This concludes the proof of (8.93). □

Proof of Theorem 8.2.3 in the case $\delta\geq 0$. Using Lemma 8.6.4 we remove the regime $a\leq N^\rho$, $|\xi|\leq N^\rho$ or $\tau\in[0,N^{\rho/2}a^{-1}]$ in (8.81). Then, using the expansion for $G_{1,N}$ and $G_{2,N}$ in (8.76)-(8.78) in the remaining regime of (8.81), combining Lemma 8.5.2 and Lemma 8.6.3 we conclude Theorem 8.2.3. □

8.6.2 Case $-CN^{-1/2}\leq\delta<0$.

Now we summarize the necessary changes for the case $\delta<0$. Similarly to the case $0\leq\delta\leq 1$, all along this section we assume that $E'<0$ in (8.35)–(8.36), i.e. $E'=-E$ with $0\leq E\lesssim N^{-3/2}$.

Let x_* be the real stationary point of f, i.e. x_* at leading order is given by

$$x_*\approx\begin{cases}3|\delta|^{-1}/2 & \text{if}\quad E\ll|\delta|^3,\\ \mu(c)E^{-1/3} & \text{if}\quad E=c|\delta|^3,\\ E^{-1/3} & \text{if}\quad E\gg|\delta|^3,\end{cases}$$

for some function $\mu(c)>0$ and any fixed constant $c>0$ independent of N, E, and δ. As in the complex case, we can treat the regime $0<-\delta\lesssim N^{-1/2}$ similarly to the regime $0\leq\delta\lesssim N^{-1/2}$, since for $|\delta|\lesssim N^{-1/2}$ the only δ-dependent terms, i.e. the term δa^{-1} in the expansion of $g(a,1,\eta)=f(a)$ in (8.75) and the term $(1-\tau)\delta(a\tau)^{-1}$ in the expansion of $g(a,\tau,\eta)-g(a,1,\eta)$, do not play any role in the estimates of the (a,τ) integral in the regime $a\geq N^\rho$, $\tau\in[N^{\rho/2}a^{-1},1]$. Note that this is also the case for $0\leq\delta\leq N^{-1/2}$, when the estimates (8.84) and (8.89) were derived. For this reason, unlike the case $0\leq\delta\leq 1$, in the present case, $\delta<0$, $|\delta|\lesssim N^{-1/2}$ and $E\lesssim c(N)$, we do not deform the ξ-contour through the leading order of the saddle x_*, but we always deform it as $\Gamma=\Gamma_{\xi_*}:=\Gamma_{1,\xi_*}\cup\Gamma_{2,\xi_*}$, where $\xi_* := E^{-1/3}$, with $\Gamma_{1,\xi_*}, \Gamma_{2,\xi_*}$ defined in (8.48a) replacing z_* by ξ_*. Note that we could have done the same choice in the case $0 \leq \delta \lesssim N^{-1/2}$, but not for the regime $N^{-1/2} \ll \delta \leq 1$, hence, in order to treat the regime $0 \leq \delta \leq 1$ in the same way for any δ, in Section 8.6.1 we deformed the contour Γ through (8.80). For any $E \lesssim c(N)$ and $0 < -\delta \lesssim N^{-1/2}$ we have $\xi_* \gg 1$, hence we prove that the main contribution to (8.81) comes from the regime when $a, |\xi| \geq N^\rho$. Moreover, similarly to the case $0 \leq \delta \leq 1$, the contribution to (8.81) in the regime $(a, \tau, \xi) \in (\Lambda \setminus \widetilde{\Lambda}) \times [0, N^{\rho/2}a^{-1}] \times (\Gamma \setminus \widetilde{\Gamma})$ will be exponentially small. Hence, in order to estimate the leading order of (8.81), we expand f, $g(\cdot, 1, \eta)$ and G_N for large a and $|\xi|$ arguments as in (8.75)–(8.78).

The phase functions f and g, defined in (8.35) and (8.36), respectively, satisfy the properties i and ii of Lemma 8.6.1, but not the ones in iii and iv. Instead, it is easy to see that the following lemma holds true.

Lemma 8.6.5. *Let f and g be the phase functions defined in* (8.35) *and* (8.36), *respectively. Then, the following properties hold true:*

(iii') For any $a \in [0, +\infty)$, we have that $g(a, \tau, \eta) \geq g(a, \tau, 0)$ and that

$$e^{-Ng(a,\tau,0)} \lesssim e^{-Ng(a,\tau_0,0)},$$

for any fixed $\tau_0 \in [0, 1]$ and any $\tau \in [0, \tau_0]$.

(iv') The function $a \mapsto g(a, 1, 0)$ is strictly decreasing on $[0, |\delta|^{-1}/2]$.

Since $|\delta| \lesssim N^{-1/2}$, by iv of Lemma 8.6.5 it clearly follows that the function $a \mapsto g(a, 1, \eta)$ is strictly decreasing on $[0, N^\rho]$. Note that for $|\delta| \lesssim N^{-1/2}$ we have

$$e^{-N(1-\tau)\left[\frac{1}{a^2\tau^2} + \frac{\delta}{a\tau}\right]} \lesssim e^{-\frac{(1-\tau)N}{2a^2\tau^2}}, \quad e^{-N\left[\frac{\delta}{a} + \frac{1}{2a^2}\right]} \lesssim e^{-\frac{N}{4a^2}}, \tag{8.100}$$

for any $a \in [N^\rho, +\infty)$ and $\tau \in [N^{\rho/2}a^{-1}, 1]$. Using (8.100), inspecting the proof of (8.84) and (8.89) in Lemma 8.6.2 and Lemma 8.6.3, respectively, and noticing that in the regime $0 \leq \delta \leq N^{-1/2}$ the sign of δ did not play any role we conclude that the bounds (8.84), (8.89) hold true for the case $\delta < 0$, $|\delta| \lesssim N^{-1/2}$ as well. Then, similarly to the case $0 \leq \delta \leq 1$, by i–ii of Lemma 8.6.1 and iii–iv of Lemma 8.6.5, using the bound in (8.89) to estimate the (a, τ)-integral in the regime $(a, \tau) \in [N^\rho, +\infty) \times [N^{\rho/2}a^{-1}, 1]$ and the ones in Lemma 8.5.2 to estimate the ξ-integral in the regime $|\xi| \geq N^\rho$ we conclude that the contribution to (8.81) in the regime when either $a \in [0, N^\rho]$ or $|\xi| \leq N^\rho$ and in the regime $[0, N^{\rho/2}a^{-1}]$ is exponentially small, i.e. Lemma 8.6.4 holds true once δ is replaced by $|\delta|$ everywhere.

Proof of Theorem 8.2.3 in the case $-CN^{-1/2} \leq \delta < 0$. By combining (8.89), Lemma 8.5.2 and Lemma 8.6.4, using the expansion of G_N in (8.76)–(8.78), we conclude the proof of Theorem 8.2.3 also in the case $-CN^{-1/2} \leq \delta < 0$. □

8.A Superbosonisation formula for meromorphic functions

The superbosonisation formulas [142, Eq. (1.10) and (1.13)] (see also [9, Corollary 2.6] for more precise conditions) are stated under the condition that

$$F(\Phi^*\Phi) = F\begin{pmatrix} \langle s, s\rangle & \langle s, \chi\rangle \\ \langle \chi, s\rangle & \langle \chi, \chi\rangle \end{pmatrix},$$

viewed as a function of four independent variables, is holomorphic and decays faster than any inverse power at real $+\infty$ in the $\langle s,s\rangle$ variable (for definiteness, we discuss the complex case; the argument for the real case is analogous). Our function F defined in (8.26) has a pole at $\langle s,s\rangle = \mathrm{i}N$ and $\langle \chi,\chi\rangle = \mathrm{i}N$ using the definitions (8.20)–(8.21) after expanding the inverse of the matrix $1+\frac{\mathrm{i}}{N}\Phi^*\Phi$ in the Grassmannian variables but this pole is far away from the integration domain on both sides of (8.22). We now outline a standard approximation procedure to verify the superbosonisation formula for such meromorphic functions; for simplicity we consider only our concrete function from (8.26).

In the first step notice that the integration at infinity on the non-compact domain for the boson-boson variable is absolutely convergent on both sides as guaranteed by the $\exp(\mathrm{i}w\,\mathrm{STr}\,\Phi^*\Phi)$ regularization, since $\Im w > 0$.

Second, in the LHS of (8.22) using Taylor expansions, we expand F into a finite polynomial in the Grassmannian variables with meromorphic coefficient functions in the variable $\langle s,s\rangle$. Algebraically, we perform exactly the same expansion in the RHS of (8.22). For the fermionic variables σ,τ these expansions naturally terminate after finitely many terms. From the formulas (8.28) it is clear that only the geometric expansion $(1+y)^{-1} = 1-y+y^2-\dots$ may result in an infinite power series instead of a finite polynomial. However, owing to the contour integral in y and that the integrand has a pole of at most finite order ($\approx N$) at zero, we may replace this power series with its finite truncation without changing the value of the RHS of (8.22). We choose the order of truncation sufficiently large that the remaining formula contains all non-zero terms on both sides. We denote this new truncated function by $\widetilde{F}$.

Now we are in the situation where on both sides of (8.22), with $\widetilde{F}$ replacing F, we have the same finite polynomial in the variables $\langle s,\chi\rangle, \langle\chi,s\rangle$ and $\langle\chi,\chi\rangle$ in the LHS as in the variables σ,τ,y in the RHS, with coefficients that are meromorphic in $\langle s,s\rangle$, resp. in x. All coefficient functions $h_k(x)$ are analytic in a neighborhood of the positive real axis (their possible pole is at -1) and they have an exponential decay $\approx \exp(-(\Im w)\langle s,s\rangle)$ in the LHS, resp. $\exp(-(\Im w)x)$ in the RHS, at infinity from the regularization observed in the first step.

Finally, in the third step, dropping the k index temporarily, we write each coefficient function as $h(x) = g(x)e^{-\alpha x}$ with $\alpha = \frac{1}{2}\Im w$. For any given $\epsilon > 0$ we approximate $g(x)$ via classical (rescaled) Laguerre polynomials $p_n(x)$ of degree n with weight function $e^{-\alpha x}$ such that $\int_0^\infty |g(x)-p_n(x)|^2 e^{-\alpha x}\,\mathrm{d}x \le (\epsilon\alpha)^2$, where n depends on ϵ and $\Im w$. By completeness of the Laguerre polynomials in $L^2(\mathbf{R}_+, e^{-\alpha x}\,\mathrm{d}x)$ and by $\int |g(x)|^2 e^{-\alpha x}\,\mathrm{d}x = \int |h(x)|^2 e^{\alpha x}\,\mathrm{d}x < \infty$ such approximating polynomial exists. Therefore, with a Schwarz inequality, we have

$$\int_0^\infty |h(x) - p_n(x)e^{-\alpha x}|\,\mathrm{d}x = \int_0^\infty |g(x)-p_n(x)|e^{-\alpha x}\,\mathrm{d}x \le \epsilon.$$

Since there are only finitely many coefficient functions $h(x) = h_k(x)$ in $\widetilde{F}$, we can replace each of them with an entire function (namely with a polynomial times $e^{-\alpha x}$) with at most an ϵ error in the RHS of (8.22). The same estimates hold on the LHS. But for these replacements the superbosonisation formula [142, Eq. (1.10)] is applicable since the new functions are entire. The error is at most ϵ on both sides, but this argument is valid for arbitrary $\epsilon > 0$. This proves the superbosonisation formula for the function (8.26).

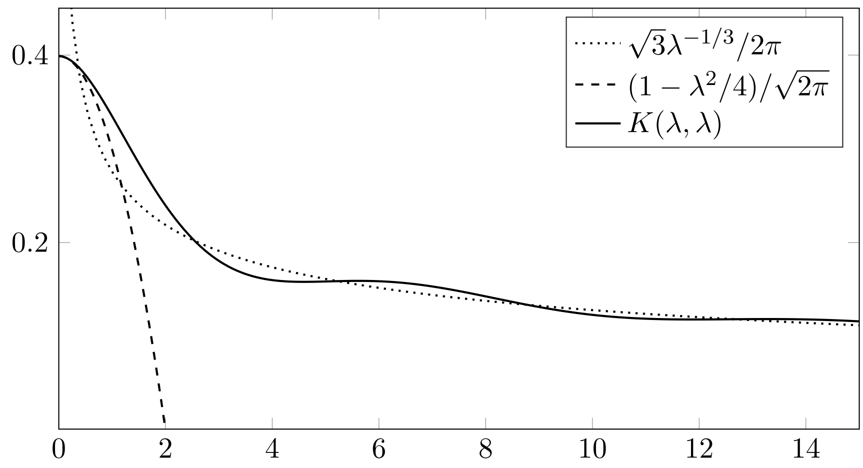


FIGURE 8.7: Plot of the 1-point function $K(\lambda,\lambda)=\pi^{-1}\Im q_0(\lambda)$ in the complex case with $|z|=1$. The dotted and dashed lines show the large λ and small λ asymptotes, respectively.

8.B Explicit formulas for the real symmetric integral representation

Here we collect the explicit formulas for the polynomials of a,ξ,τ in the definition of G_N in (8.34).

$$\begin{aligned}
p_{2,0,0} &:= a^4\tau^2+2a^3\xi\tau+4a^3\tau-a^2\xi^2\tau+4a^2\xi^2+8a^2\xi+2a^2\tau\\
&\qquad +4a^2+2a\xi^3+8a\xi^2+10a\xi+4a+\xi^4+4\xi^3+6\xi^2+4\xi+1\\
p_{1,0,0} &:= -a^4\xi\tau^2+a^4\tau^2-2a^3\xi^2\tau-2a^3\xi\tau+4a^3\tau-a^2\xi^3\tau-3a^2\xi^2\tau\\
&\qquad -2a^2\xi\tau+4a^2\xi+2a^2\tau+4a^2+2a\xi^2+6a\xi+4a+\xi^3+3\xi^2+3\xi+1\\
p_{2,2,0} &:= 4(a+1)\left(a^2\tau+a\xi\tau+2a\tau+\xi^2+2\xi+1\right)\\
p_{1,2,0} &:= 4(a+1)\left(a^2\tau+a\xi\tau+2a\tau+\xi+1\right)\\
p_{2,0,1} &:= 2(a^3\tau^2+2a^2\xi\tau+4a^2\tau+2a\xi^2+2a\xi\tau\\
&\qquad +4a\xi+3a\tau+2a+\xi^3+4\xi^2+5\xi+2)\\
p_{1,0,1} &:= 2(a^3\tau^2+2a^2\xi\tau+4a^2\tau+a\xi^2\tau+3a\xi\tau\\
&\qquad +2a\xi+3a\tau+2a+\xi^2+3\xi+2)\\
p_{2,2,1} &:= 4(a+1)(a+\xi+2)\\
p_{2,0,2} &:= a^2\tau+2a\xi+4a+\xi^2+4\xi+4
\end{aligned}$$

8.C Comparison with the contour-integral derivation

In [28] the correlation kernel of $(X-z)(X-z)^*$ for complex Ginibre matrices X has been derived using contour-integral methods. Earlier, the joint eigenvalue density for the general *Laguerre ensemble* had been obtained in the physics literature [107, 116] via supersymmetric methods, see also [206] with orthogonal polynomials. Adapting [28] to our scaling, and choosing $y_i=\pm 1$, it follows from [28, Theorem 7.1] that for $|z|=1$ the rescaled kernel

$K_N(N^{-3/2}\lambda, N^{-3/2}\mu)$ is given by

$$\frac{N^3}{\mathrm{i}\pi}\int_\Gamma \mathrm{d}x\int_\gamma \mathrm{d}y K_B(2N^{1/4}x\sqrt{\lambda}, 2N^{1/4}y\sqrt{\mu})e^{N(h(y)-h(x))}\Big(1-\frac{1}{1-x^2}\frac{1}{1-y^2}\Big)xy,$$
$$K_B(x,y) := \frac{xI_1(x)I_0(y)-yI_0(x)I_1(y)}{x^2-y^2}, \quad h(x) = x^2+\log(1-x^2) = -\frac{x^4}{2}+\mathcal{O}(x^5), \tag{8.101}$$

where I_0, I_1 are the modified Bessel function of 0-th and 1-st kind. The contour Γ is any contour encircling $[-1,1]$ in a counter-clockwise direction (in contradiction to the contours depicted in [28, Figure 8.1]) and the contour γ is composed of two straight half-lines $[0,\mathrm{i}\infty)$ and $[0,-\mathrm{i}\infty)$. The main contribution in (8.101) comes from the $|x|\lesssim N^{-1/4}$ and $|y|\lesssim N^{-1/4}$ regime which motivates the change of variables $x\mapsto N^{-1/4}x, y\mapsto N^{-1/4}y$. Together with the expansion of $1-(1-x^2)^{-1}(1-y^2)^{-1} = -x^2-y^2+\mathcal{O}(x^4+y^4)$ it follows that

$$K_N(N^{-3/2}\lambda, N^{-3/2}\mu)\approx N^{3/2}K(\lambda,\mu),$$

where

$$K(\lambda,\mu) := \frac{\mathrm{i}}{\pi}\int_{\Gamma'}\mathrm{d}x\int_\gamma \mathrm{d}y K_B(2x\sqrt{\lambda}, 2y\sqrt{\mu})e^{x^4/2-y^4/2}xy(x^2+y^2)$$

and Γ' consists of four straight half-lines $(e^{\mathrm{i}\pi/4}\infty, 0]$, $[0, e^{3\mathrm{i}\pi/4}\infty)$, $(e^{5\mathrm{i}\pi/4}\infty, 0]$, $[0, e^{7\mathrm{i}\pi/4}\infty)$.

We now compare the limiting 1-point function $K(\lambda,\lambda)$ with the asymptotic expansion we derived in Theorem 8.2.1, which in the case $|z|=1$, i.e. $\delta=0$, simplifies to

$$q_0(\lambda) = \frac{\lambda^{1/3}}{2\pi\mathrm{i}}\int \mathrm{d}x\oint \mathrm{d}y e^{\lambda^{2/3}\left(-y+1/(2y^2)+x-1/(2x^2)\right)}\Big(\frac{1}{x^3}+\frac{1}{x^2y}+\frac{1}{xy^2}\Big).$$

The resulting 1-point function, given by $\pi^{-1}\Im q_0(\lambda)$, coincides precisely with $K(\lambda,\lambda)$ and is plotted in Figure 8.7.

References

[1] A. Adhikari and J. Huang, *Dyson Brownian motion for general β and potential at the edge*, Probab. Theory Related Fields **178**, 893–950 (2020), MR4168391.

[2] M. Adler, P. L. Ferrari, and P. van Moerbeke, *Airy processes with wanderers and new universality classes*, Ann. Probab. **38**, 714–769 (2010), MR2642890.

[3] I. Afanasiev, *On the correlation functions of the characteristic polynomials of non-Hermitian random matrices with independent entries*, J. Stat. Phys. **176**, 1561–1582 (2019), MR4001834.

[4] O. H. Ajanki, L. Erdős, and T. Krüger, *Singularities of solutions to quadratic vector equations on the complex upper half-plane*, Comm. Pure Appl. Math. **70**, 1672–1705 (2017), MR3684307.

[5] O. H. Ajanki, L. Erdős, and T. Krüger, *Stability of the matrix Dyson equation and random matrices with correlations*, Probab. Theory Related Fields **173**, 293–373 (2019), MR3916109.

[6] O. H. Ajanki, L. Erdős, and T. Krüger, *Universality for general Wigner-type matrices*, Probab. Theory Related Fields **169**, 667–727 (2017), MR3719056.

[7] O. H. Ajanki, L. Erdős, and T. Krüger, *Quadratic vector equations on complex upper half-plane*, Mem. Amer. Math. Soc. **261**, v+133 (2019), MR4031100.

[8] G. Akemann, J. Osborn, K. Splittorff, and J. Verbaarschot, *Unquenched QCD Dirac operator spectra at nonzero baryon chemical potential*, Nuclear Phys. B **712**, 287–324 (2005), MR2128558.

[9] A. Alldridge and Z. Shaikh, *Superbosonization via Riesz superdistributions*, Forum Math. Sigma **2**, e9, 64 (2014), MR3264247.

[10] J. Alt, *Singularities of the density of states of random Gram matrices*, Electron. Commun. Probab. **22** (2017).

[11] J. Alt, L. Erdős, and T. Krüger, *Local inhomogeneous circular law*, Ann. Appl. Probab. **28**, 148–203 (2018), MR3770875.

[12] J. Alt, L. Erdős, and T. Krüger, *Local law for random Gram matrices*, Electron. J. Probab. **22** (2017).

[13] J. Alt, L. Erdős, and T. Krüger, *Spectral radius of random matrices with independent entries*, preprint (2019), arXiv:1907.13631.

[14] J. Alt, L. Erdős, and T. Krüger, *The Dyson equation with linear self-energy: spectral bands, edges and cusps*, Doc. Math. **25**, 1421–1540 (2020), MR4164728.

[15] J. Alt, L. Erdős, T. Krüger, and D. Schröder, *Correlated random matrices: Band rigidity and edge universality*, Ann. Probab. **48**, 963–1001 (2020), MR4089499.

[16] J. Alt and T. Krüger, *Inhomogeneous Circular Law for Correlated Matrices*, arXiv preprint arXiv:2005.13533 (2020).

[17] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, Vol. 118, Cambridge Studies in Advanced Mathematics (Cambridge University Press, Cambridge, 2010), pp. xiv+492, MR2760897.

[18] Z. D. Bai, *Circular law*, Ann. Probab. **25**, 494–529 (1997), MR1428519.

[19] Z. D. Bai and J. Yao, *On the convergence of the spectral empirical process of Wigner matrices*, Bernoulli **11**, 1059–1092 (2005), MR2189081.

[20] Z. D. Bai and Y. Q. Yin, *Limiting behavior of the norm of products of random matrices and two problems of Geman-Hwang*, Probab. Theory Related Fields **73**, 555–569 (1986), MR863545.

[21] Z. Bai and J. W. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of information-plus-noise type matrices*, Random Matrices Theory Appl. **1**, 1150004, 44 (2012), MR2930382.

[22] Z. D. Bai and J. W. Silverstein, *CLT for linear spectral statistics of large-dimensional sample covariance matrices*, Ann. Probab., 553–605 (2004).

[23] Z. Bao and L. Erdős, *Delocalization for a class of random block band matrices*, Probab. Theory Related Fields **167**, 673–776 (2017), MR3627427.

[24] Z. Bao, K. Schnelli, and Y. Xu, *Central limit theorem for mesoscopic eigenvalue statistics of the free sum of matrices*, preprint (2020), arXiv:2001.07661.

[25] R. Bauerschmidt, P. Bourgade, M. Nikula, and H.-T. Yau, *The two-dimensional Coulomb plasma: quasi-free approximation and central limit theorem*, Adv. Theor. Math. Phys. **23**, 841–1002 (2019), MR4063572.

[26] F. Bekerman, T. Leblé, and S. Serfaty, *CLT for fluctuations of β-ensembles with general potential*, Electron. J. Probab. **23**, Paper no. 115, 31 (2018), MR3885548.

[27] F. Bekerman and A. Lodhia, *Mesoscopic central limit theorem for general β-ensembles*, Ann. Inst. Henri Poincaré Probab. Stat. **54**, 1917–1938 (2018), MR3865662.

[28] G. Ben Arous and S. Péché, *Universality of local eigenvalue statistics for some sample covariance matrices*, Comm. Pure Appl. Math. **58**, 1316–1357 (2005), MR2162782.

[29] F. A. Berezin, *Introduction to superanalysis*, Vol. 9, Mathematical Physics and Applied Mathematics, Edited and with a foreword by A. A. Kirillov, With an appendix by V. I. Ogievetsky, Translated from the Russian by J. Niederle and R. Kotecký, Translation edited by Dimitri Leĭtes (D. Reidel Publishing Co., Dordrecht, 1987), pp. xii+424, MR914369.

[30] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Isotropic local laws for sample covariance and generalized Wigner matrices*, Electron. J. Probab **19**, 1–53 (2014).

[31] A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin, *On the principal components of sample covariance matrices*, Probab. Th. Related Fields **164**, 459–552 (2016).

[32] O. Bohigas, M.-J. Giannoni, and C. Schmit, *Characterization of chaotic quantum spectra and universality of level fluctuation laws*, Phys. Rev. Lett. **52**, 1–4 (1984), MR730191.

[33] C. Bordenave, P. Caputo, D. Chafaï, and K. Tikhomirov, *On the spectral radius of a random matrix: An upper bound without fourth moment*, Ann. Probab. **46**, 2268–2286 (2018), MR3813992.

[34] C. Bordenave and D. Chafaï, *Around the circular law*, Probab. Surv. **9**, 1–89 (2012), MR2908617.

[35] A. Borodin and C. D. Sinclair, *The Ginibre ensemble of real random matrices and its scaling limits*, Comm. Math. Phys. **291**, 177–224 (2009), MR2530159.

[36] A. Borodin, *CLT for spectra of submatrices of Wigner random matrices*, Mosc. Math. J. **14**, 29–38 (2014).

[37] G. Borot and A. Guionnet, *Asymptotic expansion of beta matrix models in the multi-cut regime*, preprint (2013), arXiv:1303.1045.

[38] P. Bourgade and G. Dubach, *The distribution of overlaps between eigenvectors of Ginibre matrices*, Probab. Theory Related Fields **177**, 397–464 (2020), MR4095019.

[39] P. Bourgade and H.-T. Yau, *The eigenvector moment flow and local quantum unique ergodicity*, Comm. Math. Phys. **350**, 231–278 (2017), MR3606475.

[40] P. Bourgade, *Extreme gaps between eigenvalues of Wigner matrices*, preprint (2018), arXiv:1812.10376.

[41] P. Bourgade, L. Erdős, and H.-T. Yau, *Edge universality of beta ensembles*, Comm. Math. Phys. **332**, 261–353 (2014), MR3253704.

[42] P. Bourgade, L. Erdős, H.-T. Yau, and J. Yin, *Fixed energy universality for generalized Wigner matrices*, Comm. Pure Appl. Math. **69**, 1815–1881 (2016), MR3541852.

[43] P. Bourgade, L. Erdős, H.-T. Yau, and J. Yin, *Universality for a class of random band matrices*, Adv. Theor. Math. Phys. **21**, 739–800 (2017), MR3695802.

[44] P. Bourgade, H.-T. Yau, and J. Yin, *Local circular law for random matrices*, Probab. Theory Related Fields **159**, 545–595 (2014), MR3230002.

[45] P. Bourgade, H.-T. Yau, and J. Yin, *Random band matrices in the delocalized phase, I: Quantum unique ergodicity and universality*, preprint (2018), arXiv:1807.01559.

[46] P. Bourgade, H.-T. Yau, and J. Yin, *The local circular law II: The edge case*, Probab. Theory Related Fields **159**, 619–660 (2014), MR3230004.

[47] A. Boutet de Monvel and A. Khorunzhy, *Asymptotic distribution of smoothed eigenvalue density. I. Gaussian random matrices*, Random Oper. Stochastic Equations **7**, 1–22 (1999), MR1678012.

[48] A. Boutet de Monvel and A. Khorunzhy, *Asymptotic distribution of smoothed eigenvalue density. II. Wigner random matrices*, Random Oper. Stochastic Equations **7**, 149–168 (1999), MR1689027.

[49] E. Brézin and S. Hikami, *Level spacing of random matrices in an external source*, Phys. Rev. E (3) **58**, 7176–7185 (1998), MR1662382.

[50] E. Brézin and S. Hikami, *Universal singularity at the closure of a gap in a random matrix theory*, Phys. Rev. E (3) **57**, 4140–4149 (1998), MR1618958.

[51] D. Chafaï, *Around the circular law: An update*, (2018) http://djalil.chafai.net/blog/2018/11/04/around-the-circular-law-an-update/ (visited on 12/03/2018).

[52] J. T. Chalker and B. Mehlig, *Eigenvector statistics in non-Hermitian random matrix ensembles*, Phys. Rev. Lett. **81**, 3367–3370 (1998).

[53] Z. Che and B. Landon, *Local spectral statistics of the addition of random matrices*, Probab. Theory Related Fields **175**, 579–654 (2019), MR4009717.

[54] Z. Che and P. Lopatto, *Universality of the least singular value for sparse random matrices*, Electron. J. Probab. **24**, Paper No. 9, 53 (2019), MR3916329.

[55] Z. Che and P. Lopatto, *Universality of the least singular value for the sum of random matrices*, preprint (2019), arXiv:1908.04060.

[56] G. Cipolloni and L. Erdős, *Fluctuations for differences of linear eigenvalue statistics for sample covariance matrices*, Random Matrices: Theory and Applications **9**, 2050006 (2020).

[57] G. Cipolloni, L. Erdős, T. Krüger, and D. Schröder, *Cusp universality for random matrices, II: The real symmetric case*, Pure Appl. Anal. **1**, 615–707 (2019), MR4026551.

[58] G. Cipolloni, L. Erdős, and D. Schröder, *Central limit theorem for linear eigenvalue statistics of non-Hermitian random matrices*, Accepted to Communications on Pure and Applied Mathematics (2020), arXiv:1912.04100.

[59] G. Cipolloni, L. Erdős, and D. Schröder, *Edge universality for non-Hermitian random matrices*, Probability Theory and Related Fields, 1–28 (2020).

[60] G. Cipolloni, L. Erdős, and D. Schröder, *Fluctuation around the circular law for random matrices with real entries*, preprint (2020), arXiv:2002.02438.

[61] G. Cipolloni, L. Erdős, and D. Schröder, *Optimal lower bound on the least singular value of the shifted ginibre ensemble*, Accepted to Probability and Mathemtical Physics (2020), arXiv:1908.01653.

[62] T. Claeys, A. B. J. Kuijlaars, K. Liechty, and D. Wang, *Propagation of Singular Behavior for Gaussian Perturbations of Random Matrices*, Comm. Math. Phys. **362**, 1–54 (2018), MR3833603.

[63] N. Cook, *Lower bounds for the smallest singular value of structured random matrices*, Ann. Probab. **46**, 3442–3500 (2018), MR3857860.

[64] O. Costin and J. L. Lebowitz, *Gaussian fluctuation in random matrices*, Phys. Rev. Lett. **75**, 69–72 (1995), MR3155254.

[65] N. Coston and S. O'Rourke, *Gaussian fluctuations for linear eigenvalue statistics of products of independent iid random matrices*, J. Theoret. Probab. **33**, 1541–1612 (2020), MR4125967.

[66] R. Couillet and M. Debbah, *Random matrix methods for wireless communications* (Cambridge University Press, Cambridge, 2011), pp. xxii+539, MR2884783.

[67] P. Desrosiers and P. J. Forrester, *Asymptotic correlations for Gaussian and Wishart matrices with external source*, Int. Math. Res. Not., Art. ID 27395, 43 (2006), MR2250019.

[68] M. Disertori, H. Pinson, and T. Spencer, *Density of states for random band matrices*, Comm. Math. Phys. **232**, 83–124 (2002), MR1942858.

[69] M. Disertori and M. Lager, *Density of states for random band matrices in two dimensions*, Ann. Henri Poincaré **18**, 2367–2413 (2017), MR3665217.

[70] M. Disertori, M. Lohmann, and S. Sodin, *The density of states of 1D random band matrices via a supersymmetric transfer operator*, preprint (2018), arXiv:1810.13150.

[71] R. B. Dozier and J. W. Silverstein, *On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices*, J. Multivariate Anal. **98**, 678–694 (2007), MR2322123.

[72] M. Duits and K. Johansson, *On mesoscopic equilibrium for linear statistics in Dyson's Brownian motion*, Mem. Amer. Math. Soc. **255**, v+118 (2018), MR3852256.

[73] I. Dumitriu and E. Paquette, *Spectra of overlapping Wishart matrices and the Gaussian free field*, Random Matrices: Theory and Applications **7**, 1850003 (2018).

[74] F. J. Dyson, *A Brownian-motion model for the eigenvalues of a random matrix*, J. Mathematical Phys. **3**, 1191–1198 (1962), MR148397.

[75] A. Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl. **9**, 543–560 (1988), MR964668.

[76] A. Edelman, *The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law*, J. Multivariate Anal. **60**, 203–232 (1997), MR1437734.

[77] A. Edelman, E. Kostlan, and M. Shub, *How many eigenvalues of a random matrix are real?*, J. Amer. Math. Soc. **7**, 247–267 (1994), MR1231689.

[78] A. Edelman and N. R. Rao, *Random matrix theory*, Acta Numer. **14**, 233–297 (2005), MR2168344.

[79] K. Efetov, *Supersymmetry in disorder and chaos* (Cambridge University Press, Cambridge, 1997), pp. xiv+441, MR1628498.

[80] L. Erdős, A. Knowles, and H.-T. Yau, *Averaging fluctuations in resolvents of random band matrices*, **14**, 1837–1926 (2013).

[81] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *The local semicircle law for a general class of random matrices*, Electron. J. Probab. **18**, no. 59, 58 (2013), MR3068390.

[82]L. Erdős, T. Krüger, and D. Renfrew, *Power law decay for systems of randomly coupled differential equations*, SIAM J. Math. Anal. **50**, 3271–3290 (2018), MR3816180.

[83]L. Erdős, T. Krüger, and D. Schröder, *Cusp universality for random matrices I: local law and the complex Hermitian case*, Comm. Math. Phys. **378**, 1203–1278 (2020), MR4134946.

[84]L. Erdős, T. Krüger, and D. Schröder, *Random matrices with slow correlation decay*, Forum Math. Sigma **7**, e8, 89 (2019), MR3941370.

[85]L. Erdős, S. Péché, J. A. Ramírez, B. Schlein, and H.-T. Yau, *Bulk universality for Wigner matrices*, Comm. Pure Appl. Math. **63**, 895–925 (2010), MR2662426.

[86]L. Erdős, B. Schlein, and H.-T. Yau, *Universality of random matrices and local relaxation flow*, Invent. Math. **185**, 75–119 (2011), MR2810797.

[87]L. Erdős, B. Schlein, H.-T. Yau, and J. Yin, *The local relaxation flow approach to universality of the local statistics for random matrices*, Ann. Inst. Henri Poincaré Probab. Stat. **48**, 1–46 (2012), MR2919197.

[88]L. Erdős and K. Schnelli, *Universality for random matrix flows with time-dependent density*, Ann. Inst. Henri Poincaré Probab. Stat. **53**, 1606–1656 (2017), MR3729630.

[89]L. Erdős and D. Schröder, *Fluctuations of rectangular Young diagrams of interlacing Wigner eigenvalues*, Int. Math. Res. Not. IMRN, 3255–3298 (2018), MR3805203.

[90]L. Erdős and H.-T. Yau, *A dynamical approach to random matrix theory*, Vol. 28, Courant Lecture Notes in Mathematics (Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017), pp. ix+226, MR3699468.

[91]L. Erdős and H.-T. Yau, *Gap universality of generalized Wigner and β-ensembles*, J. Eur. Math. Soc. (JEMS) **17**, 1927–2036 (2015), MR3372074.

[92]L. Erdős, H.-T. Yau, and J. Yin, *Bulk universality for generalized Wigner matrices*, Probab. Theory Related Fields **154**, 341–407 (2012), MR2981427.

[93]L. Erdős, H.-T. Yau, and J. Yin, *Rigidity of eigenvalues of generalized Wigner matrices*, Adv. Math. **229**, 1435–1515 (2012), MR2871147.

[94]P. Erdős and A. Hajnal, *On chromatic number of graphs and set-systems*, Acta Math. Acad. Sci. Hungar. **17**, 61–99 (1966), MR193025.

[95]P. J. Forrester, *Fluctuation formula for complex random matrices*, J. Phys. A **32**, L159–L163 (1999), MR1687948.

[96]P. J. Forrester, *The spectrum edge of random matrix ensembles*, Nuclear Phys. B **402**, 709–728 (1993), MR1236195.

[97]P. Forrester and T. Nagao, *Eigenvalue statistics of the real Ginibre ensemble*, Phys. Rev. Lett. **99**, 050603 (2007), PMID17930739.

[98]P. J. Forrester and E. M. Rains, *Matrix averages relating to Ginibre ensembles*, Journal of Physics A: Mathematical and Theoretical **42**, 385205 (2009).

[99]Y. V. Fyodorov, *On statistics of bi-orthogonal eigenvectors in real and complex Ginibre ensembles: combining partial Schur decomposition with supersymmetry*, Comm. Math. Phys. **363**, 579–603 (2018), MR3851824.

[100]Y. V. Fyodorov, J. Grela, and E. Strahov, *On characteristic polynomials for a generalized chiral random matrix ensemble with a source*, J. Phys. A **51**, 134003, 30 (2018), MR3780342.

[101]S. Geman, *The spectral radius of large random matrices*, Ann. Probab. **14**, 1318–1328 (1986), MR866352.

[102]J. Ginibre, *Statistical ensembles of complex, quaternion, and real matrices*, J. Mathematical Phys. **6**, 440–449 (1965), MR173726.

[103]V. L. Girko, *The circular law*, Teor. Veroyatnost. i Primenen. **29**, 669–679 (1984), MR773436.

[104]V. Gorin and L. Zhang, *Interlacing adjacent levels of β-Jacobi corners processes*, Probab. Theory and Rel. Fields (2016).

[105]F. Götze and A. Tikhomirov, *The circular law for random matrices*, Ann. Probab. **38**, 1444–1491 (2010), MR2663633.

[106]T. Guhr, "Supersymmetry", in *The Oxford handbook of random matrix theory* (Oxford Univ. Press, Oxford, 2011), pp. 135–154, MR2932627.

[107]T. Guhr and T. Wettig, *An Itzykson-Zuber-like integral and diffusion for complex ordinary and supermatrices*, J. Math. Phys. **37**, 6395–6413 (1996), MR1419177.

[108]T. Guhr and T. Wettig, *Universal spectral correlations of the Dirac operator at finite temperature*, Nuclear Phys. B **506**, 589–611 (1997), MR1488590.

[109]W. Hachem, A. Hardy, and J. Najim, *Large complex correlated Wishart matrices: The Pearcey kernel and expansion at the hard edge*, Electron. J. Probab. **21**, Paper No. 1, 36 (2016), MR3485343.

[110]Y. He, *Mesoscopic linear statistics of Wigner matrices of mixed symmetry class*, J. Stat. Phys. **175**, 932–959 (2019), MR3959983.

[111]Y. He and A. Knowles, *Mesoscopic eigenvalue density correlations of Wigner matrices*, Probab. Theory Related Fields **177**, 147–216 (2020), MR4095015.

[112]Y. He and A. Knowles, *Mesoscopic eigenvalue statistics of Wigner matrices*, Ann. Appl. Probab. **27**, 1510–1550 (2017), MR3678478.

[113] J. W. Helton, R. R. Far, and R. Speicher, *Operator-valued semicircular elements: solving a quadratic matrix equation with positivity constraints*, International Mathematics Research Notices **2007**, rnm086–rnm086 (2007).

[114] J. Huang and B. Landon, *Rigidity and a mesoscopic central limit theorem for Dyson Brownian motion for general β and potentials*, Probab. Theory Related Fields **175**, 209–253 (2019), MR4009708.

[115] J. Huang, B. Landon, and H.-T. Yau, *Bulk universality of sparse random matrices*, J. Math. Phys. **56**, 123301, 19 (2015), MR3429490.

[116] A. Jackson, M. Şener, and J. Verbaarschot, *Finite volume partition functions and Itzykson-Zuber integrals*, Phys. Lett. B **387**, 355–360 (1996), MR1413913.

[117] K. Johansson, *On fluctuations of eigenvalues of random Hermitian matrices*, Duke Math. J. **91**, 151–204 (1998), MR1487983.

[118] K. Johansson, *Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices*, Comm. Math. Phys. **215**, 683–705 (2001), MR1810949.

[119] D. Jonsson, *Some limit theorems for the eigenvalues of a sample covariance matrix*, J. Multivariate Anal. **12**, 1–38 (1982).

[120] O. Kallenberg, *Foundations of modern probability*, Second edition, Probability and its Applications (New York) (Springer-Verlag, New York, 2002), pp. xx+638, MR1876169.

[121] E. Kanzieper and G. Akemann, *Statistics of real eigenvalues in Ginibre's ensemble of random real matrices*, Phys. Rev. Lett. **95**, 230201, 4 (2005), MR2185860.

[122] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, Vol. 113, Graduate Texts in Mathematics (Springer-Verlag, New York, 1988), pp. xxiv+470, MR917065.

[123] J. Keating, "The Riemann zeta-function and quantum chaology", in *Quantum chaos (Varenna, 1991)*, Proc. Internat. School of Phys. Enrico Fermi, CXIX (North-Holland, Amsterdam, 1993), pp. 145–185, MR1246830.

[124] A. M. Khorunzhy, B. A. Khoruzhenko, and L. A. Pastur, *Asymptotic properties of large random matrices with independent entries*, J. Math. Phys. **37**, 5033–5060 (1996), MR1411619.

[125] M. Kieburg, H.-J. Sommers, and T. Guhr, *A comparison of the superbosonization formula and the generalized Hubbard-Stratonovich transformation*, J. Phys. A **42**, 275206, 23 (2009), MR2512124.

[126] P. Kopel, *Linear statistics of non-Hermitian matrices matching the real or complex Ginibre ensemble to four moments*, preprint (2015), arXiv:1510.02987.

[127] G. Lambert, M. Ledoux, and C. Webb, *Quantitative normal approximation of linear statistics of β-ensembles*, Ann. Probab. **47**, 2619–2685 (2019), MR4021234.

[128] B. Landon and P. Sosoe, *Applications of mesoscopic CLTs in random matrix theory*, preprint (2018), arXiv:1811.05915.

[129] B. Landon, P. Sosoe, and H.-T. Yau, *Fixed energy universality of Dyson Brownian motion*, Adv. Math. **346**, 1137–1332 (2019), MR3914908.

[130] B. Landon and H.-T. Yau, *Convergence of local statistics of Dyson Brownian motion*, Comm. Math. Phys. **355**, 949–1000 (2017), MR3687212.

[131] B. Landon and H.-T. Yau, *Edge statistics of Dyson Brownian motion*, preprint (2017), arXiv:1712.03881.

[132] T. Leblé and S. Serfaty, *Fluctuations of two dimensional Coulomb gases*, Geom. Funct. Anal. **28**, 443–508 (2018), MR3788208.

[133] J. O. Lee and K. Schnelli, *Edge universality for deformed Wigner matrices*, Rev. Math. Phys. **27**, 1550018, 94 (2015), MR3405746.

[134] J. O. Lee and K. Schnelli, *Local law and Tracy-Widom limit for sparse random matrices*, Probab. Theory Related Fields **171**, 543–616 (2018), MR3800840.

[135] J. O. Lee and K. Schnelli, *Tracy-Widom distribution for the largest eigenvalue of real sample covariance matrices with general population*, Ann. Appl. Probab. **26**, 3786–3839 (2016), MR3582818.

[136] J. O. Lee, K. Schnelli, B. Stetler, and H.-T. Yau, *Bulk universality for deformed Wigner matrices*, Ann. Probab. **44**, 2349–2425 (2016), MR3502606.

[137] N. Lehmann and H.-J. Sommers, *Eigenvalue statistics of random real matrices*, Phys. Rev. Lett. **67**, 941–944 (1991), MR1121461.

[138] Y. Li, K. Schnelli, and Y. Xu, *Central limit theorem for mesoscopic eigenvalue statistics of deformed Wigner matrices and sample covariance matrices*, preprint (2019), arXiv:1909.12821.

[139] Y. Li and Y. Xu, *On fluctuations of global and mesoscopic linear eigenvalue statistics of generalized Wigner matrices*, preprint (2020), arXiv:2001.08725.

[140] D. R. Lick and A. T. White, *k-degenerate graphs*, Canadian J. Math. **22**, 1082–1096 (1970), MR266812.

[141] E. H. Lieb and M. Loss, *Analysis*, Second, Vol. 14, Graduate Studies in Mathematics (American Mathematical Society, Providence, RI, 2001), pp. xxii+346, MR1817225.

[142] P. Littelmann, H.-J. Sommers, and M. Zirnbauer, *Superbosonization of invariant random matrix ensembles*, Comm. Math. Phys. **283**, 343–395 (2008), MR2430637.

[143] A. Lytova and L. Pastur, *Central limit theorem for linear eigenvalue statistics of random matrices with independent entries*, Ann. Probab. **37**, 1778–1840 (2009), MR2561434.

[144] V. A. Marčenko and L. A. Pastur, *Distribution of eigenvalues for some sets of random matrices*, Math. USSR Sbornik **1**, 457 (1967).

[145] R. May, *Will a large complex system be stable?*, Nature **238**, 413–4 (1972), PMID4559589.

[146] M. L. Mehta, *Random matrices and the statistical theory of energy levels* (Academic Press, New York-London, 1967), pp. x+259, MR0220494.

[147] M. L. Mehta and M. Gaudin, *On the density of eigenvalues of a random matrix*, Nuclear Phys. **18**, 420–427 (1960), MR0112895.

[148] M. Y. Mo, *Rank 1 real Wishart spiked model*, Comm. Pure Appl. Math. **65**, 1528–1638 (2012), MR2969495.

[149] A. B. de Monvel and A. Khorunzhy, *Asymptotic distribution of smoothed eigenvalue density. II. Wigner random matrices*, Random Operators and Stochastic Equations **7**, 149–168 (1999).

[150] H. H. Nguyen and V. Vu, *Random matrices: law of the determinant*, Ann. Probab. **42**, 146–167 (2014), MR3161483.

[151] I. Nourdin and G. Peccati, *Universal Gaussian fluctuations of non-Hermitian matrix ensembles: from weak convergence to almost sure CLTs*, ALEA Lat. Am. J. Probab. Math. Stat. **7**, 341–375 (2010), MR2738319.

[152] S. O'Rourke and D. Renfrew, *Central limit theorem for linear eigenvalue statistics of elliptic random matrices*, J. Theoret. Probab. **29**, 1121–1191 (2016), MR3540493.

[153] J. Osborn, D. Toublan, and J. Verbaarschot, *From chiral random matrix theory to chiral perturbation theory*, Nuclear Physics B **540**, 317–344 (1999).

[154] G. Pan and W. Zhou, *Circular law, extreme singular values and potential theory*, J. Multivariate Anal. **101**, 645–656 (2010), MR2575411.

[155] L. Pastur and M. Shcherbina, *Bulk universality and related properties of Hermitian matrix models*, J. Stat. Phys. **130**, 205–250 (2008), MR2375744.

[156] L. Pastur and M. Shcherbina, *On the edge universality of the local eigenvalue statistics of matrix models*, Mat. Fiz. Anal. Geom. **10**, 335–365 (2003), MR2012268.

[157] T. Pearcey, *The structure of an electromagnetic field in the neighbourhood of a cusp of a caustic*, Philos. Mag. (7) **37**, 311–317 (1946), MR0020857.

[158] K. Rajan and L. Abbott, *Eigenvalue spectra of random matrices for neural networks*, Phys. Rev. Lett. **97**, 188104 (2006), PMID17155583.

[159] N. R. Rao, J. A. Mingo, R. Speicher, and A. Edelman, *Statistical eigen-inference from large Wishart matrices*, Ann. Stat. **36**, 2850–2885 (2008).

[160] C. Recher, M. Kieburg, T. Guhr, and M. Zirnbauer, *Supersymmetry approach to Wishart correlation matrices: Exact results*, J. Stat. Phys. **148**, 981–998 (2012), MR2975518.

[161] B. Rider, *Deviations from the circular law*, Probab. Theory Related Fields **130**, 337–367 (2004), MR2095933.

[162] B. Rider and J. W. Silverstein, *Gaussian fluctuations for non-Hermitian random matrix ensembles*, Ann. Probab. **34**, 2118–2143 (2006), MR2294978.

[163] B. Rider and B. Virág, *Complex determinantal processes and H^1 noise*, Electron. J. Probab. **12**, no. 45, 1238–1257 (2007), MR2346510.

[164] B. Rider and B. Virág, *The noise in the circular law and the Gaussian free field*, Int. Math. Res. Not. IMRN, Art. ID rnm006, 33 (2007), MR2361453.

[165] M. Rudelson, *Invertibility of random matrices: Norm of the inverse*, Ann. of Math. (2) **168**, 575–600 (2008), MR2434885.

[166] M. Rudelson, R. Vershynin, et al., *Delocalization of eigenvectors of random matrices with independent entries*, Duke Mathematical Journal **164**, 2507–2538 (2015).

[167] M. Rudelson and R. Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218**, 600–633 (2008), MR2407948.

[168] A. Sankar, D. A. Spielman, and S.-H. Teng, *Smoothed analysis of the condition numbers and growth factors of matrices*, SIAM J. Matrix Anal. Appl. **28**, 446–476 (2006), MR2255338.

[169] M. Shcherbina, *Central limit theorem for linear eigenvalue statistics of the Wigner and sample covariance random matrices*, Zh. Mat. Fiz. Anal. Geom. **7**, 176–192, 197, 199 (2011), MR2829615.

[170] M. Shcherbina, *Fluctuations of linear eigenvalue statistics of β matrix models in the multi-cut regime*, J. Stat. Phys. **151**, 1004–1034 (2013), MR3063494.

[171] M. Shcherbina and T. Shcherbina, *Characteristic polynomials for 1D random band matrices from the localization side*, Comm. Math. Phys. **351**, 1009–1044 (2017), MR3623245.

[172] M. Shcherbina and T. Shcherbina, *Transfer matrix approach to 1D random band matrices: Density of states*, J. Stat. Phys. **164**, 1233–1260 (2016), MR3541181.

[173] M. Shcherbina and T. Shcherbina, *Universality for 1D random band matrices*, preprint (2019), arXiv:1910.02999.

[174] T. Shcherbina, *On the correlation functions of the characteristic polynomials of the Hermitian sample covariance matrices*, Probab. Theory Related Fields **156**, 449–482 (2013), MR3055265.

[175] T. Shcherbina, *Characteristic polynomials for random band matrices near the threshold*, preprint (2019), arXiv:1905.08136.

[176] T. Shcherbina, *On the correlation function of the characteristic polynomials of the Hermitian Wigner ensemble*, Comm. Math. Phys. **308**, 1–21 (2011), MR2842968.

[177] T. Shcherbina, *On the second mixed moment of the characteristic polynomials of 1D band matrices*, Comm. Math. Phys. **328**, 45–82 (2014), MR3196980.

[178] S. Sheffield, *Gaussian free fields for mathematicians*, Probab. Theory Related Fields **139**, 521–541 (2007), MR2322706.

[179] N. J. Simm, *Central limit theorems for the real eigenvalues of large Gaussian random matrices*, Random Matrices Theory Appl. **6**, 1750002, 18 (2017), MR3612267.

[180] Y. G. Sinai and A. B. Soshnikov, *A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge for random symmetric matrices*, Functional Analysis and Its Applications **32**, 114 (1998).

[181] S. Sodin, *The spectral edge of some random band matrices*, Ann. of Math. (2) **172**, 2223–2251 (2010), MR2726110.

[182] H.-J. Sommers and B. A. Khoruzhenko, *Schur function averages for the real Ginibre ensemble*, J. Phys. A **42**, 222002, 8 (2009), MR2515561.

[183] H.-J. Sommers and W. Wieczorek, *General eigenvalue correlations for the real Ginibre ensemble*, J. Phys. A **41**, 405003, 24 (2008), MR2439268.

[184] H. Sompolinsky, A. Crisanti, and H. Sommers, *Chaos in random neural networks*, Phys. Rev. Lett. **61**, 259–262 (1988), PMID10039285.

[185] A. Soshnikov, *Gaussian limit for determinantal random point fields*, Ann. Probab. **30**, 171–187 (2002), MR1894104.

[186] A. Soshnikov, *Universality at the edge of the spectrum in Wigner random matrices*, Comm. Math. Phys. **207**, 697–733 (1999), MR1727234.

[187] P. Sosoe and P. Wong, *Regularity conditions in the CLT for linear eigenvalue statistics of Wigner matrices*, Adv. Math. **249**, 37–87 (2013), MR3116567.

[188] T. Spencer, "Random banded and sparse matrices", in *The Oxford handbook of random matrix theory* (Oxford Univ. Press, Oxford, 2011), pp. 471–488, MR2932643.

[189] E. M. Stein and G. Weiss, *Fractional integrals on n-dimensional Euclidean space*, J. Math. Mech. **7**, 503–514 (1958), MR0098285.

[190] M. Stephanov, *Random matrix model of QCD at finite density and the nature of the quenched limit*, Phys Rev Lett. **76**, 4472–4475 (1996), PMID10061300.

[191] T. Tao and V. Vu, *Random matrices: The circular law*, Commun. Contemp. Math. **10**, 261–307 (2008), MR2409368.

[192] T. Tao and V. Vu, *Random matrices: The distribution of the smallest singular values*, Geom. Funct. Anal. **20**, 260–297 (2010), MR2647142.

[193] T. Tao and V. Vu, *Random matrices: Universality of local eigenvalue statistics*, Acta Math. **206**, 127–204 (2011), MR2784665.

[194] T. Tao and V. Vu, *Random matrices: Universality of local eigenvalue statistics up to the edge*, Comm. Math. Phys. **298**, 549–572 (2010), MR2669449.

[195] T. Tao and V. Vu, *Random matrices: universality of local spectral statistics of non-Hermitian matrices*, Ann. Probab. **43**, 782–874 (2015), MR3306005.

[196] T. Tao and V. Vu, *Smooth analysis of the condition number and the least singular value*, Math. Comp. **79**, 2333–2352 (2010), MR2684367.

[197] T. Tao and V. Vu, "The condition number of a randomly perturbed matrix", in *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing* (ACM, New York, 2007), pp. 248–255, MR2402448.

[198] T. Tao, V. Vu, M. Krishnapur, et al., *Random matrices: Universality of ESDs and the circular law*, The Annals of Probability **38**, 2023–2065 (2010).

[199] T. Tao and V. H. Vu, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, Ann. of Math. (2) **169**, 595–632 (2009), MR2480613.

[200] N. M. Temme, *Uniform asymptotics for the incomplete gamma functions starting from negative values of the parameters*, Methods Appl. Anal. **3**, 335–344 (1996), MR1421474.

[201] K. Tikhomirov, *Invertibility via distance for non-centered random matrices with continuous distributions*, preprint (2017), arXiv:1707.09656.

[202] C. A. Tracy and H. Widom, *Level-spacing distributions and the Airy kernel*, Comm. Math. Phys. **159**, 151–174 (1994), MR1257246.

[203] C. A. Tracy and H. Widom, *On orthogonal and symplectic matrix ensembles*, Comm. Math. Phys. **177**, 727–754 (1996), MR1385083.

[204] C. A. Tracy and H. Widom, *The Pearcey process*, Comm. Math. Phys. **263**, 381–400 (2006), MR2207649.

[205] J. Verbaarschot, H. Weidenmüller, and M. Zirnbauer, *Grassmann integration in stochastic quantum physics: The case of compound-nucleus scattering*, Phys. Rep. **129**, 367–438 (1985), MR820690.

[206] J. Verbaarschot and I. Zahed, *Spectral density of the QCD Dirac operator near zero virtuality*, Phys Rev Lett. **70**, 3852–3855 (1993), PMID10053982.

[207] J. Verbaarschot and T. Wettig, *Random matrix theory and chiral symmetry in QCD*, Annu. Rev. Nucl. Part. Sci. **50**, 343–410 (2000).

[208] H. Wang, *Quantitative universality for the largest eigenvalue of sample covariance matrices*, preprint (2019), arXiv:1912.05473.

[209] E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math. (2) **62**, 548–564 (1955), MR0077805.

[210] E. P. Wigner, *On the distribution of the roots of certain symmetric matrices*, Ann. of Math. (2) **67**, 325–327 (1958), MR0095527.

[211] T. Wilke, T. Guhr, and T. Wettig, *Microscopic spectrum of the QCD Dirac operator with finite quark masses*, Phys. Rev. D **57**, 6486–6495 (1998).

[212] J. Wishart, *The generalised product moment distribution in samples from a normal multivariate population*, Biometrika, 32–52 (1928).

[213] J. Yin, *The local circular law III: General case*, Probab. Theory Related Fields **160**, 679–732 (2014), MR3278919.