

hexagon:2610.00675v1

Smallest Gaps between Eigenvalues of Wigner Matrices without Smoothness Assumptions

Patrick Lopatto

AUTHOR'S NOTE

With the exception of this note, the entire document is machine-written. I have carefully supervised its preparation to ensure that the proofs are readable and that previous literature is appropriately cited. I welcome suggestions concerning mathematical corrections, omitted citations, and attribution of ideas.

The author was partially supported by NSF grant DMS-2450004.

ABSTRACT

We determine the asymptotic distribution of the smallest gaps between eigenvalues in the bulk of the spectrum of real symmetric Wigner matrices, including Bernoulli matrices. After multiplication by $N^{3/2}$, the gaps and their locations converge to a Poisson point process with the same intensity as for the Gaussian orthogonal ensemble. Since changing a single entry can move eigenvalues by an amount comparable to the smallest gaps, standard perturbative comparison estimates do not directly resolve this scale for arbitrary entry distributions. Instead, we use that resampling a few random entries does not change the law of the matrix, and we show that the resampling moves each pair of nearby eigenvalues by a random walk that satisfies a local limit theorem; this requires excluding arithmetic structure in the eigenvectors. We also prove a law of large numbers for the number of gaps at all scales between $N^{-3/2}$ and N^{-1} , and we sketch the proof of the analogous smallest-gap result for complex Hermitian Wigner matrices.

Contents

1	Introduction	1
2	Overview of the proof	9
3	Preliminaries	22
4	Reduction to three propositions	42
5	A comparison estimate for pair statistics	60
6	Pair statistics at the scales $N^{-1-\alpha}$	80
7	Actual and linearized gaps	101
8	A local limit theorem for the linearized increments	107
9	Complex Hermitian matrices	130
A	Pair statistics of the Gaussian orthogonal ensemble	155
	References	176

Introduction

1.1 Background

The statistics of spacings between consecutive energy levels have been studied since Wigner proposed to model the spectra of heavy nuclei by the eigenvalues of large random matrices [78, 80, 81]. A basic feature of such spectra is the rarity of very close neighboring levels, which is known as *level repulsion*. The origin of level repulsion goes back to an observation of von Neumann and Wigner [76]. Real symmetric matrices with a repeated eigenvalue form a set of codimension two, and complex Hermitian matrices with a repeated eigenvalue form a set of codimension three. Heuristically, the probability that a given spacing of a random matrix is smaller than s therefore behaves like $s^{\beta+1}$, where $\beta = 1$ in the real symmetric case and $\beta = 2$ in the complex Hermitian case, and the density of spacings vanishes like s^β near zero [36, 37]. The same exponent appears in Wigner's surmise for the spacing distribution [80] and in the factor $\prod_{i < j} |\lambda_i - \lambda_j|^\beta$ in the eigenvalue densities of Dyson's ensembles [20, 21, 56].

The real symmetric case corresponds to systems with time-reversal symmetry. For such systems, Bohigas, Giannoni, and Schmit conjectured that quantum spectra whose classical dynamics is chaotic share the local statistics of the Gaussian orthogonal ensemble (GOE) [8], in contrast with the Poisson statistics, without repulsion, expected for integrable systems [6]. This picture is supported by experimental data (see, for example, [38, 66] for nuclear spectra and microwave billiards).

This paper concerns the smallest spacings in a spectrum of N levels. If the spacings were independent, with a density proportional to s^β near zero, then the smallest spacing would be of order $N^{-1/(\beta+1)}$ times the mean spacing, and the smallest spacings, suitably rescaled, would form a Poisson process with an intensity proportional to $s^\beta ds$. This heuristic is described in [35, 63], and numerical simulations agree with it [49, 63]. (We emphasize that this Poisson process describes the rare small spacings. Its intensity vanishes at zero because of level repulsion, and it should not be confused with the Poisson level statistics of integrable systems.) Smallest gaps also appear in numerical linear algebra. For example, the rate at which the Toda flow, which is closely related to the QR algorithm, diagonalizes a symmetric matrix is governed by its smallest eigenvalue gap [3, 17, 18], and numerical experiments suggest that the resulting running times are universal for Wigner matrices, including matrices with discrete entries [59].

We call real symmetric or complex Hermitian random matrices with independent entries on and above the diagonal *Wigner matrices*. The class considered in this paper is specified in Definition 1.1 below. In mathematics, the local eigenvalue statistics of Wigner matrices are now known to be universal, in the sense that under mild conditions on the entry distributions they agree asymptotically with those of the Gaussian ensembles. This was established through the work of many authors. We mention [25–28, 41, 70] for the averaged bulk statistics, [22, 69] for the distribution of a single gap, and [12, 46, 47] for statistics at a fixed energy, and we refer to the book [23] for an account of the dynamical approach. Matrices with discrete entries are natural in this context. Wigner’s first paper on the semicircle law concerns real symmetric matrices whose off-diagonal entries are ± 1 with random signs [79], and the centered and rescaled adjacency matrix of a dense Erdős–Rényi graph has independent Bernoulli entries above the diagonal.

Universality of a single gap does not determine the smallest gap among N gaps, since the latter depends on rare events. For the classical ensembles, the smallest gaps have been computed from explicit formulas for the correlation functions. When these functions are determinants, as for the circular unitary ensemble (CUE) and the Gaussian unitary ensemble (GUE), Vinson found the limiting law of the smallest gap [75]. Building on Soshnikov’s analysis of extreme spacings in determinantal point processes [65], Ben Arous and Bourgade [3] obtained the joint limit of the smallest gaps of the CUE and the GUE together with their locations, and Figalli and Guionnet carried these results over to several-matrix models [34]. Determinantal structure is also behind the results of Shi and Jiang for complex Ginibre matrices [62] and those of Charlier for the two-dimensional Coulomb gas at $\beta = 2$ with a general potential, which describes the eigenvalues of random normal matrices [14].

For the Gaussian orthogonal ensemble (GOE), the correlation functions are Pfaffians, and the smallest gaps were determined by Feng, Tian, and Wei [32]. Feng, Li, and Yao [33] gave a detailed treatment of the Gaussian symplectic ensemble, together with a sketch of a GOE version that also records the locations of the gaps, and the smallest gaps of real Gaussian matrices were studied in [52]. Feng and Wei [30] treated the circular β -ensembles for every positive integer β . For the largest gaps of the CUE and the GUE, Ben Arous and Bourgade [3] determined the leading order, and Feng and Wei [31] determined the fluctuations.

For Wigner matrices, Bourgade [10] proved universality of the smallest gaps for both real symmetric and complex Hermitian generalized Wigner matrices under smoothness assumptions on the entry distributions. Bourgade also proved universality of the largest gaps in the complex Hermitian case under a smoothness assumption. Landon, Lopatto, and Marcinek [48] proved a comparison theorem that removes this smoothness assumption. In both works, the explicit largest-gap limit laws were stated for complex Hermitian matrices because the corresponding limiting distribution for the GOE was not known at

the time; their comparison arguments also apply to real symmetric matrices. Whether the smallest gaps are universal without smoothness assumptions was raised in [10]. Recently, Zhang [82] proved universality of the smallest gaps for complex Hermitian Wigner matrices whose entries have real and imaginary parts with sufficiently many well-separated atoms (at least ten suffice), through a comparison argument that matches many moments of the entries. In [82, Conjecture 1.6], he conjectured the analogous result for real symmetric matrices, noting that it does not seem accessible by his methods, and he further conjectured that two atoms suffice.

For real symmetric matrices with discrete entries, the strongest results available were lower bounds. Nguyen, Tao, and Vu [57] proved tail bounds for gaps that imply, in the normalization of this paper, that the smallest gap is at least $N^{-2-o(1)}$ with probability tending to one, and related bounds were obtained in [13, 51, 64]. The absence of repeated eigenvalues was proved in [13, 53, 71]. These lower bounds are smaller than the true order $N^{-3/2}$ of the smallest gaps by a factor $N^{1/2}$.

1.2 Main results

We now state our results. We begin with the definition of the matrices we consider.

Definition 1.1. Let X and Z be real random variables such that $\mathbb{E}X = \mathbb{E}Z = 0$, $\mathbb{E}X^2 = 1$, and $\mathbb{E}Z^2 > 0$, and such that there exist constants $c_0, C_0 > 0$ with

$$\mathbb{P}(|X| > y) + \mathbb{P}(|Z| > y) \leq C_0 e^{-c_0 y^2} \quad \text{for all } y \geq 0. \quad (1.1)$$

For every integer $N \geq 2$, let $(X_{ab})_{1 \leq a < b \leq N}$ be independent copies of X , and let $(Z_a)_{1 \leq a \leq N}$ be independent copies of Z , independent of the X_{ab} . We call the $N \times N$ real symmetric matrix $H = H_N$ with entries

$$H_{ab} = H_{ba} = N^{-1/2} X_{ab} \quad (1 \leq a < b \leq N), \quad H_{aa} = N^{-1/2} Z_a \quad (1 \leq a \leq N), \quad (1.2)$$

the *real symmetric Wigner matrix* associated with X and Z .

No assumption on the symmetry, smoothness, or support of the laws of X and Z is made beyond (1.1). In particular, X may take only two values, and $\mathbb{E}X^3$ may be nonzero; for example, X may be a centered and normalized Bernoulli variable with any parameter. The law of Z may equal that of X , in which case all entries on and above the diagonal are identically distributed.

For a matrix H as in Definition 1.1, let $\lambda_1 \leq \dots \leq \lambda_N$ denote its eigenvalues. For every positive integer n , we set $[n] = \{1, \dots, n\}$. For $i \in [N-1]$, we let $g_i = \lambda_{i+1} - \lambda_i$ denote the i -th gap. We let

$$\rho_{\text{sc}}(E) = \frac{1}{2\pi} \sqrt{\max\{4 - E^2, 0\}}, \quad c(E) = \frac{\pi^2}{6} \rho_{\text{sc}}(E)^3 \quad (E \in \mathbb{R}) \quad (1.3)$$

denote the semicircle density and the function that determines the intensity in our main result. We identify a locally finite point configuration on $[0, \infty) \times (-2, 2)$ with its counting measure. We equip the set of such measures with the *vague topology*, which we define as the coarsest topology for which the maps $\mu \mapsto \int f d\mu$ are continuous for all continuous functions f with compact support in $[0, \infty) \times (-2, 2)$. We use the convention that the minimum over the empty set is $+\infty$. Our main result is the following theorem.

Theorem 1.2. *Let H be as in Definition 1.1. Then the point processes*

$$\Xi_N = \sum_{i \in [N-1]: \lambda_i \in (-2, 2)} \delta_{(N^{3/2}g_i, \lambda_i)} \quad (1.4)$$

on $[0, \infty) \times (-2, 2)$ converge in distribution, with respect to the vague topology, to a Poisson point process with intensity

$$\nu(dx dE) = c(E) x dx dE. \quad (1.5)$$

Consequently, for every nonempty interval I whose closure is contained in $(-2, 2)$ and every $x \geq 0$,

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(N^{3/2} \min_{i \in [N-1]: \lambda_i \in I} g_i > x \right) = \exp \left(- \frac{\pi^2 x^2}{12} \int_I \rho_{\text{sc}}(E)^3 dE \right). \quad (1.6)$$

Theorem 1.2 identifies the smallest gaps in the bulk of the spectrum for every matrix as in Definition 1.1, including Bernoulli matrices. It settles [82, Conjecture 1.6], together with the stronger version with two atoms, for the matrices of Definition 1.1. The conjecture concerns real generalized Wigner matrices whose entries have tails at most $c^{-1}e^{-x^c}$ and laws supported on sufficiently many well-separated atoms (two atoms in the stronger version), whereas Theorem 1.2 requires sub-Gaussian entries, $\mathbb{E}Z^2 > 0$, and identically distributed entries above the diagonal. The limit agrees with the one obtained for the GOE in [32, 33] and with the one obtained under smoothness assumptions in [10]. For every interval I of positive length whose closure is contained in $(-2, 2)$, the Poisson limit determines the joint limiting law of the k smallest gaps in I for every k . We assign $+\infty$ to the k -th smallest gap when fewer than k gaps have left endpoints in I . For example, as in [10, Corollary 1.3], which gives the limiting law of the k -th smallest gap, the k -th smallest value of $N^{3/2}(2^{-1} \int_I c(E) dE)^{1/2} g_i$ over i with $\lambda_i \in I$ converges in law to the distribution with density $2x^{2k-1}e^{-x^2}/(k-1)!$ on $(0, \infty)$.

We have not attempted to optimize the assumptions. The sub-Gaussian condition (1.1) is used in several places, for instance through the results of [57]. The condition $\mathbb{E}Z^2 > 0$ is used through results on generalized Wigner matrices [10, 29], which assume that all entry variances are bounded below by a multiple of N^{-1} ; in particular, Theorem 1.2 does not apply directly to centered adjacency matrices of random graphs, whose diagonal vanishes. We assume that the entries above the diagonal are identically distributed because the resampling argument of this paper and the results of [57] use this property.

Our second result is a law of large numbers for the number of gaps at scales between the smallest gaps and the typical spacing. It is the main input to the proof of Theorem 1.2 and may be of independent interest. We let $E_i = (\lambda_i + \lambda_{i+1})/2$ denote the midpoint of the i -th gap. We extend every function on $(-2, 2)$ by zero to a function on $\mathbb{R}$, and every function on $[0, \infty) \times (-2, 2)$ by zero to a function on $[0, \infty) \times \mathbb{R}$.

Theorem 1.3. *Let H be as in Definition 1.1, and let $\alpha \in (0, 1/2)$. Then for every continuous function $\varphi: (-2, 2) \rightarrow \mathbb{R}$ with compact support and every continuous function $f: [0, \infty) \rightarrow \mathbb{R}$ with compact support,*

$$N^{2\alpha-1} \sum_{i=1}^{N-1} \varphi(E_i) f(N^{1+\alpha} g_i) \longrightarrow \int_{-2}^2 \varphi(E) c(E) dE \cdot \int_0^\infty y f(y) dy \quad (1.7)$$

in probability as $N \rightarrow \infty$.

We call a compact interval contained in $(-2, 2)$ a *bulk interval*. In words, Theorem 1.3 states that for every bulk interval I and every $y > 0$, the number of gaps at most $yN^{-1-\alpha}$ with midpoints in I is asymptotic to $N^{1-2\alpha}(y^2/2) \int_I c(E) dE$ in probability. By Theorem 1.2, this number tends to zero in probability for $\alpha > 1/2$, and it converges in law to a Poisson random variable for $\alpha = 1/2$, so a law of large numbers cannot hold for $\alpha \geq 1/2$. For the GOE, the corresponding law of large numbers for pair statistics with $\alpha < 1/4$ (Proposition 3.6) follows from the Pfaffian formulas for its correlation functions. We prove it in Appendix A, where it serves as the main input to the base case of our proof.

Our methods also apply to complex Hermitian matrices, as follows.

Definition 1.4. Let Z be a real random variable such that $\mathbb{E}Z = 0$ and $\mathbb{E}Z^2 > 0$, and let X_C be a complex random variable such that $\mathbb{E}X_C = 0$, $\mathbb{E}|X_C|^2 = 1$, and $\mathbb{E}X_C^2 = 0$. Suppose that there exist constants $c_0, C_0 > 0$ such that $\mathbb{P}(|X_C| > y) + \mathbb{P}(|Z| > y) \leq C_0 e^{-c_0 y^2}$ for all $y \geq 0$. For every integer $N \geq 2$, let $(X_{C,ab})_{1 \leq a < b \leq N}$ be independent copies of X_C , and let $(Z_a)_{1 \leq a \leq N}$ be independent copies of Z , independent of the $X_{C,ab}$. We call the $N \times N$ complex Hermitian matrix H with entries $H_{ab} = N^{-1/2} X_{C,ab}$ and $H_{ba} = \overline{H_{ab}}$ for $a < b$, and $H_{aa} = N^{-1/2} Z_a$, the *complex Hermitian Wigner matrix* associated with X_C and Z .

The condition $\mathbb{E}X_C^2 = 0$ states that the real and imaginary parts of X_C are uncorrelated and have the same variance. The third moments of X_C are arbitrary. For example, the real and imaginary parts of X_C may be independent copies of $2^{-1/2}X$ for every X as in Definition 1.1.

Theorem 1.5. *Let H be as in Definition 1.4, with eigenvalues $\lambda_1 \leq \dots \leq \lambda_N$ and gaps $g_i = \lambda_{i+1} - \lambda_i$, and let $c_2(E) = (\pi^2/3)\rho_{sc}(E)^4$. Then the point processes*

$$\sum_{i \in [N-1]: \lambda_i \in (-2, 2)} \delta_{(N^{4/3} g_i, \lambda_i)} \quad (1.8)$$

on $[0, \infty) \times (-2, 2)$ converge in distribution, with respect to the vague topology, to a Poisson point process with intensity $c_2(E) x^2 dx dE$. Consequently, for every nonempty interval I whose closure is contained in $(-2, 2)$ and every $x \geq 0$,

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(N^{4/3} \min_{i \in [N-1]: \lambda_i \in I} g_i > x \right) = \exp \left(- \frac{\pi^2 x^3}{9} \int_I \rho_{\text{sc}}(E)^4 dE \right). \quad (1.9)$$

The limit agrees with the one for the GUE [3]. Theorem 1.5 removes the smoothness assumption of [10] and the assumption on the number of atoms of [82] for complex Hermitian Wigner matrices with identically distributed entries above the diagonal. We give the proof of Theorem 1.5 in less detail than that of Theorem 1.2. In Section 9, we sketch how the arguments of Sections 5 and 6 and Appendix A adapt to complex Hermitian matrices; the resampling argument is not needed in this case. We give the details of the two estimates that are new. The first bounds, for $\kappa \in (0, 1/8)$, the number of triples of eigenvalues in a bulk interval within distance $N^{-1-\kappa}$ of each other by $N^{1-8\kappa+o(1)}$, with probability tending to one. This is small compared with the number $N^{1-3\alpha'}$ of gaps at a smaller scale $N^{-1-\alpha'}$ when $\alpha' < 8\kappa/3$, which is the comparison that the induction needs. The second bounds the probability that a Gaussian perturbation of a 3×3 diagonal matrix has a small gap.

1.3 Methods and relation to previous work

We give an overview of the proofs in Section 2. Here we describe how their ingredients relate to the literature.

In [10], the smoothness of the entry distributions is used through the reverse heat flow [25], which represents the matrix, up to a small error in total variation, as a matrix with a Gaussian component. Dyson Brownian motion [19] must run for a time at least $N^{-1/2+2\varepsilon}$, for some $\varepsilon > 0$, for its relaxation to resolve the gaps at the scale $N^{-3/2}$. This time corresponds to a Gaussian component of standard deviation $N^{-1/4+\varepsilon}$ in the variables X_{ab} , and this is where smoothness on the scale $N^{-1/4+\varepsilon}$ is needed.

The alternative is a comparison argument, which replaces the entries one at a time. As explained in Section 2.1, for real symmetric matrices a single entry moves a gap by an amount comparable to the smallest gaps, so the standard perturbative comparison estimate does not give a small error at this scale. Zhang gives a different reason for restricting his result to the complex Hermitian case, which is consistent with ours. The error in his Green function comparison, which matches a fixed number of moments, is small only for spectral parameters at distance much larger than $N^{-3/2}$ from the real axis (see [82, (11)] and the paragraph after it), while the smallest gaps of real symmetric matrices are of order $N^{-3/2}$. We avoid comparison at the scale $N^{-3/2}$.

Instead, we use the fact that resampling a small random set of entries does not change the law of the matrix. Resampling a few random entries of a Wigner matrix, and the

heuristic that each resampled entry moves an eigenvalue by one step of a random walk, were used by Bordenave, Lugosi, and Zhivotovskiy to study the noise sensitivity of the top eigenvector [9]; see also [16] for a related dynamics. We use resampling instead to smooth the law of the gaps while preserving the law of the matrix. The relevant random walk acts on the traceless part of the 2×2 compression of the matrix to the span of two consecutive eigenvectors. This traceless part can be identified with a vector in $\mathbb{R}^2$ whose norm is half of the gap, which reflects the codimension-two phenomenon of [76].

The analysis of this random walk requires a local limit theorem that resolves the law of the walk at the scale of a single step, which for discrete entries requires excluding arithmetic structure in the eigenvectors. For this, we use net and decoupling arguments developed for the invertibility of random matrices and the structure of their eigenvectors by Rudelson and Vershynin [60], Vershynin [74], Nguyen, Tao, and Vu [57], and Tao and Vu [71]. The local limit theorem itself is proved by Fourier analysis with smooth test functions, as in the classical theorems of Shepp and Stone [61, 67], and the passage from local limits to Poisson convergence uses the Bonferroni inequalities, as in the factorial moment method used for extreme gaps in [3, 65]. We also use the tail bound of Nguyen, Tao, and Vu for three consecutive eigenvalues [57, Corollary 2.5] at several points.

The resampling argument requires the number of gaps at a scale slightly larger than $N^{-3/2}$, which is the content of Theorem 1.3. Its proof uses a comparison argument along the Ornstein–Uhlenbeck flow. Our comparison estimate follows the continuity estimate of Bourgade and Yau [11, Lemma A.1] (see also [23, Chapter 15]), which descends from the Lindeberg method of Chatterjee [15], the four moment theorem of Tao and Vu [70], and the Green function comparison of Erdős, Yau, and Yin [28]; the expansion in a single entry is the cumulant expansion of Khorunzhy, Khoruzhenko, and Pastur [44], in the form of [54, Proposition 3.1].

Comparison arguments for statistics that involve all eigenvalues appear in [48], which uses regularizations of eigenvalues through the Helffer–Sjöstrand formula following Knowles and Yin [45], and in [82], which uses resolvent expansions with spectral parameters at distance less than N^{-1} from the real axis, together with many matched moments. In these works, derivatives of the observables are controlled through resolvent or perturbation formulas. Our observables are instead bounded by one in a complex neighborhood of the real entries, which follows from a trace representation of the pair statistic, the argument principle, and the local semicircle law [4, 24, 29], and we bound their derivatives by Cauchy’s estimates. We are not aware of an earlier use of this device in comparison arguments.

The proof of Theorem 1.3 is an induction on α . Its base case $\alpha < 1/8$ uses the coupling of Bourgade [10, Corollary 3.3] between Dyson Brownian motion started from H and from a GOE matrix, together with the GOE result of Appendix A, which relies on the Pfaffian formulas for the GOE correlation functions [1, 56, 72] in the form given in [33].

To go beyond $\alpha = 1/8$, we analyze matrices with a small Gaussian component conditionally on the eigenvectors of H , as in the work of Johansson [41] and Erdős, Péché, Ramírez, Schlein, and Yau [25] (see also [26] and [23, Chapter 12]); the regime of Dyson Brownian motion run for times much smaller than N^{-1} , which is the one relevant here, was studied by von Soosten and Warzel [77]. Our perturbation lemma for entrywise small perturbations of diagonal matrices is a variant of the quadratic residual bounds of Mathias [55] and Li and Li [50], with entrywise bounds in place of operator norms. The computation of gap counts at a smaller scale through independent 2×2 Gaussian blocks makes rigorous, in this regime, the 2×2 heuristic behind Wigner's surmise. The induction on the scale is in the spirit of the multiscale arguments for local laws (see [4, Section 5]) and of the induction on the size of index sets in [5].

1.4 Notation

Constants $c, C > 0$ may depend on the laws of X and Z in Definition 1.1 (or of X_C and Z in Definition 1.4) and on fixed parameters such as bulk intervals, test functions, integers, and small exponents, but never on N ; their values may change from line to line. The function $c(E)$ of (1.3) always appears with its argument, which distinguishes it from these constants. The letter ε denotes a small positive exponent. Its value is fixed within each argument and may change between arguments. When the dependence on a parameter matters, we indicate it by a subscript, as in C_k .

We say that an event holds *with overwhelming probability* if for every $D > 0$ there exists $C_D > 0$ such that its complement has probability at most $C_D N^{-D}$ for all N . We let $o(1)$ denote a deterministic quantity that tends to zero as $N \rightarrow \infty$, $o_{\mathbb{P}}(1)$ a random quantity that tends to zero in probability, and $O_{\mathbb{P}}(1)$ a family of random quantities that is tight.

For $x, y \in \mathbb{R}^n$, we let $|x|$ denote the Euclidean norm of x and $x \cdot y$ the Euclidean inner product. For $v \in \mathbb{R}^N$, we let $\|v\|_2$ and $\|v\|_\infty$ denote its Euclidean and supremum norms. For a matrix A , we let $\|A\|$ denote its operator norm, and A^T and A^* its transpose and conjugate transpose. For an $N \times N$ matrix A and a set $\mathcal{K} \subset [N]$, we let $A_{\mathcal{K}\mathcal{K}} = (A_{jk})_{j,k \in \mathcal{K}}$ denote the corresponding principal submatrix. We let Id_n denote the $n \times n$ identity matrix and e_a the a -th standard basis vector of $\mathbb{R}^N$.

For a real symmetric or complex Hermitian $N \times N$ matrix A , we let $\lambda_1(A) \leq \dots \leq \lambda_N(A)$ denote its eigenvalues, counted with multiplicity. For the matrix H of Definition 1.1, we write $\lambda_i = \lambda_i(H)$, and we fix a measurable choice of orthonormal eigenvectors $u_1, \dots, u_N$ with $Hu_i = \lambda_i u_i$; we let $u_i(a) = e_a^T u_i$ denote the a -th coordinate of u_i . Probabilities and expectations conditional on H are conditional on H and on this choice of eigenvectors.

Overview of the proof

This section explains the proofs of Theorem 1.2 and Theorem 1.3, and, in Section 2.6, the changes needed for Theorem 1.5. It is written to be read on its own. The precise statements and proofs occupy Sections 3 to 8 and Appendix A; a roadmap is given in Section 2.7.

Throughout this section, except in Section 2.6, H denotes a matrix as in Definition 1.1. It is real symmetric, its entries above the diagonal are $N^{-1/2}X_{ab}$, where the X_{ab} are independent copies of a centered, sub-Gaussian random variable X of variance one, and its diagonal entries are independent, centered, and sub-Gaussian, with positive variance. Let $\lambda_1 \leq \dots \leq \lambda_N$ denote the eigenvalues of H , and let $u_1, \dots, u_N$ denote corresponding orthonormal eigenvectors. For $i \in [N-1]$, let $g_i = \lambda_{i+1} - \lambda_i$ denote the i -th gap and $E_i = (\lambda_i + \lambda_{i+1})/2$ its midpoint. Recall that $c(E) = (\pi^2/6)\rho_{\text{sc}}(E)^3$, where ρ_{sc} is the semicircle density. Recall also that we call a compact interval contained in $(-2, 2)$ a bulk interval, and that an event holds with overwhelming probability if for every $D > 0$ its complement has probability at most $C_D N^{-D}$. In this section, statements involving the exponent $\varepsilon > 0$ hold for all sufficiently small ε . Approximate statements about random quantities hold with probability tending to one, and in such statements $o(1)$ denotes a random error that is bounded by a deterministic sequence tending to zero, with probability tending to one.

2.1 The main difficulty

We set $d = N^{-3/2}$, and we call d the *critical scale*. Theorem 1.2 states that the smallest gaps of H in the bulk of the spectrum are of order d , and that for every bulk interval I and every $x > 0$, the number of indices i with $g_i \leq xd$ and $\lambda_i \in I$ converges to a Poisson random variable with mean

$$\int_I \int_0^x c(E)y \, dy \, dE = \frac{x^2}{2} \int_I c(E) \, dE. \quad (2.1)$$

The usual route to universality of local eigenvalue statistics compares H with a matrix that has a Gaussian component, by replacing the entries one at a time and showing that each replacement changes the observable of interest by a small amount [15, 28, 70]. For the smallest gaps of real symmetric matrices, the standard perturbative comparison

estimate does not give a small error at this scale. First-order perturbation theory suggests that changing the entries H_{ab} and H_{ba} by an amount h moves λ_i by approximately $2hu_i(a)u_i(b)$. The entries are of order $N^{-1/2}$, and the eigenvectors are delocalized, in the sense that $\max_{i,a} |u_i(a)| \leq N^{-1/2+\varepsilon}$ with overwhelming probability for every $\varepsilon > 0$ [4, 29]. So a single entry moves an eigenvalue, and therefore a gap, by about $N^{-3/2} = d$, which is the scale of the smallest gaps. Observables that resolve gaps at the scale d are therefore not stable under the replacement of a single entry.

In [10], this obstacle was avoided by assuming that the entry distributions are smooth, which allows one to replace H by a matrix with a Gaussian component without a comparison at the scale d . For complex Hermitian matrices, the smallest gaps are of order $N^{-4/3}$, which is much larger than $N^{-3/2}$, and the obstacle does not arise; we return to this case in Section 2.6.

Our argument has two parts. At the scale d , we do not compare H with another ensemble. Instead, we use the sensitivity of the gaps to single entries through a resampling argument, which we describe in Sections 2.2 to 2.4. This argument requires the number of gaps at a scale slightly larger than d . At the scales $N^{-1-\alpha}$ with $\alpha \in (0, 1/2)$, which lie between d and the typical spacing N^{-1} , the typical gap displacement caused by changing a single entry, of order $N^{-3/2}$, is small compared to this scale. In Section 2.5, we prove a law of large numbers for the number of gaps at such scales, using a comparison argument together with an induction on α .

Four steps of the proof are new: the resampling argument, together with the reduction of its local limit theorem to the exclusion of eigenvectors whose coordinates take few values, which replaces the use of smooth entry distributions in [10]; the holomorphic bound on derivatives in the comparison argument, which replaces bounds obtained from eigenvalue perturbation formulas or from resolvent regularizations of eigenvalues; the induction on α , which computes gap counts through independent 2×2 Gaussian blocks and replaces the direct combination of coupling and comparison, which reaches only $\alpha < 1/8$ when $\mathbb{E}X^3 \neq 0$; and, for complex Hermitian matrices, a bound on triples of nearby eigenvalues relative to the number of gaps, together with an estimate for Gaussian perturbations of 3×3 diagonal matrices, which replaces the exclusion of such triples.

2.2 Resampling

Let $\tau = N^{1/100}$ and $m = 2\lfloor \tau/8 \rfloor$. Then $m = \tau/4 + O(1)$. We choose positions $(a_1, b_1), \dots, (a_m, b_m)$ independently and uniformly from the set of pairs (a, b) with $1 \leq a < b \leq N$, and we choose independent copies $Y_1, \dots, Y_m$ of X , all independent of H . We call $(a_1, b_1), \dots, (a_m, b_m)$ the *selected positions*. Let $\tilde{H}$ denote the matrix obtained from H by replacing, for $l = 1, \dots, m$ in turn, the entries in the positions (a_l, b_l) and (b_l, a_l) by $N^{-1/2}Y_l$. Each replacement preserves the joint law of the entries, so $\tilde{H}$ has the

same law as H . It therefore suffices to prove Theorem 1.2 for $\tilde{H}$, and we may perform all computations conditionally on H .

Heuristically, one replacement moves a gap by an amount of order d , so the m replacements should move it by a random walk with $m \approx \tau/4$ steps of size of order d . To make this precise, we use the following elementary fact. The eigenvalues of a real symmetric 2×2 matrix with diagonal entries h_1, h_2 and off-diagonal entry h_{12} differ by

$$2 \left| \left(\frac{h_1 - h_2}{2}, h_{12} \right) \right|. \quad (2.2)$$

Fix $i \in [N - 1]$, and let $\tilde{g}_i$ denote the gap between the i -th and $(i + 1)$ -th eigenvalues of $\tilde{H}$. For $a < b$, we define the vector $D_{ab}^{(i)} \in \mathbb{R}^2$ by

$$D_{ab}^{(i)} = N(u_i(a)u_i(b) - u_{i+1}(a)u_{i+1}(b), u_i(a)u_{i+1}(b) + u_i(b)u_{i+1}(a)), \quad (2.3)$$

and we set

$$z_i = \left(-\frac{\tilde{g}_i}{2d}, 0 \right), \quad S_i = \sum_{l=1}^m (Y_l - X_{a_l b_l}) D_{a_l b_l}^{(i)}. \quad (2.4)$$

We apply (2.2) to the compression of $\tilde{H}$ to the span of u_i and u_{i+1} , which has diagonal entries $u_i^T \tilde{H} u_i$ and $u_{i+1}^T \tilde{H} u_{i+1}$ and off-diagonal entry $u_i^T \tilde{H} u_{i+1}$. When the selected positions are distinct, which holds with probability $1 - O(\tau^2 N^{-2})$, this compression has eigenvalue gap exactly $2d|z_i + S_i|$. If we neglect the interaction of u_i and u_{i+1} with the other eigenvectors, which we justify in Section 2.3, we therefore obtain the approximation $\tilde{g}_i \approx 2d|z_i + S_i|$. We call S_i the *linearized increment* and $2d|z_i + S_i|$ the *linearized gap* of the pair $(i, i + 1)$.

Conditionally on H , the vector S_i is a sum of m independent copies of the random vector $(Y - X_{ab}) D_{ab}^{(i)}$, where (a, b) is a uniformly chosen position with $a < b$ and Y is an independent copy of X . By the delocalization of the eigenvectors, and since all entries X_{ab} are bounded by N^ε with overwhelming probability, we have $|(Y - X_{ab}) D_{ab}^{(i)}| \leq 2N^{2\varepsilon}(|Y| + N^\varepsilon)$. The conditional mean of this vector is small by the eigenvalue equation, and its conditional covariance is $4\text{Id}_2 + o(1)$. Heuristically, the factor $(Y - X_{ab})^2$ has average close to 2 over the uniformly chosen position, and the orthonormality of u_i and u_{i+1} makes the average of $D_{ab}^{(i)} (D_{ab}^{(i)})^T$ over $a < b$ close to 2Id_2 . The entries X_{ab} are not independent of the eigenvectors, so these two averages cannot simply be multiplied. The two are decoupled by the bound $\|\tilde{X}\| \leq N^{1/2+\varepsilon}$, which holds with overwhelming probability, for the matrix $\tilde{X}$ with zero diagonal and off-diagonal entries $X_{ab}^2 - 1$; Lemma 8.4 contains the computation.

So S_i is approximately a centered Gaussian vector with covariance $4m\text{Id}_2 \approx \tau\text{Id}_2$, whose density at $y \in \mathbb{R}^2$ is $(2\pi\tau)^{-1} e^{-|y|^2/(2\tau)}$. Suppose for the moment that S_i satisfies a

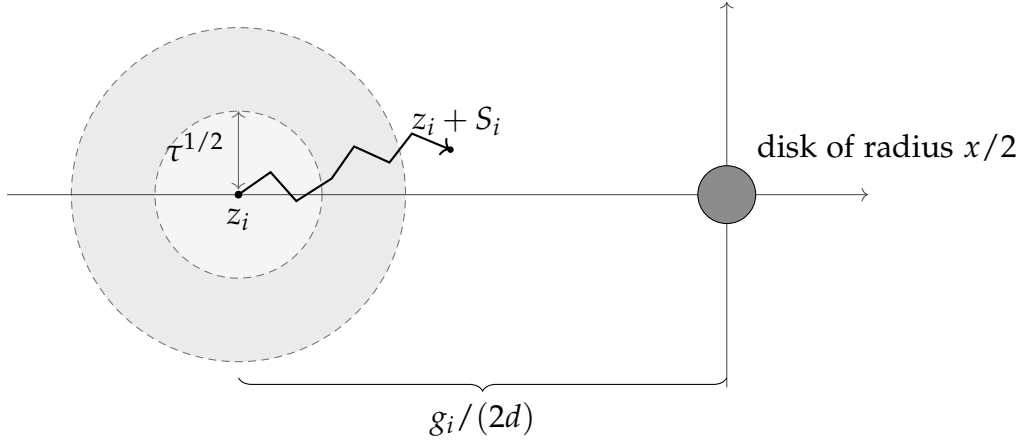


Figure 2.1. The linearized picture of a pair $(i, i+1)$. The point $z_i = (-g_i/(2d), 0)$ encodes the gap g_i of H . The resampling moves it by the linearized increment S_i , a sum of $m \approx \tau/4$ conditionally independent steps of order one, each with covariance close to 4Id_2 , whose law is close to a centered Gaussian law with covariance τId_2 (dashed circles of radius $\tau^{1/2}$ and $2\tau^{1/2}$). Up to an error $o(d)$, the gap of $\tilde{H}$ is at most xd exactly when $z_i + S_i$ lies in the disk of radius $x/2$ around the origin; by (2.6), this has probability close to $(x^2/(8\tau)) \exp(-g_i^2/(8\ell^2))$.

local limit theorem that resolves its law at the scale of a single step, in the sense that for every smooth function $\psi: \mathbb{R}^2 \rightarrow \mathbb{R}$ with compact support,

$$\tau \mathbb{E}[\psi(z_i + S_i) \mid H] = \frac{1}{2\pi} \exp\left(-\frac{|z_i|^2}{2\tau}\right) \int_{\mathbb{R}^2} \psi(y) dy + o(1). \quad (2.5)$$

We set $\ell = d\tau^{1/2} = N^{-299/200}$. Approximating the indicator function of a disk by smooth functions, we find for every $x > 0$ that

$$\mathbb{P}(\tilde{g}_i \leq xd \mid H) \approx \mathbb{P}\left(|z_i + S_i| \leq \frac{x}{2} \mid H\right) \approx \frac{\pi x^2}{4} \cdot \frac{1}{2\pi\tau} \exp\left(-\frac{|z_i|^2}{2\tau}\right) = \frac{x^2}{8\tau} \exp\left(-\frac{g_i^2}{8\ell^2}\right). \quad (2.6)$$

This mechanism is illustrated in Figure 2.1. A gap g_i of order ℓ has a probability of order τ^{-1} to produce a gap of $\tilde{H}$ of order d , and gaps g_i much larger than ℓ contribute negligibly.

To sum (2.6) over i , we need the number of gaps of H at the scale ℓ . For $\alpha = 99/200$, we have $N^{-1-\alpha} = \ell$ and $N^{1-2\alpha} = \tau$, so Theorem 1.3 with this value of α counts the gaps of H at the scale ℓ . Theorem 1.3 concerns continuous test functions with compact support. To apply it to the indicator function of a bulk interval I and to the Gaussian function $y \mapsto e^{-y^2/8}$, we approximate these functions by continuous functions with compact support. We control the tail of the Gaussian function by an upper bound that we prove together with Theorem 1.3 (see Proposition 4.1). It states that for every bulk interval I ,

there exists $C_I > 0$ such that with probability tending to one,

$$\sum_{1 \leq i < j \leq N: (\lambda_i + \lambda_j)/2 \in I} \left(1 + \frac{(\lambda_j - \lambda_i)^2}{\ell^2}\right)^{-61} \leq C_I \tau. \quad (2.7)$$

We obtain

$$\frac{1}{\tau} \sum_{i: E_i \in I} \exp\left(-\frac{g_i^2}{8\ell^2}\right) \longrightarrow 4 \int_I c(E) dE \quad (2.8)$$

in probability, where we used that $\int_0^\infty y e^{-y^2/8} dy = 4$. Multiplying by $x^2/8$, we recover the mean (2.1). We note that the dependence on τ cancels, as it must.

Summing (2.6) over all pairs requires care. The local limit theorem (2.5) has an error $o(\tau^{-1})$ for each pair, while the sums run over all pairs whose gap is small enough for the linearization to apply. These are the pairs with $g_i \leq N^{-7/5}$ (see Section 2.3), and there are about $N^{1/5}$ of them, many more than τ . We therefore restrict the sums to the pairs with $g_i \leq R\ell$ for a large constant R , of which there are $O_R(\tau)$ by (2.7). We treat the remaining pairs with a large-deviation bound. For $g_i \geq \ell$, the conditional probability that $|z_i + S_i| \leq x/2$ is at most $C\tau^{-1}e^{-cg_i/\ell}$ (Proposition 4.4), and these bounds are summable against (2.7), with a sum that tends to zero as $R \rightarrow \infty$.

To obtain a Poisson limit, we need the joint behavior of $S_{i_1}, \dots, S_{i_k}$ for indices $i_1 < \dots < i_k$ such that $i_{j+1} \geq i_j + 2$ for all j , so that the pairs $(i_j, i_j + 1)$ are disjoint. This restriction suffices, since we show in Section 2.3 that, with probability tending to one, the pairs with small gaps are disjoint. The vectors $S_{i_1}, \dots, S_{i_k}$ are built from the same selected positions and the same variables $Y_1, \dots, Y_m$, so they are not conditionally independent. However, their joint conditional covariance is close to $4m\text{Id}_{2k}$, again by the orthonormality of the eigenvectors, and we prove a joint local limit theorem for $(S_{i_1}, \dots, S_{i_k})$. Joint local limits, rather than independence, suffice to compute, through the Bonferroni inequalities, the conditional Laplace functional of the point process of the rescaled linearized gaps $2|z_i + S_i|$, which approximate the rescaled gaps of $\tilde{H}$. Finally, since $\tilde{H}$ and H have the same law, the Poisson limit for $\tilde{H}$ is also the Poisson limit for H .

The choice of τ is a compromise. The local limit theorem needs many steps, so τ must tend to infinity. On the other hand, the error in the linearization grows with τ , as we explain next. The choice $\tau = N^{1/100}$ is convenient, but not essential.

2.3 Comparing actual and linearized gaps

We now explain why the approximation $\tilde{g}_i \approx 2d|z_i + S_i|$ holds up to an error $o(d)$, simultaneously for all relevant i . Let $\Delta = \tilde{H} - H$, let U denote the orthogonal matrix with columns $u_1, \dots, u_N$, and let $W = U^T \Delta U$ denote the perturbation in the eigenbasis

of H . With probability tending to one, the matrix Δ has at most $2m$ nonzero entries, each of absolute value at most $2N^{-1/2+\varepsilon}$, so the delocalization of the eigenvectors gives

$$\max_{i,j} |W_{ij}| \leq \sum_{a < b} |\Delta_{ab}| |u_i(a)u_j(b) + u_i(b)u_j(a)| \leq 4mN^{-3/2+3\varepsilon} \leq w_N, \quad (2.9)$$

$$w_N = \tau N^{-3/2+3\varepsilon},$$

where we used that $4m \leq \tau$. This bound is much smaller than the typical spacing N^{-1} , but it is larger than d . So the first-order approximation is not enough, and we must control the interaction of each pair with the rest of the spectrum.

Let $s = N^{-7/5}$. We call a maximal set $\mathcal{K} \subset [N]$ of consecutive integers such that $\lambda_{j+1} - \lambda_j \leq s$ whenever $j, j+1 \in \mathcal{K}$ a *cluster*. By a result of Nguyen, Tao, and Vu [57], the expected number of indices $i \in [N-2]$ with $\lambda_{i+2} - \lambda_i \leq r$ is $O(N(Nr)^3)$, uniformly in $r \geq N^{-2}$. Applying this with $r = 4s$, we see that $\lambda_{i+2} - \lambda_i > 4s$ for all i , and in particular that no cluster has three or more elements, with probability $1 - O(N(Ns)^3) = 1 - O(N^{-1/5})$.

For each cluster $\mathcal{K}$, a Schur complement argument (Lemma 3.8) shows that $H + \Delta$ has exactly $|\mathcal{K}|$ eigenvalues near the eigenvalues λ_j with $j \in \mathcal{K}$, and that their ordered values differ from the ordered eigenvalues of the block $(\Lambda + W)_{\mathcal{K}\mathcal{K}}$, where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$, by at most

$$2|\mathcal{K}| \left(\max_{i,j} |W_{ij}| \right)^2 \sup_w \sum_{j \notin \mathcal{K}} \frac{1}{|\lambda_j - w|} \leq 4w_N^2 \mathcal{L}_N = \frac{\delta_N}{2}, \quad (2.10)$$

$$\mathcal{L}_N = 2s^{-1} \log N, \quad \delta_N = 8w_N^2 \mathcal{L}_N.$$

Here the supremum is over the complex numbers w in the rectangle obtained by enlarging the interval $[\lambda_{\min \mathcal{K}}, \lambda_{\max \mathcal{K}}]$ by $s/4$ in each direction. We show in Section 7 that the sum is bounded by $\mathcal{L}_N$, because every eigenvalue λ_j with $j \notin \mathcal{K}$ is at distance at least $3s/4$ from such w , and at most two eigenvalues lie in every interval of length $4s$, since $\lambda_{j+2} - \lambda_j > 4s$ for all j . Inserting the values of τ , s , and d , we find

$$\frac{\delta_N}{d} = 16N^{-2/25+6\varepsilon} \log N, \quad (2.11)$$

which tends to zero when $\varepsilon < 1/75$. In Section 7, we take $\varepsilon = 1/100$.

For a two-element cluster $\mathcal{K} = \{i, i+1\}$, the block $(\Lambda + W)_{\mathcal{K}\mathcal{K}}$ is exactly the compression considered in Section 2.2, and by (2.2) its eigenvalue gap equals $2d|z_i + S_i|$ when the selected positions are distinct. Hence $\tilde{g}_i = 2d|z_i + S_i| + o(d)$ for all two-element clusters at once, and all other gaps of $\tilde{H}$ exceed $s/2$. The same estimate, applied with a Gaussian perturbation in place of Δ , is used in Section 2.5.

2.4 A local limit theorem at the scale of a single step

The local limit theorem (2.5) is needed at unit scale, which is the size of a single increment $(Y - X_{ab})D_{ab}^{(i)}$. At this resolution, the lack of smoothness of X matters again. For example, if X takes the values ± 1 , then $Y - X_{ab} \in \{-2, 0, 2\}$. If the linearized increments were supported on a fixed lattice with mesh bounded away from zero at unit resolution, the local limit theorem could fail. The fixed-lattice example at the beginning of Section 8 illustrates this obstruction. The central limit theorem holds regardless, but the local limit theorem requires a quantitative upper bound on the modulus of the conditional characteristic function

$$\Phi_H(\xi) = \mathbb{E}[\exp(i\xi \cdot (Y - X_{ab})D_{ab}^{(i)}) \mid H], \quad \xi \in \mathbb{R}^2, \quad (2.12)$$

for ξ in a range of intermediate frequencies, where (a, b) is a uniformly chosen position and Y is an independent copy of X . Small frequencies are handled by the Gaussian approximation, and large frequencies by the decay of the Fourier transform of ψ in (2.5).

The first step is a symmetrization, which removes the old entries X_{ab} (whose values are fixed once we condition on H) and avoids any symmetry assumption on X . Let Y' denote an independent copy of Y . Averaging over the uniformly chosen position (a, b) and using the convexity of $z \mapsto |z|^2$, we find

$$1 - |\Phi_H(\xi)|^2 \geq \left(\frac{N}{2}\right)^{-1} \sum_{a < b} \mathbb{E}[1 - \cos((Y - Y') \xi \cdot D_{ab}^{(i)})]. \quad (2.13)$$

For $b \in [N]$, let $r_b = N^{1/2}(u_i(b), u_{i+1}(b)) \in \mathbb{R}^2$. For $\xi = (\xi(1), \xi(2)) \in \mathbb{R}^2$, let $\Gamma(\xi)$ denote the symmetric 2×2 matrix with rows $(\xi(1), \xi(2))$ and $(\xi(2), -\xi(1))$. A direct computation from (2.3) gives $\xi \cdot D_{ab}^{(i)} = r_a^T \Gamma(\xi) r_b$. Suppose that the right side of (2.13) is small. Since $\text{Var } X = 1$, the variable $|Y - Y'|$ lies in a fixed compact subset of $(0, \infty)$ with positive probability, and for such values of $Y - Y'$ the phases $(Y - Y') r_a^T \Gamma(\xi) r_b$ are close to $2\pi\mathbb{Z}$ for most pairs (a, b) . We choose two indices a', a'' such that the 2×2 matrix Q with rows $r_{a'}^T$ and $r_{a''}^T$ is well conditioned and such that most phases in these two rows are close to $2\pi\mathbb{Z}$. Then for most b , there exists $\mathbf{n}_b \in \mathbb{Z}^2$ such that

$$(Y - Y') Q \Gamma(\xi) r_b = 2\pi \mathbf{n}_b + (\text{small error}), \quad \text{so that} \quad r_b \approx \frac{2\pi}{Y - Y'} \Gamma(\xi)^{-1} Q^{-1} \mathbf{n}_b. \quad (2.14)$$

When $N^{-c} \leq |\xi| \leq N^c$ for a small constant $c > 0$, the vectors $\mathbf{n}_b$ range over a set of size at most N^{Cc} . Consequently, $N^{1/2}u_i$ is close, in the norm $v \mapsto (N^{-1} \sum_b v(b)^2)^{1/2}$, to a vector whose coordinates take at most N^{Cc} distinct values. For k pairs at once, the same argument applies with $2k$ indices in place of a', a'' .

The second step excludes this. We show that with overwhelming probability, no delocalized unit eigenvector of H is close, in this norm, to a vector whose coordinates take

few distinct values (Lemma 8.5). The proof replaces the eigenvector and its eigenvalue by nearby elements of a deterministic net of approximate eigenvectors and approximate eigenvalues, and it combines a union bound over this net with anticoncentration.

For a given vector v and a given energy E in the net, we split $[N]$ into a set of about $N/2$ indices on which $|v|$ is largest and its complement. For an index a in the complement, the coordinate $N^{1/2}((H - E)v)(a)$ equals the sum of $X_{ab}v(b)$ over the indices b in the first set, plus a term that does not involve the entries $X_{\tilde{a}b}$ with $\tilde{a}$ in the complement and b in the first set. These coordinates are independent over a once the other entries are fixed, and each lies in a given short interval with small probability, by the Berry–Esseen theorem. If $(H - E)v$ is small, then at least $N/4$ of these coordinates lie in such intervals. The product of the corresponding probabilities is much smaller than the inverse of the size of the net, which is exponential in N only through the number of possible assignments of the few distinct values. This part follows the net and decoupling arguments for structured approximate eigenvectors in [57, 60, 74].

2.5 Gap counts at the scales $N^{-1-\alpha}$

It remains to explain Theorem 1.3, which supplies (2.8). For $\alpha \in (0, 1/2)$, let $\ell_\alpha = N^{-1-\alpha}$ and $M_\alpha = N^{1-2\alpha}$. Theorem 1.3 states that for every $y > 0$ and every bulk interval I , the number of gaps $g_i \leq y\ell_\alpha$ with $E_i \in I$ is asymptotic to $M_\alpha(y^2/2) \int_I c(E) dE$. At the scale ℓ_α , a single entry moves a gap by about $N^{-3/2} = o(\ell_\alpha)$, so comparison arguments are available in principle. We use the Ornstein–Uhlenbeck interpolation

$$H_t = e^{-t/2}H + (1 - e^{-t})^{1/2}G, \quad t \in [0, 1], \quad (2.15)$$

where G is a matrix from the Gaussian orthogonal ensemble (GOE), independent of H and normalized so that its off-diagonal entries have variance $1/N$ and its diagonal entries have variance $2/N$.

We work with smoothed pair statistics. Let $p \geq 2$ be an integer. We define $f_p(w) = (1 + w^2)^{-p}$ for $w \in \mathbb{C} \setminus \{i, -i\}$. Let φ be a smooth nonnegative function with compact support in $(-2, 2)$, and let φ_N denote the convolution of φ with the centered Gaussian density of standard deviation $N^{-1/10}$. The function φ_N extends to an entire function, which we also denote by φ_N . Let $\omega_p > 0$ be a small constant depending on p . For an $N \times N$ complex matrix A whose eigenvalues $w_1, \dots, w_N$, listed with algebraic multiplicity, satisfy $|\operatorname{Im} w_j| < \omega_p \ell_\alpha$ for all j , and for $\theta > 0$, we set

$$\mathcal{X}_\alpha(A) = \sum_{1 \leq i < j \leq N} \varphi_N\left(\frac{w_i + w_j}{2}\right) f_p\left(\frac{w_j - w_i}{\ell_\alpha}\right), \quad F(A) = \exp\left(-\frac{\theta \mathcal{X}_\alpha(A)}{M_\alpha}\right). \quad (2.16)$$

The sum does not depend on the order of the eigenvalues, since f_p is even. We call $\mathcal{X}_\alpha(A)$ the *pair statistic* of A . (In Section 5, it is denoted $\mathcal{X}_\alpha(A; \varphi_N, f_p)$.) For a real symmetric

matrix A , it is a weighted count of the pairs of eigenvalues at distance of order ℓ_α . We will show that

$$\frac{\mathcal{X}_\alpha(H)}{M_\alpha} \longrightarrow \int_0^\infty y f_p(y) dy \cdot \int_{-2}^2 \varphi(E) c(E) dE \quad (2.17)$$

in probability, for all sufficiently large p and all φ as above.

For $p \geq 2$, the pairs (i, j) with $j \geq i + 2$ and the pairs at distance much larger than ℓ_α contribute $o(M_\alpha)$ to $\mathcal{X}_\alpha(H)$. Further, linear combinations of the functions f_p approximate every continuous function on $[0, \infty)$ with compact support, and the smoothing of φ can be removed. Hence (2.17) implies Theorem 1.3 at the scale ℓ_α . We carry out this deduction in the proof of Theorem 1.3 in Section 6.4.

Base case. For $\alpha < 1/8$, we use a coupling of Bourgade [10, Corollary 3.3]. Fix $c_* \in (0, 1/2)$. For $N^{-1+\varepsilon} \leq t < 1$, there exists a coupling of H_t with a GOE matrix G' such that, with overwhelming probability, the gaps $\lambda_{i+1}(H_t) - \lambda_i(H_t)$ and $\lambda_{i+1}(G') - \lambda_i(G')$ differ by at most $N^\varepsilon/(N^2 t)$ for all i with $c_* N \leq i \leq (1 - c_*)N$ (Proposition 3.5). We take $t = T$, where $T = N^{-3/4-\alpha}$. Then this error is $N^{-5/4+\alpha+\varepsilon}$, which is $o(\ell_\alpha)$ when $\alpha < 1/8$, so (2.17) with H_T in place of H follows from the corresponding statement for the GOE. We prove the latter in Appendix A from the Pfaffian formulas for the GOE correlation functions. It then remains to compare H_T with H .

Comparison. We compare the Laplace transforms $\mathbb{E}F(H_t)$ and $\mathbb{E}F(H)$. A comparison requires bounds on derivatives of F with respect to the entries of the matrix. Derivatives of eigenvalues of order two and higher contain denominators $(\lambda_i(A) - \lambda_j(A))^{-1}$, so bounds obtained from perturbation formulas would require control of small gaps, which is the information we are trying to obtain.

We avoid this through holomorphy. For $a < b$, let $V^{ab} = e_a e_b^T + e_b e_a^T$. Let A be a real symmetric matrix such that $\|A\| \leq 3$, every unit eigenvector of A has all coordinates bounded by $N^{-1/2+\varepsilon}$, and every interval of length N^{-1} contains at most N^ε eigenvalues of A . By the local semicircle law [4, 29], these conditions hold for H_t with overwhelming probability (Proposition 3.2). We show that there exists $c = c_p > 0$ such that, for all sufficiently large N and every complex ζ with $|\zeta| \leq cN^{-\alpha-4\varepsilon}$, all eigenvalues w of $A + \zeta V^{ab}$ satisfy $|\operatorname{Im} w| < \omega_p \ell_\alpha$ (Lemma 5.2). Hence $F(A + \zeta V^{ab})$ is defined. Since $\mathcal{X}_\alpha$ is a symmetric function of the eigenvalues, $F(A + \zeta V^{ab})$ is a holomorphic function of ζ (Lemma 5.5). Since the imaginary parts of the eigenvalues are much smaller than $N^{-1/5}$, the square of the smoothing scale $N^{-1/10}$ of φ_N , and ω_p is small, every term of the sum in (2.16) has nonnegative real part (Lemmas 5.1 and 5.5). Therefore $|F(A + \zeta V^{ab})| \leq 1$, and Cauchy's estimates give, for all real ζ with $|\zeta| \leq cN^{-\alpha-4\varepsilon}/2$ and all integers $n \geq 1$,

$$\left| \frac{d^n}{d\zeta^n} F(A + \zeta V^{ab}) \right| \leq n! \left(\frac{cN^{-\alpha-4\varepsilon}}{2} \right)^{-n}. \quad (2.18)$$

No information about gaps enters this bound. With (2.18), a Taylor expansion in one entry at a time along the interpolation (2.15), in the style of the continuity estimate of Bourgade and Yau [11], gives for every $D > 0$

$$|\mathbb{E}F(H_t) - \mathbb{E}F(H)| \leq CtN^{1/2+3\alpha+12\varepsilon} + C_D N^{-D}. \quad (2.19)$$

The factor $N^{1/2+3\alpha+12\varepsilon}$ comes from summing third-order Taylor terms, each of size $N^{-3/2} \cdot N^{3(\alpha+4\varepsilon)}$, over the N^2 entries. We do not assume that $\mathbb{E}X^3 = 0$. For $t = T$, the right side of (2.19) is $CN^{-1/4+2\alpha+12\varepsilon} + C_D N^{-D}$, which tends to zero when $\alpha < 1/8$. Since this holds for all $\theta > 0$, the statement (2.17) for H_T implies (2.17) for H . This proves Theorem 1.3 for $\alpha < 1/8$.

Induction on α . Let α and α' satisfy $1/10 \leq \alpha < \alpha' < 1/2$, and suppose that (2.17) holds. To prove (2.17) with α' in place of α , we use the Gaussian component of H_t to transfer information from the scale ℓ_α to the smaller scale $\ell_{\alpha'}$, and then compare back. Let $t_\alpha = N\ell_\alpha^2 = N^{-1-2\alpha}$, let U and Λ be as in Section 2.3, and let $W' = (1 - e^{-t_\alpha})^{1/2} U^T G U$. Then

$$U^T H_{t_\alpha} U = e^{-t_\alpha/2} \Lambda + W', \quad (2.20)$$

so H_{t_α} has the same eigenvalues as $e^{-t_\alpha/2} \Lambda + W'$. (In Section 6, the matrices W' and $e^{-t_\alpha/2} \Lambda$ are denoted W and D_y .) By the orthogonal invariance of G , conditionally on H , the matrix W' is a GOE matrix multiplied by $(1 - e^{-t_\alpha})^{1/2}$, and its entries are of order ℓ_α .

We group the eigenvalues of H into clusters as in Section 2.3, with the threshold $q = N^{-1-\kappa}$ in place of s , where $\kappa = \alpha - \alpha'/6$. We apply the estimate (2.10) with $e^{-t_\alpha/2} \Lambda$, W' , and q in place of Λ , W , and s . Since the entries of W' are bounded by $N^\varepsilon \ell_\alpha$ with overwhelming probability, it shows that for each two-element cluster $\mathcal{K} = \{i, i+1\}$, the gap of H_{t_α} equals the eigenvalue gap of the block $(e^{-t_\alpha/2} \Lambda + W')_{\mathcal{K}\mathcal{K}}$ up to an error of order $N^{C\varepsilon} \ell_\alpha^2 / q$. By (2.2), the gap of this block equals

$$|(e^{-t_\alpha/2} g_i + W'_{i+1,i+1} - W'_{ii}, 2W'_{i,i+1})|. \quad (2.21)$$

Here $W'_{i+1,i+1} - W'_{ii}$ and $2W'_{i,i+1}$ are independent centered Gaussian variables with variance $4(1 - e^{-t_\alpha})/N = 4\ell_\alpha^2(1 + o(1))$, and the blocks of different clusters are conditionally independent. Therefore, conditionally on H , the number of gaps of the blocks at the scale $\ell_{\alpha'}$, which approximates the corresponding number for H_{t_α} , is a sum of independent bounded variables. Its conditional mean is a Gaussian smoothing, at the scale ℓ_α , of the gap counts of H at the scale ℓ_α . These counts are known by the induction hypothesis, so we obtain (2.17) with α' in place of α for H_{t_α} . The comparison (2.19), with α' in place of α and $t = t_\alpha$, transfers it to H provided that $t_\alpha N^{1/2+3\alpha'} = N^{-1/2-2\alpha+3\alpha'}$ tends to zero.

Two further conditions are needed. We have

$$\frac{\ell_\alpha^2}{q} = N^{-1-2\alpha+\kappa}, \quad N(Nq)^3 = N^{1-3\kappa}. \quad (2.22)$$

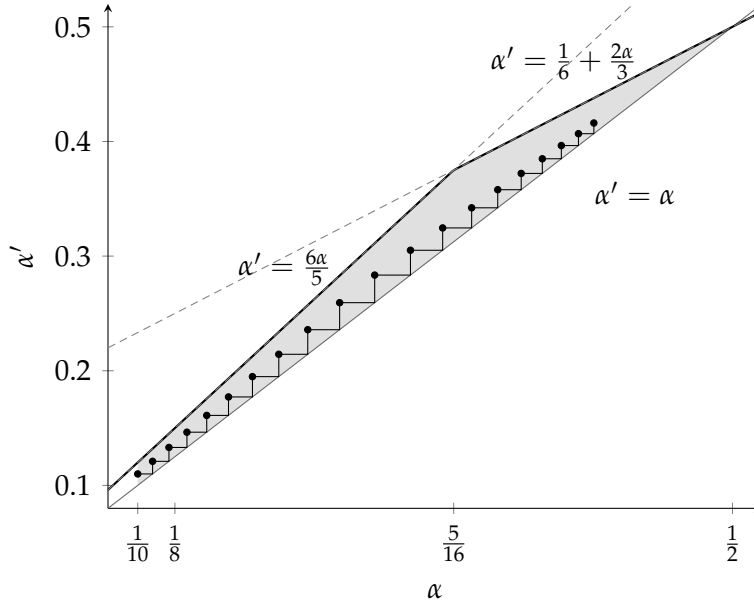


Figure 2.2. The induction on α . The shaded region is the set of pairs (α, α') allowed by (2.23). The staircase shows the first steps of the sequence $\alpha^{(n+1)} = \alpha^{(n)} + \min\{\alpha^{(n)}, 1/2 - \alpha^{(n)}\}/10$ starting from $\alpha^{(0)} = 1/10$, which increases to $1/2$; the base case covers all $\alpha < 1/8$.

The error in the Schur complement estimate, which is of order ℓ_α^2/q up to factors $N^{C\varepsilon}$, must be $o(\ell_{\alpha'})$, which holds when $\kappa < 2\alpha - \alpha'$. Clusters with three or more elements must be negligible. By the bound of Nguyen, Tao, and Vu, their expected number is $O(N(Nq)^3)$, which must be $o(M_{\alpha'})$, and this holds when $3\kappa > 2\alpha'$. For $\kappa = \alpha - \alpha'/6$, both conditions are equivalent to $\alpha' < 6\alpha/5$. All conditions therefore hold when

$$\frac{1}{10} \leq \alpha < \alpha' < \min\left\{\frac{6\alpha}{5}, \frac{1}{6} + \frac{2\alpha}{3}\right\}. \quad (2.23)$$

The bound $p \geq 61$, for which (2.17) is proved, and the lower bound $\alpha \geq 1/10$ are used only in the bound (6.47) in the proof of Proposition 6.7, where $2p(\alpha' - \kappa) > 1 + 2\alpha'$ is needed and holds because $\alpha' - \kappa > \alpha/6 \geq 1/60$. Starting from $\alpha = 1/10$, which is covered by the base case, finitely many steps reach every $\alpha < 1/2$, as shown in Figure 2.2.

2.6 Hermitian matrices

In this section only, H denotes a complex Hermitian Wigner matrix as in Definition 1.4, with eigenvalues $\lambda_1 \leq \dots \leq \lambda_N$ and gaps $g_i = \lambda_{i+1} - \lambda_i$, and the matrix G in (2.15) is taken from the Gaussian unitary ensemble (GUE), normalized so that $\mathbb{E}|G_{ab}|^2 = 1/N$ and $\mathbb{E}G_{ab}^2 = 0$ for $a < b$ and its diagonal entries have variance $1/N$. The smallest gaps are of order $N^{-4/3}$, which is the critical scale in this case, while a single entry moves a gap by only about $N^{-3/2}$. Comparison at the critical scale is therefore possible, and

no resampling is needed. The scale $N^{-4/3}$ corresponds to $\alpha = 1/3$, and the analogue of (2.19) at this scale has the right side

$$CtN^{1/2}(N \cdot N^{-4/3})^{-3}N^{C\varepsilon} = CtN^{3/2+C\varepsilon} \quad (2.24)$$

when the third moments of the entries do not vanish. So the Gaussian component must be taken at a time $t = N^{-1-2\alpha_1}$ with $\alpha_1 > 1/4$. We also take $\alpha_1 < 1/3$, so that the entries of the Gaussian part of H_t , which are of order $N^{-1-\alpha_1}$, are larger than the critical scale.

An induction on α as in Section 2.5, with Gaussian vectors in $\mathbb{R}^3$ in place of $\mathbb{R}^2$ in (2.21), gives the Hermitian analogue of (2.17). With $c_2(E)$ as in Theorem 1.5, it states that $\mathcal{X}_\alpha(H)/N^{1-3\alpha}$ converges in probability to $\int_0^\infty y^2 f_p(y) dy \cdot \int_{-2}^2 \varphi(E) c_2(E) dE$. We prove this along a finite chain of exponents that starts below $1/8$ and ends at $\alpha_1 = 8/25$, which is all that the argument needs; the induction step differs from the real case in its treatment of triples of eigenvalues, described below. At the time $t = N^{-1-2\alpha_1}$, the 2×2 block computation then gives the statistics of the gaps of H_t at the scale $N^{-4/3}$, provided that clusters of three eigenvalues are under control and clusters of four or more eigenvalues are absent.

This is where nonzero third moments cause difficulty. We again group the eigenvalues into clusters, at a threshold $q = N^{-1-\kappa}$ with $\kappa > 0$ to be chosen. For the GUE, the three-point correlation function vanishes quadratically in each of the three distances between its arguments, so the expected number of indices i with $\lambda_{i+2} - \lambda_i \leq q$ and λ_i in a bulk interval is of order $N(Nq)^{2+6} = N^{1-8\kappa}$. Excluding such triples therefore requires $\kappa > 1/8$. However, transferring information at the scale $N^{-1-\kappa}$ from the GUE to H through Bourgade's coupling and the comparison estimate requires $\kappa < 1/8$.

Instead of excluding triples, we show that they rarely produce small gaps. For $\kappa < 1/8$, the number of indices i with $\lambda_{i+2} - \lambda_i \leq N^{-1-\kappa}$ and λ_i in a bulk interval is at most $N^{1-8\kappa+C\varepsilon}$ with probability tending to one. Further, a Gaussian perturbation at time $t = N^{-1-2\alpha_1}$ turns a given triple into a gap of order $N^{-4/3}$ with probability at most $N^{C\varepsilon}(N^{-4/3}/N^{-1-\alpha_1})^3 = N^{-1+3\alpha_1+C\varepsilon}$. The expected number of small gaps produced by triples is therefore bounded by

$$N^{1-8\kappa+C\varepsilon} \cdot N^{-1+3\alpha_1+C\varepsilon} = N^{3\alpha_1-8\kappa+2C\varepsilon}, \quad (2.25)$$

which tends to zero when $\kappa > 3\alpha_1/8$. This is compatible with $\kappa < 1/8$ because $\alpha_1 < 1/3$. Clusters of four or more eigenvalues within distance $N^{-1-\kappa}$ are absent with probability tending to one when $\kappa > 1/15$, by the analogous count of quadruples. A similar bound on triples is used in the steps of the induction. We sketch the details in Section 9, where we take $\alpha_1 = 8/25$.

2.7 Organization of the paper

Section 3 collects estimates on eigenvalues and eigenvectors from the literature, a perturbation lemma for matrices that are entrywise small perturbations of diagonal matrices, and an estimate for Gaussian kernels. In Section 4, we state three propositions (the gap counts at the scale ℓ , the local limit theorem for linearized increments, and the comparison of actual and linearized gaps) and deduce Theorem 1.2 from them. Section 5 contains the holomorphic comparison estimate, which is the main input to Section 6, where we prove Theorem 1.3 and Proposition 4.1 through the base case and the induction on α . In Section 7, we compare actual and linearized gaps, and in Section 8, we prove the local limit theorem. Section 9 sketches the proof of Theorem 1.5. Appendix A computes the gap counts of the GOE needed for the base case.

Preliminaries

In this section, we collect the facts about eigenvalues and eigenvectors that are used in the rest of the paper. Section 3.1 contains regularity and rigidity estimates for H , H_t , and G , a level repulsion estimate for H , a coupling of the gaps of H_t with those of the GOE, and the pair statistics of the GOE. Section 3.2 contains a deterministic lemma on the eigenvalues of a diagonal matrix perturbed by a Hermitian matrix with small entries, which we apply in Section 6, Section 7, and Section 9. Section 3.3 contains an elementary Gaussian estimate used in Section 6 and Section 9.

We first fix some notation for the rest of the paper. We write $\sigma_Z^2 = \mathbb{E}Z^2$, which is positive by Definition 1.1, and we set $X_{aa} = Z_a$ for $a \in [N]$, so that $X_{ab} = N^{1/2}H_{ab}$ for all $a \leq b$. We define the real symmetric $N \times N$ matrices

$$V^{ab} = e_a e_b^T + e_b e_a^T \quad (1 \leq a < b \leq N), \quad V^{aa} = e_a e_a^T \quad (a \in [N]), \quad (3.1)$$

so that every real symmetric $N \times N$ matrix A satisfies $A = \sum_{a \leq b} A_{ab} V^{ab}$. Throughout the paper, G denotes a GOE matrix independent of H , whose upper-triangular entries are independent centered Gaussian random variables with variance N^{-1} off the diagonal and $2N^{-1}$ on the diagonal, and we set

$$H_t = e^{-t/2}H + (1 - e^{-t})^{1/2}G, \quad t \in [0, 1]. \quad (3.2)$$

For $i \in [N]$, we let γ_i denote the classical location defined by $\int_{-2}^{\gamma_i} \rho_{\text{sc}}(E) dE = i/N$. For events indexed by a parameter, we say that they hold with overwhelming probability *uniformly* in the parameter if the constant C_D in the definition of overwhelming probability can be chosen independent of the parameter.

3.1 Estimates for eigenvalues and eigenvectors

We begin with the set of matrices on which our deterministic arguments take place.

Definition 3.1. Let $\varepsilon > 0$. We define $\mathcal{G}_\varepsilon$ to be the set of $N \times N$ real symmetric matrices A with the following three properties.

- (i) $\|A\| \leq 3$.
- (ii) Every unit eigenvector $v \in \mathbb{R}^N$ of A satisfies $\|v\|_\infty \leq N^{-1/2+\varepsilon}$.

(iii) For all $E \in \mathbb{R}$, we have $\#\{i \in [N] : \lambda_i(A) \in [E, E + N^{-1}]\} \leq N^\varepsilon$.

We call the elements of $\mathcal{G}_\varepsilon$ *regular matrices*.

Each of the three conditions becomes weaker as ε increases, so $\mathcal{G}_\varepsilon \subset \mathcal{G}_{\varepsilon'}$ for $\varepsilon < \varepsilon'$. (Using the continuity of eigenvalues and of total eigenprojections [43], one checks that $\mathcal{G}_\varepsilon$ is a Borel set.)

Proposition 3.2. *The following statements hold.*

(a) For all $\varepsilon > 0$ and $D > 0$, there exists $C_{\varepsilon,D} > 0$ such that $\mathbb{P}(H_t \notin \mathcal{G}_\varepsilon) \leq C_{\varepsilon,D} N^{-D}$ for all $N \geq 2$ and $t \in [0, 1]$.

(b) For all $c_* \in (0, 1/2)$, $\varepsilon > 0$, and $D > 0$, there exists $C > 0$ such that for all $N \geq 2$ and $t \in [0, 1]$,

$$\mathbb{P}\left(\max_{c_* N \leq i \leq (1-c_*)N} |\lambda_i(H_t) - \gamma_i| > N^{-1+\varepsilon}\right) \leq C N^{-D}. \quad (3.3)$$

(c) For every integer $n \geq 1$, there exists $C_n > 0$ such that $\mathbb{E}\|H_t\|^n \leq C_n$ for all $N \geq 2$ and $t \in [0, 1]$.

The same statements hold with H_t replaced by G .

The next lemma is an edge-inclusive rigidity estimate with subpolynomial precision, due to Erdős, Yau, and Yin [29]. It is used only for the coupling, in the proof of Proposition 3.5 and of its Hermitian analogue in Section 9.

Lemma 3.3. *There exists $A_0 > 0$ such that, uniformly in $t \in [0, 1]$, with overwhelming probability,*

$$|\lambda_i(H_t) - \gamma_i| \leq e^{A_0(\log \log N)^2} N^{-2/3} \min(i, N+1-i)^{-1/3} \quad \text{for all } i \in [N]. \quad (3.4)$$

The same statement holds with H_t replaced by G .

Proof of Proposition 3.2. We deduce the proposition from the results of Benaych-Georges and Knowles [4]. Their Definition 2.2 concerns a sequence $A = A^{(N)}$ of $N \times N$ Hermitian random matrices with the following three properties.

- (1) The entries A_{ab} with $a \leq b$ are independent.
- (2) We have $\mathbb{E}A_{ab} = 0$ for all a, b , $\mathbb{E}|A_{ab}|^2 = N^{-1}$ for $a \neq b$, and $\mathbb{E}|A_{aa}|^2 \leq C N^{-1}$ for all a .
- (3) For every integer $n \geq 1$, there exists $C_n > 0$ such that $\mathbb{E}|N^{1/2}A_{ab}|^n \leq C_n$ for all N, a , and b .

We first check that (1)–(3) hold for $(H_{t_N})_{N \geq 2}$, for every sequence (t_N) in $[0, 1]$, with constants independent of the sequence. The upper-triangular entries of H_t are independent and centered, since H and G are independent and each has independent centered upper-triangular entries. For $a \leq b$, we have

$$N^{1/2}(H_t)_{ab} = e^{-t/2}X_{ab} + (1 - e^{-t})^{1/2}N^{1/2}G_{ab}, \quad (3.5)$$

and therefore

$$\mathbb{E}(H_t)_{ab}^2 = \frac{1}{N} \quad (a \neq b), \quad \mathbb{E}(H_t)_{aa}^2 = \frac{e^{-t}\sigma_Z^2 + 2(1 - e^{-t})}{N} \leq \frac{\max(\sigma_Z^2, 2)}{N}.$$

Using (3.5), Minkowski's inequality, and the sub-Gaussian tails of X , Z , and $N^{1/2}G_{ab}$, we see that the moments of $N^{1/2}(H_t)_{ab}$ are bounded uniformly in t , N , a , and b . The GOE matrix G satisfies (1)–(3) by [4, Example 2.3].

Next, let A be a sequence with the properties (1)–(3). We claim that for all $\varepsilon \in (0, 1/4]$, $c_* \in (0, 1/2)$, and $D > 0$, there exists N_0 such that for all $N \geq N_0$,

$$\mathbb{P}(A \notin \mathcal{G}_\varepsilon) + \mathbb{P}\left(\max_{c_*N \leq i \leq (1-c_*)N} |\lambda_i(A) - \gamma_i| > N^{-1+\varepsilon}\right) \leq N^{-D}. \quad (3.6)$$

We first show that the claim implies (a) and (b). The events in (a) and (b) decrease as ε increases, so it suffices to consider $\varepsilon \leq 1/4$. Suppose that (a) or (b) fails for some ε , c_* , and D . Then for infinitely many N , there exists $t_N \in [0, 1]$ such that the left side of (3.6) with $A = H_{t_N}$ is larger than N^{-D} , since otherwise a suitable constant would give the bounds in (a) and (b) for all N and t . Setting $t_N = 0$ for all other N , we obtain a sequence (H_{t_N}) with the properties (1)–(3) for which (3.6) fails, which is a contradiction. Applying (3.6) to $A = G$, we also obtain (a) and (b) for G .

We now prove the claim, and we begin with the norm and the diagonal resolvent entries. By [4, Theorem G.1], we have $\|A\| \leq 3$ with overwhelming probability. For $w = E + i\eta$ with $\eta > 0$, let

$$m_{\text{sc}}(w) = \int_{-2}^2 \frac{\rho_{\text{sc}}(E')}{E' - w} dE', \quad \mathfrak{r}(w) = \left(\frac{\text{Im } m_{\text{sc}}(w)}{N\eta}\right)^{1/2} + \frac{1}{N\eta}$$

denote the Stieltjes transform of the semicircle law and the error parameter of [4, (2.16)], and recall that $|m_{\text{sc}}(w)| \leq 1$. We also set

$$\mathcal{W}_\varepsilon = \{E + i\eta : |E| \leq 4/\varepsilon, N^{-1+\varepsilon/4} \leq \eta \leq 4/\varepsilon\}.$$

The local semicircle law [4, Theorem 2.6], in the form [4, (2.17)] that holds simultaneously for all spectral parameters, applied with the spectral domain $\mathcal{W}_\varepsilon$, shows that with overwhelming probability

$$|((A - w)^{-1})_{aa} - m_{\text{sc}}(w)| \leq N^{\varepsilon/4} \mathfrak{r}(w) \quad \text{for all } a \in [N] \text{ and } w \in \mathcal{W}_\varepsilon.$$

We have $\mathfrak{r} \leq 2$ on $\mathcal{W}_\varepsilon$, as $\text{Im } m_{\text{sc}} \leq 1$ and $N\eta \geq 1$ there. Therefore, with overwhelming probability,

$$\max_{a \in [N]} \sup_{w \in \mathcal{W}_\varepsilon} |((A - w)^{-1})_{aa}| \leq 3N^{\varepsilon/4}. \quad (3.7)$$

Let $\mathcal{U}$ denote the event that $\|A\| \leq 3$ and (3.7) hold. We will show that conditions (ii) and (iii) of Definition 3.1 hold on $\mathcal{U}$ for $N \geq N_0(\varepsilon)$. These deductions from the local

law are standard (compare the proof of [4, Theorem 2.10]). We write them out because condition (ii) concerns every unit eigenvector, including the eigenvectors of multiple eigenvalues.

We first derive (ii). For an eigenvalue λ of A , let Π_λ denote the orthogonal projection onto its eigenspace, so that

$$(A - w)^{-1} = \sum_{\lambda'} \frac{\Pi_{\lambda'}}{\lambda' - w}, \quad w \in \mathbb{C} \setminus \mathbb{R},$$

where the sum runs over the distinct eigenvalues of A . Let v be a unit eigenvector of A with eigenvalue λ , and let $a \in [N]$. For all $\eta > 0$, we have

$$|v(a)|^2 = |(\Pi_\lambda e_a) \cdot v|^2 \leq (\Pi_\lambda)_{aa} \leq \eta \sum_{\lambda'} \frac{\eta (\Pi_{\lambda'})_{aa}}{(\lambda' - \lambda)^2 + \eta^2} = \eta \operatorname{Im} ((A - \lambda - i\eta)^{-1})_{aa}. \quad (3.8)$$

Here the first equality follows from $v = \Pi_\lambda v$, the first inequality follows from the Cauchy-Schwarz inequality and $\|\Pi_\lambda e_a\|_2^2 = (\Pi_\lambda)_{aa}$, and the second inequality follows from the fact that $(\Pi_{\lambda'})_{aa} \geq 0$ for every eigenvalue λ' .

We now use (3.8) on $\mathcal{U}$. There we have $|\lambda| \leq 3 \leq 4/\varepsilon$, so $\lambda + iN^{-1+\varepsilon/4} \in \mathcal{W}_\varepsilon$. Using (3.8) with $\eta = N^{-1+\varepsilon/4}$ and (3.7), we obtain

$$|v(a)|^2 \leq 3N^{-1+\varepsilon/2} \leq N^{-1+2\varepsilon}$$

for $N \geq N_0(\varepsilon)$, which gives (ii).

Next, we derive (iii). Fix $E \in \mathbb{R}$, and note that the set in (iii) is empty on $\mathcal{U}$ unless $[E, E + N^{-1}]$ intersects $[-3, 3]$, so we may assume this. Then $E' = E + (2N)^{-1}$ satisfies $|E'| \leq 4$. We set $\eta = N^{-1+\varepsilon/4}$, so that $E' + i\eta \in \mathcal{W}_\varepsilon$. Every i with $\lambda_i(A) \in [E, E + N^{-1}]$ satisfies

$$|\lambda_i(A) - E'| \leq (2N)^{-1} \leq \eta, \quad \text{and hence} \quad \frac{\eta}{(\lambda_i(A) - E')^2 + \eta^2} \geq \frac{1}{2\eta}.$$

Summing over these i , we obtain

$$\begin{aligned} \#\{i : \lambda_i(A) \in [E, E + N^{-1}]\} &\leq 2\eta \sum_{i=1}^N \frac{\eta}{(\lambda_i(A) - E')^2 + \eta^2} \\ &= 2\eta \sum_{a=1}^N \operatorname{Im} ((A - E' - i\eta)^{-1})_{aa} \leq 6N^{\varepsilon/2}, \end{aligned}$$

where we used (3.7) in the last step. The right side is bounded by N^ε for $N \geq N_0(\varepsilon)$, which gives (iii).

As $\|A\| \leq 3$ on $\mathcal{U}$, we conclude that the first term in (3.6) is bounded by $\mathbb{P}(\mathcal{U}^c) \leq N^{-D}/2$ for $N \geq N_0$.

We now prove the rigidity estimate in (3.6). The eigenvalues in [4] are labeled in decreasing order, and the typical location γ_i^{BK} of the i -th largest eigenvalue is defined there by the first relation in

$$N \int_{\gamma_i^{\text{BK}}}^2 \rho_{\text{sc}}(E) dE = i - \frac{1}{2}, \quad N \int_{-2}^{\gamma'_i} \rho_{\text{sc}}(E) dE = i - \frac{1}{2}.$$

The second relation holds for $\gamma'_i = \gamma_{N+1-i}^{\text{BK}}$. After relabeling, [4, Theorem 2.9] and a union bound over i show that with overwhelming probability

$$|\lambda_i(A) - \gamma'_i| \leq N^{\varepsilon/2} N^{-2/3} \min(i, N+1-i)^{-1/3} \quad \text{for all } i \in [N].$$

Let $c_* N \leq i \leq (1 - c_*)N$ and $N \geq 2/c_*$. Then i/N and $(i - 1/2)/N$ belong to $[c_*/2, 1 - c_*/2]$, and ρ_{sc} is bounded below on the energies whose quantiles lie in this interval, so the mean value theorem gives $|\gamma_i - \gamma'_i| \leq CN^{-1}$. Since $\min(i, N+1-i) \geq c_* N$, we conclude that with overwhelming probability

$$|\lambda_i(A) - \gamma_i| \leq |\lambda_i(A) - \gamma'_i| + |\gamma'_i - \gamma_i| \leq N^{\varepsilon/2} c_*^{-1/3} N^{-1} + CN^{-1} \leq N^{-1+\varepsilon}$$

for all such i and all $N \geq N_0$. This completes the proof of the claim.

Finally, we prove (c). By [4, Theorem H.1], if an $N \times N$ Hermitian matrix A with independent centered upper-triangular entries satisfies $\mathbb{E} \exp(c' N |A_{ab}|^2) \leq 1/c'$ for all a, b and some $c' > 0$, then

$$\mathbb{P}(\|A\| \geq x) \leq \exp(-c(x^2 - C)N) \quad \text{for all } x > 0,$$

where $c, C > 0$ depend only on c' . By (3.5), we have

$$N|(H_t)_{ab}|^2 \leq 2X_{ab}^2 + 2NG_{ab}^2,$$

so the independence of H and G and the sub-Gaussian tails of X, Z , and $N^{1/2}G_{ab}$ show that this hypothesis holds for H_t with some $c' > 0$ independent of t , and also for G . Setting $x_0 = (2C)^{1/2}$, we find

$$\mathbb{E}\|H_t\|^n = \int_0^\infty nx^{n-1} \mathbb{P}(\|H_t\| \geq x) dx \leq x_0^n + \int_{x_0}^\infty nx^{n-1} e^{-cx^2 N/2} dx \leq C_n,$$

where we used that $x^2 - C \geq x^2/2$ for $x \geq x_0$. The same computation applies to G . $\square$

Proof of Lemma 3.3. We use [29, Theorem 2.2]. It concerns an $N \times N$ real symmetric matrix A with independent centered upper-triangular entries whose variances $\sigma_{ab}^2 = \mathbb{E}A_{ab}^2$ satisfy the following three conditions.

- (1) The variances in each column sum to one, which is [29, (2.2)].

- (2) There exist constants $0 < C_- \leq C_+$ with $C_- \leq N\sigma_{ab}^2$ and $N\sigma_{ab}^2 \leq C_+$ for all $a, b \in [N]$. By [29, Example 1], this implies the assumption [29, (2.3)]. Further, it implies [29, (2.4)] directly.
- (3) The variables A_{ab}/σ_{ab} have uniformly subexponential tails in the sense of [29, (2.17)].

The classical locations in [29] coincide with our γ_i , and the eigenvalues there are labeled in increasing order. Let $A_1 > 1$ be the constant in the lower bound on the parameter L in [29, (2.18)]. Under the three conditions above, [29, (2.25)] with $L = A_1 \log \log N$ states that

$$\begin{aligned} \mathbb{P}\left(\exists i \in [N] : |\lambda_i(A) - \gamma_i| \geq e^{A_1(\log \log N)^2} N^{-2/3} \min(i, N+1-i)^{-1/3}\right) \\ \leq C \exp\left(-c e^{c_1 A_1(\log \log N)^2}\right) \end{aligned} \quad (3.9)$$

for $N \geq N_0$, where $c_1 \in (0, 1)$, and where A_1, c, c_1, C , and N_0 depend only on $C_\pm$ and on the constants in [29, (2.17)]. Here we used that $(\log N)^L = e^{A_1(\log \log N)^2}$, and that the upper bound on L in [29, (2.18)] holds for N large. Since $e^{c_1 A_1(\log \log N)^2}$ grows faster than every multiple of $\log N$, the right side of (3.9) is bounded by $C_D N^{-D}$ for every $D > 0$.

The matrix H_t violates the first condition by a term of order N^{-1} , and we correct this by a scalar factor. Let

$$\sigma_t^2 = e^{-t} \sigma_Z^2 + 2(1 - e^{-t}), \quad r_{N,t} = 1 + \frac{\sigma_t^2 - 1}{N},$$

so that σ_t^2 is the variance of $N^{1/2}(H_t)_{aa}$. By the computation of the variances in the proof of Proposition 3.2, we have

$$\sum_{a=1}^N \mathbb{E}(H_t)_{ab}^2 = \frac{N-1}{N} + \frac{\sigma_t^2}{N} = r_{N,t} \quad \text{for all } b \in [N],$$

so the matrix $A = r_{N,t}^{-1/2} H_t$ satisfies the first condition.

Recall that $\sigma_Z^2 > 0$ (this is needed for the lower bound below). It follows that for $N \geq 2$,

$$0 < \min(\sigma_Z^2, 2) \leq \sigma_t^2 \leq \max(\sigma_Z^2, 2), \quad \frac{1}{2} \leq r_{N,t} \leq 1 + \max(\sigma_Z^2, 2).$$

The variances of the entries of A satisfy

$$N\sigma_{ab}^2 = r_{N,t}^{-1} \quad (a \neq b), \quad N\sigma_{aa}^2 = \sigma_t^2 r_{N,t}^{-1},$$

so we obtain the second condition in the form

$$\frac{\min(\sigma_Z^2, 1)}{1 + \max(\sigma_Z^2, 2)} \leq N\sigma_{ab}^2 \leq 2 \max(\sigma_Z^2, 2) \quad \text{for all } a, b \in [N].$$

We next check the third condition. By (3.5), we have

$$\frac{A_{ab}}{\sigma_{ab}} = e^{-t/2} X_{ab} + (1 - e^{-t})^{\frac{1}{2}} N^{\frac{1}{2}} G_{ab} \quad (a \neq b), \quad \frac{A_{aa}}{\sigma_{aa}} = \frac{e^{-t/2} Z_a + (1 - e^{-t})^{\frac{1}{2}} N^{\frac{1}{2}} G_{aa}}{\sigma_t}.$$

These variables have sub-Gaussian tails with constants independent of t , N , a , and b , as $\sigma_t^2 \geq \min(\sigma_Z^2, 2)$. This implies [29, (2.17)] with constants independent of t . Hence (3.9) holds for $A = r_{N,t}^{-1/2} H_t$, with constants independent of t .

We now return to H_t . For $N \geq N_0$, we have $|\lambda_i(A)| \leq 3$ for all i on the complement of the event in (3.9). Since $\lambda_i(H_t) = r_{N,t}^{1/2} \lambda_i(A)$, we find on this complement that

$$\begin{aligned} |\lambda_i(H_t) - \lambda_i(A)| &= |r_{N,t}^{1/2} - 1| |\lambda_i(A)| \leq 3|r_{N,t} - 1| \\ &\leq CN^{-1} \leq CN^{-2/3} \min(i, N+1-i)^{-1/3}, \quad i \in [N], \end{aligned}$$

where we used $|r_{N,t}^{1/2} - 1| \leq |r_{N,t} - 1|$. Combining this with (3.9), we obtain on the same complement that, for N large and all $i \in [N]$,

$$\begin{aligned} |\lambda_i(H_t) - \gamma_i| &\leq (e^{A_1(\log \log N)^2} + C) N^{-2/3} \min(i, N+1-i)^{-1/3} \\ &\leq e^{(A_1+1)(\log \log N)^2} N^{-2/3} \min(i, N+1-i)^{-1/3}. \end{aligned}$$

This is (3.4) with $A_0 = A_1 + 1$. For G , the same argument applies with σ_t^2 replaced by 2. $\square$

We next record the tail bound of Nguyen, Tao, and Vu [57] for the distance between eigenvalues whose labels differ by two.

Proposition 3.4. *For every $C_1 > 0$, there exists $C > 0$ such that*

$$\sup_{1 \leq i \leq N-2} \mathbb{P}(\lambda_{i+2} - \lambda_i \leq \delta/N) \leq C\delta^3 \quad \text{for all } \delta \geq N^{-C_1}. \quad (3.10)$$

Further, for $r > 0$, let $\mathcal{N}_3(r)$ denote the number of $i \in [N-2]$ such that $\lambda_{i+2} - \lambda_i \leq r$. Then $\mathbb{E}\mathcal{N}_3(r) \leq CN(Nr)^3$ for all $r \geq N^{-1-C_1}$.

Proof. This is [57, Corollary 2.5] with $l = 2$, for which the exponent c_l there equals 3, translated to our normalization. That corollary concerns real symmetric Wigner matrices in the sense of [57, Section 1], whose strictly upper-triangular entries are independent and identically distributed sub-Gaussian random variables with mean zero and variance one, and whose diagonal entries are independent sub-Gaussian random variables with mean zero and variance at most $N^{1-o(1)}$, independent of the strictly upper-triangular entries. These hypotheses allow atomic entry distributions. The matrix $N^{1/2}H$, whose entries are the X_{ab} and the Z_a , is of this form, since Z_a has variance σ_Z^2 .

For every $C_1 > 0$, [57, Corollary 2.5] gives $C > 0$ such that

$$\sup_{1 \leq i \leq N-2} \mathbb{P}(\lambda_{i+2}(N^{1/2}H) - \lambda_i(N^{1/2}H) \leq \delta N^{-1/2}) \leq C\delta^3 \quad \text{for all } \delta \geq N^{-C_1}.$$

As $\lambda_i(N^{1/2}H) = N^{1/2}\lambda_i$, this is (3.10). Applying (3.10) with $\delta = Nr$, we obtain

$$\mathbb{E}\mathcal{N}_3(r) = \sum_{i=1}^{N-2} \mathbb{P}(\lambda_{i+2} - \lambda_i \leq r) \leq CN(Nr)^3, \quad r \geq N^{-1-C_1}.$$

□

We next describe the coupling of the gaps of H_t with those of the GOE, which is due to Bourgade [10]. Let $B = (B(t))_{t \geq 0}$ be an $N \times N$ real symmetric matrix of Brownian motions, independent of H and G , such that the processes B_{ab} with $a < b$ and $2^{-1/2}B_{aa}$ with $a \in [N]$ are independent standard Brownian motions. Let $(H_t^{\text{dyn}})_{t \geq 0}$ denote the matrix Ornstein–Uhlenbeck process

$$H_t^{\text{dyn}} = e^{-t/2}H + N^{-1/2} \int_0^t e^{-(t-t')/2} dB(t'), \quad (3.11)$$

which is the solution of $dH_t^{\text{dyn}} = N^{-1/2} dB(t) - 2^{-1}H_t^{\text{dyn}} dt$ with $H_0^{\text{dyn}} = H$. For each $t \geq 0$, the stochastic integral in (3.11) is a centered Gaussian real symmetric matrix, independent of H , whose upper-triangular entries are independent with variance $(1 - e^{-t})N^{-1}$ off the diagonal and $2(1 - e^{-t})N^{-1}$ on the diagonal. Hence H_t^{dyn} has the law of H_t for $t \in [0, 1]$. We write $\lambda_i(t) = \lambda_i(H_t^{\text{dyn}})$.

Let $\mathcal{E}_0$ denote the event that the eigenvalues of H are distinct. As recalled in [10, Section 1.4], the eigenvalues of H_t^{dyn} form, on $\mathcal{E}_0$, the unique strong solution of the system

$$d\lambda_i(t) = \frac{d\beta_i(t)}{N^{1/2}} + \left(\frac{1}{N} \sum_{j \neq i} \frac{1}{\lambda_i(t) - \lambda_j(t)} - \frac{\lambda_i(t)}{2} \right) dt, \quad i \in [N], \quad (3.12)$$

where $\beta_1, \dots, \beta_N$ are independent Brownian motions with $\mathbb{E}\beta_i(t)^2 = 2t$. We extend this definition to the complement of $\mathcal{E}_0$ by letting $\beta_1, \dots, \beta_N$ be independent Brownian motions with the same variance there, independent of all other random variables. We let $\mu(t) = (\mu_1(t), \dots, \mu_N(t))$ denote the solution of (3.12) driven by the same β_i , with initial condition $\mu_i(0) = \lambda_i(G)$. The spectrum of G is simple almost surely, and the β_i are independent of G , since B is independent of (H, G) . The law of the spectrum of G is invariant for (3.12) (see the discussion of [10, (2.2)]), so $\mu(t)$ has the law of the spectrum of G for every $t \geq 0$. This is the coupling of [10, (1.7)–(1.8)], with $\lambda(t)$ and $\mu(t)$ corresponding to the parameter values 0 and 1 of the interpolation there.

Proposition 3.5. *For all $c_* \in (0, 1/2)$, $\varepsilon > 0$, and $D > 0$, there exists N_0 such that for all $N \geq N_0$ and $N^{-1+\varepsilon} \leq t < 1$,*

$$\mathbb{P}\left(\max_{c_*N \leq i \leq (1-c_*)N} |(\lambda_{i+1}(t) - \lambda_i(t)) - (\mu_{i+1}(t) - \mu_i(t))| > \frac{N^\varepsilon}{N^2 t}\right) \leq N^{-D}. \quad (3.13)$$

In particular, for every such t , the spectrum of H_t and that of a GOE matrix can be coupled so that the event in (3.13) has probability at most N^{-D} .

Proof. This follows from [10, Corollary 3.3] and a union bound, and we now check its hypotheses. That corollary states the following. For all $c_*, \varepsilon, D > 0$, there exist $C, N_0 > 0$ such that for all $N \geq N_0$, all t with $e^{C(\log \log N)^2} N^{-1} < t < 1$, and all i with $c_*N \leq i \leq (1 - c_*)N$,

$$\mathbb{P}\left(|(\lambda_{i+1}(t) - \lambda_i(t)) - (\mu_{i+1}(t) - \mu_i(t))| > \frac{N^\varepsilon}{N^2 t}\right) \leq N^{-D}. \quad (3.14)$$

In [10], the initial condition $\lambda(0)$ is the spectrum of a generalized Wigner matrix in the sense of [10, Definition 1.1], $\mu(0)$ is the spectrum of a GOE matrix, and the initial points are distinct.

First, the variances of the entries of a generalized Wigner matrix in [10, Definition 1.1] have row sums exactly one, whereas those of H have row sums $1 + (\sigma_Z^2 - 1)/N$. However, [10, Remarks 2.9 and 3.2] state that the results hold, for every value of the symmetry parameter in [10, (2.2)] that is at least one, whenever the initial conditions satisfy optimal rigidity estimates. Indeed, the proofs of [10, Theorem 3.1 and Corollary 3.3] use the initial spectra only through the rigidity event of [10, (2.7)] and the lower bound $1 - N^{-D}$ on its probability in [10, Lemma 2.3]. (The GOE initial condition used there also has row sums $1 + N^{-1}$.) This event requires that for every initial condition on the segment between $\lambda(0)$ and $\mu(0)$, the solution x of (3.12) with this initial condition satisfies

$$|x_i(t') - \gamma_i| < e^{C'(\log \log N)^2/2} N^{-2/3} \min(i, N+1-i)^{-1/3}$$

for all $i \in [N]$ and $t' \in [0, 1]$, where C' is a sufficiently large constant.

As explained before [10, Lemma 2.3], the lower bound on the probability of this event follows from rigidity estimates at fixed times for the spectra of $H_{t'}^{\text{dyn}}$ with $t' \in [0, 1]$ and of a GOE matrix (the solutions started at the endpoints $\lambda(0)$ and $\mu(0)$). The passage to all times and to the whole segment uses a discretization in time and along the segment, Weyl's inequality, and a maximum principle for the derivative of the solution along the segment. Recall that $H_{t'}^{\text{dyn}}$ has the law of $H_{t'}$. For our initial data, the fixed-time estimates are therefore given by Lemma 3.3, applied to $H_{t'}$ and to G , since

$$e^{A_0(\log \log N)^2} \leq \frac{1}{2} e^{C'(\log \log N)^2/2}$$

for $C' > 2A_0$ and N large. The remaining steps use only (3.11), (3.12), and bounds on $\|H_t^{\text{dyn}}\|$, which follow from Proposition 3.2, so the conclusion of [10, Lemma 2.3] holds for our initial data as well.

Second, the system (3.12) is set up in [10] for distinct initial points, which holds on $\mathcal{E}_0$. By [57, Corollary 2.2], applied to $N^{1/2}H$ as in the proof of Proposition 3.4, for every $C_1 > 0$ there exists $C_{\text{NTV}} > 0$ such that

$$\mathbb{P}(\lambda_{i+1} - \lambda_i \leq N^{-1-C_1}) \leq C_{\text{NTV}} N^{-C_1} \quad \text{for all } i \in [N-1].$$

A union bound over i gives

$$\mathbb{P}(\mathcal{E}_0^c) \leq \sum_{i=1}^{N-1} \mathbb{P}(\lambda_{i+1} - \lambda_i \leq N^{-1-C_1}) \leq C_{\text{NTV}} N^{1-C_1},$$

and since C_1 is arbitrary, $\mathcal{E}_0$ holds with overwhelming probability. (See also [71].) We apply [10, Corollary 3.3] on $\mathcal{E}_0$ and include $\mathcal{E}_0^c$ in the exceptional event.

Third, (3.14) concerns a single index i . We apply it with $D+2$ in place of D and take a union bound over the at most N indices i with $c_*N \leq i \leq (1-c_*)N$. Together with $\mathbb{P}(\mathcal{E}_0^c) \leq N^{-D-1}$ for large N , this bounds the probability in (3.13) by

$$N \cdot N^{-D-2} + \mathbb{P}(\mathcal{E}_0^c) \leq 2N^{-D-1} \leq N^{-D}.$$

Finally, $e^{C(\log \log N)^2} N^{-1} < N^{-1+\varepsilon}$ for N large, so every $t \in [N^{-1+\varepsilon}, 1)$ belongs to the range of [10, Corollary 3.3]; the upper limit $t < 1$ is strict there. The last statement of the proposition holds because $\lambda(t)$ has the law of the spectrum of H_t and $\mu(t)$ has the law of the spectrum of G . $\square$

For $\alpha \in (0, 1/2)$, we set

$$\ell_\alpha = N^{-1-\alpha}, \quad M_\alpha = N^{1-2\alpha} = N^3 \ell_\alpha^2. \quad (3.15)$$

The next proposition gives the pair statistics of the GOE at the scale ℓ_α for $\alpha < 1/4$. We use it for the base case in Section 6, and we prove it in Appendix A.

Proposition 3.6. *Let $\alpha \in (0, 1/4)$ and $p > 1$. Let $f: [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$ be a continuous function for which there exist a bulk interval I and a constant $C_f > 0$ such that $f(r, E) = 0$ for $E \notin I$ and $|f(r, E)| \leq C_f(1+r^2)^{-p}$ for all (r, E) . Then, in L^1 and hence in probability,*

$$\frac{1}{M_\alpha} \sum_{1 \leq i < j \leq N} f\left(\frac{\lambda_j(G) - \lambda_i(G)}{\ell_\alpha}, \frac{\lambda_i(G) + \lambda_j(G)}{2}\right) \longrightarrow \int_{-2}^2 \int_0^\infty c(E) r f(r, E) dr dE. \quad (3.16)$$

3.2 Eigenvalues of entrywise small perturbations

We now consider a diagonal matrix perturbed by a Hermitian matrix with small entries. We group the diagonal entries into clusters of nearby points, and we compare the eigenvalues near each cluster with those of the corresponding diagonal block of the perturbed matrix. In Section 6, Section 7, and Section 9, the diagonal matrix contains the eigenvalues of H (multiplied by $e^{-t/2}$ in Section 6 and Section 9), and the perturbation is a Gaussian matrix or the resampling perturbation Δ , written in the eigenbasis of H .

We use the following notation in this section. For an $N \times N$ matrix A and nonempty sets $\mathcal{K}, \mathcal{K}' \subset [N]$, we let $A_{\mathcal{K}\mathcal{K}'} = (A_{ij})_{i \in \mathcal{K}, j \in \mathcal{K}'}$ denote the corresponding submatrix, with rows and columns in increasing order, and we set $\mathcal{K}^c = [N] \setminus \mathcal{K}$. For a Hermitian matrix A , we let $n_-(A)$ denote the number of negative eigenvalues of A , counted with multiplicity. For a matrix M , we let $\|M\|_{\infty \rightarrow \infty} = \max_i \sum_j |M_{ij}|$ denote its operator norm on ℓ^∞ .

We will use the inertia additivity formula of Haynsworth [39]. Let M be an $N \times N$ Hermitian matrix and let $\mathcal{K} \subset [N]$ be a nonempty proper subset such that $M_{\mathcal{K}^c\mathcal{K}^c}$ is invertible. Let

$$M/M_{\mathcal{K}^c\mathcal{K}^c} = M_{\mathcal{K}\mathcal{K}} - M_{\mathcal{K}\mathcal{K}^c}M_{\mathcal{K}^c\mathcal{K}^c}^{-1}M_{\mathcal{K}^c\mathcal{K}}$$

denote the Schur complement. After a permutation of coordinates, M is congruent to the block diagonal matrix with blocks $M/M_{\mathcal{K}^c\mathcal{K}^c}$ and $M_{\mathcal{K}^c\mathcal{K}^c}$. Hence Sylvester's law of inertia [40] and the multiplicativity of the determinant give

$$n_-(M) = n_-(M_{\mathcal{K}^c\mathcal{K}^c}) + n_-(M/M_{\mathcal{K}^c\mathcal{K}^c}), \quad \det M = \det M_{\mathcal{K}^c\mathcal{K}^c} \cdot \det(M/M_{\mathcal{K}^c\mathcal{K}^c}). \quad (3.17)$$

The following lemma compares the eigenvalues of $\Lambda_0 + W$ near a group of eigenvalues of Λ_0 with those of the compression $(\Lambda_0 + W)_{\mathcal{K}\mathcal{K}}$, in order and with multiplicity. It is related to the quadratic residual bounds of [50, 55], and its proof is elementary.

Lemma 3.7. *Let Λ_0 and W be $N \times N$ Hermitian matrices, and let $\mathcal{K} \subset [N]$ be a nonempty proper subset such that $(\Lambda_0)_{\mathcal{K}\mathcal{K}^c} = 0$. Let $E_- \leq E_+$ be real numbers such that every eigenvalue of $(\Lambda_0)_{\mathcal{K}\mathcal{K}}$ belongs to $[E_-, E_+]$, let $\vartheta > 0$, and set $\mathcal{R} = [E_- - \vartheta, E_+ + \vartheta]$. Suppose that $(\Lambda_0)_{\mathcal{K}^c\mathcal{K}^c} + W_{\mathcal{K}^c\mathcal{K}^c} - x$ is invertible for all $x \in \mathcal{R}$, and that there exists $\chi \geq 0$ such that*

$$\|W_{\mathcal{K}\mathcal{K}^c}((\Lambda_0)_{\mathcal{K}^c\mathcal{K}^c} + W_{\mathcal{K}^c\mathcal{K}^c} - x)^{-1}W_{\mathcal{K}^c\mathcal{K}}\| \leq \chi \quad \text{for all } x \in \mathcal{R}, \quad \|W_{\mathcal{K}\mathcal{K}}\| + \chi < \vartheta. \quad (3.18)$$

Let i_0 denote the number of eigenvalues of $\Lambda_0 + W$ smaller than $E_- - \vartheta$, counted with multiplicity. Then the eigenvalues of $\Lambda_0 + W$ in $\mathcal{R}$, counted with multiplicity, are $\lambda_{i_0+1}(\Lambda_0 + W), \dots, \lambda_{i_0+|\mathcal{K}|}(\Lambda_0 + W)$. Moreover,

$$|\lambda_{i_0+i}(\Lambda_0 + W) - \lambda_i((\Lambda_0 + W)_{\mathcal{K}\mathcal{K}})| \leq \chi \quad \text{for all } i \in [|\mathcal{K}|], \quad (3.19)$$

and these eigenvalues belong to $[E_- - \|W_{\mathcal{K}\mathcal{K}}\| - \chi, E_+ + \|W_{\mathcal{K}\mathcal{K}}\| + \chi]$.

Proof. We abbreviate $B = (\Lambda_0 + W)_{\mathcal{K}\mathcal{K}}$, and for $x \in \mathcal{R}$, we set

$$\mathcal{S}(x) = W_{\mathcal{K}\mathcal{K}^c}((\Lambda_0)_{\mathcal{K}^c\mathcal{K}^c} + W_{\mathcal{K}^c\mathcal{K}^c} - x)^{-1}W_{\mathcal{K}^c\mathcal{K}}, \quad \mathcal{F}(x) = B - x - \mathcal{S}(x).$$

Observe that $\mathcal{S}(x)$ and $\mathcal{F}(x)$ are Hermitian, since x is real and $W_{\mathcal{K}^c\mathcal{K}} = (W_{\mathcal{K}\mathcal{K}^c})^*$. As $(\Lambda_0)_{\mathcal{K}\mathcal{K}^c} = 0$, the matrix $M = \Lambda_0 + W - x$ satisfies $M_{\mathcal{K}\mathcal{K}^c} = W_{\mathcal{K}\mathcal{K}^c}$, and $\mathcal{F}(x)$ is its Schur complement. By (3.17), we have

$$\begin{aligned} \#\{j \in [N] : \lambda_j(\Lambda_0 + W) < x\} &= n_-(\Lambda_0 + W - x) \\ &= n_-((\Lambda_0)_{\mathcal{K}^c\mathcal{K}^c} + W_{\mathcal{K}^c\mathcal{K}^c} - x) + n_-(\mathcal{F}(x)) \end{aligned} \quad (3.20)$$

for all $x \in \mathcal{R}$.

The matrix $(\Lambda_0)_{\mathcal{K}^c\mathcal{K}^c} + W_{\mathcal{K}^c\mathcal{K}^c}$ has no eigenvalue in the interval $\mathcal{R}$, so the first term on the right side of (3.20) is constant on $\mathcal{R}$. We denote this constant by i_1 .

We next compare $\mathcal{F}(x)$ with $B - x$. The bound $\|\mathcal{S}(x)\| \leq \chi$ gives

$$B - x - \chi \leq \mathcal{F}(x) \leq B - x + \chi$$

in the sense of quadratic forms. By the Courant–Fischer minimax principle [40], an inequality $A \leq A'$ between Hermitian matrices implies $\lambda_i(A) \leq \lambda_i(A')$ for all i , and hence $n_-(A) \geq n_-(A')$. Therefore

$$\#\{i : \lambda_i(B) < x - \chi\} \leq n_-(\mathcal{F}(x)) \leq \#\{i : \lambda_i(B) < x + \chi\} \quad \text{for all } x \in \mathcal{R}. \quad (3.21)$$

Using Weyl's inequality [40], the assumption on the spectrum of $(\Lambda_0)_{\mathcal{K}\mathcal{K}}$, and (3.18), we see that every eigenvalue of B satisfies

$$E_- - \vartheta + \chi < E_- - \|W_{\mathcal{K}\mathcal{K}}\| \leq \lambda_i(B) \leq E_+ + \|W_{\mathcal{K}\mathcal{K}}\| < E_+ + \vartheta - \chi. \quad (3.22)$$

We now identify the eigenvalues of $\Lambda_0 + W$ in $\mathcal{R}$. Taking $x = E_- - \vartheta$, we see from (3.22) that the right side of (3.21) vanishes, so (3.20) gives $i_0 = i_1$. Taking $x = E_+ + \vartheta$ instead, we see that the left side of (3.21) is $|\mathcal{K}|$. As $\mathcal{F}(x)$ is a $|\mathcal{K}| \times |\mathcal{K}|$ matrix, this gives $n_-(\mathcal{F}(E_+ + \vartheta)) = |\mathcal{K}|$, and $\Lambda_0 + W$ has exactly $i_0 + |\mathcal{K}|$ eigenvalues smaller than $E_+ + \vartheta$. Moreover, (3.22) gives

$$\mathcal{F}(E_+ + \vartheta) \leq B - E_+ - \vartheta + \chi < 0$$

in the sense of quadratic forms, so the determinant formula in (3.17) shows that $E_+ + \vartheta$ is not an eigenvalue of $\Lambda_0 + W$. Hence the eigenvalues of $\Lambda_0 + W$ in $\mathcal{R}$ are $\lambda_{i_0+1}(\Lambda_0 + W), \dots, \lambda_{i_0+|\mathcal{K}|}(\Lambda_0 + W)$.

It remains to prove (3.19). For brevity, we write $\lambda'_i = \lambda_{i_0+i}(\Lambda_0 + W)$ and $\lambda''_i = \lambda_i(B)$ for $i \in [|\mathcal{K}|]$. For $x \in \mathcal{R}$, the eigenvalues of $\Lambda_0 + W$ smaller than x are the i_0 eigenvalues smaller than $E_- - \vartheta$ and the λ'_i with $\lambda'_i < x$. Inserting this into (3.20) and using (3.21), we find

$$\#\{i : \lambda''_i < x - \chi\} \leq \#\{i : \lambda'_i < x\} \leq \#\{i : \lambda''_i < x + \chi\} \quad \text{for all } x \in \mathcal{R}. \quad (3.23)$$

Fix $i \in [|\mathcal{K}|]$, and suppose that $\lambda'_i < \lambda''_i - \chi$. Every $x \in (\lambda'_i, \lambda''_i - \chi)$ belongs to $\mathcal{R}$ by (3.22), since $\lambda'_i \geq E_- - \vartheta$. For such x , the middle term in (3.23) is at least i , since $\lambda'_1 \leq \dots \leq \lambda'_i < x$, while the right term is smaller than i , since $x + \chi < \lambda''_i \leq \dots \leq \lambda''_{|\mathcal{K}|}$. This contradicts (3.23). If $\lambda'_i > \lambda''_i + \chi$, then the same reasoning with $x \in (\lambda''_i + \chi, \lambda'_i)$ shows that the left term in (3.23) is at least i and the middle term is smaller than i , again a contradiction. This proves (3.19), and the final statement follows from (3.19) and (3.22). $\square$

We now apply Lemma 3.7 to a diagonal matrix with clustered entries. Let $y_1 \leq \dots \leq y_N$ be real numbers, and let $D_y = \text{diag}(y_1, \dots, y_N)$. We call a partition $\mathfrak{K}$ of $[N]$ into nonempty sets of consecutive integers a *family of clusters*, and we call its elements *clusters*. For a cluster $\mathcal{K}$ and $\vartheta > 0$, we let

$$\mathcal{R}_{\mathcal{K}} = \left\{ w \in \mathbb{C} : \min_{i \in \mathcal{K}} y_i - \vartheta \leq \text{Re } w \leq \max_{i \in \mathcal{K}} y_i + \vartheta, \quad |\text{Im } w| \leq \vartheta \right\} \quad (3.24)$$

denote the closed rectangle obtained by enlarging the interval $[\min_{i \in \mathcal{K}} y_i, \max_{i \in \mathcal{K}} y_i]$ by ϑ in each direction. We call ϑ the *margin* of the rectangles.

Lemma 3.8. *Let $y_1 \leq \dots \leq y_N$ be real numbers, let $\mathfrak{K}$ be a family of clusters, and let $k_0 \geq 1$ be such that $|\mathcal{K}| \leq k_0$ for all $\mathcal{K} \in \mathfrak{K}$. Let $\vartheta > 0$, and suppose that the rectangles $\mathcal{R}_{\mathcal{K}}$ with $\mathcal{K} \in \mathfrak{K}$ are pairwise disjoint and that there exists $\mathcal{L} \in [0, \infty)$ such that*

$$\sup_{\mathcal{K} \in \mathfrak{K}} \sup_{w \in \mathcal{R}_{\mathcal{K}}} \sum_{j \in [N] \setminus \mathcal{K}} \frac{1}{|y_j - w|} \leq \mathcal{L}, \quad (3.25)$$

with the convention $1/0 = \infty$. Let W be an $N \times N$ Hermitian matrix such that $\max_{i,j} |W_{ij}| \leq w_0$, where

$$w_0 \mathcal{L} \leq 1/2, \quad k_0 w_0 + 2k_0 w_0^2 \mathcal{L} < \vartheta. \quad (3.26)$$

Let $\mathcal{K} \in \mathfrak{K}$, and write $\mathcal{K} = \{i_0 + 1, \dots, i_0 + |\mathcal{K}|\}$. Then the eigenvalues of $D_y + W$ in $\mathcal{R}_{\mathcal{K}}$, counted with multiplicity, are $\lambda_{i_0+1}(D_y + W), \dots, \lambda_{i_0+|\mathcal{K}|}(D_y + W)$. Moreover,

$$|\lambda_{i_0+i}(D_y + W) - \lambda_i((D_y + W)_{\mathcal{K}\mathcal{K}})| \leq 2k_0 w_0^2 \mathcal{L} \quad \text{for all } i \in [|\mathcal{K}|], \quad (3.27)$$

and each of these eigenvalues is within distance $k_0 w_0 + 2k_0 w_0^2 \mathcal{L}$ of the interval $[\min_{i \in \mathcal{K}} y_i, \max_{i \in \mathcal{K}} y_i]$.

Since $|y_j - w| \geq |y_j - \text{Re } w|$, the supremum in (3.25) may be taken over the real points of the rectangles, and since $D_y + W$ is Hermitian, its eigenvalues in $\mathcal{R}_{\mathcal{K}}$ are those in $\mathcal{R}_{\mathcal{K}} \cap \mathbb{R}$. In applications, the main condition to verify is (3.25), which requires that the points outside each cluster are not too dense near it.

Proof. Fix $\mathcal{K} \in \mathfrak{K}$, and set $E_- = \min_{i \in \mathcal{K}} y_i$ and $E_+ = \max_{i \in \mathcal{K}} y_i$. We first treat the case $\mathcal{K} = [N]$, in which $N \leq k_0$. Bounding the operator norm by the Frobenius norm and using Weyl's inequality, we find

$$|\lambda_i(D_y + W) - y_i| \leq \|W\| \leq Nw_0 \leq k_0w_0 < \vartheta \quad \text{for all } i \in [N],$$

and the lemma follows, since $(D_y + W)_{\mathcal{K}\mathcal{K}} = D_y + W$.

Now suppose that $\mathcal{K}^c$ is nonempty. We will apply Lemma 3.7 with $\Lambda_0 = D_y$, $\chi = 2k_0w_0^2\mathcal{L}$, and the interval

$$\mathcal{R} = \mathcal{R}_{\mathcal{K}} \cap \mathbb{R} = [E_- - \vartheta, E_+ + \vartheta].$$

The matrix D_y is diagonal, so $(D_y)_{\mathcal{K}\mathcal{K}^c} = 0$, and the eigenvalues of $(D_y)_{\mathcal{K}\mathcal{K}}$ are the y_i with $i \in \mathcal{K}$, which belong to $[E_-, E_+]$.

We now check the invertibility hypothesis of Lemma 3.7. Fix $x \in \mathcal{R}$. The matrix $(D_y - x)_{\mathcal{K}^c\mathcal{K}^c}$ is diagonal with entries $y_j - x$, $j \in \mathcal{K}^c$. These entries are nonzero by (3.25), and

$$\|W_{\mathcal{K}^c\mathcal{K}^c}(D_y - x)^{-1}_{\mathcal{K}^c\mathcal{K}^c}\|_{\infty \rightarrow \infty} = \max_{i \in \mathcal{K}^c} \sum_{j \in \mathcal{K}^c} \frac{|W_{ij}|}{|y_j - x|} \leq w_0\mathcal{L} \leq \frac{1}{2}.$$

Hence the Neumann series shows that $\text{Id} + W_{\mathcal{K}^c\mathcal{K}^c}(D_y - x)^{-1}_{\mathcal{K}^c\mathcal{K}^c}$ is invertible. We denote its inverse by $\Theta(x)$, and the Neumann series also gives $\|\Theta(x)\|_{\infty \rightarrow \infty} \leq 2$. Using the factorization

$$(D_y + W)_{\mathcal{K}^c\mathcal{K}^c} - x = (\text{Id} + W_{\mathcal{K}^c\mathcal{K}^c}(D_y - x)^{-1}_{\mathcal{K}^c\mathcal{K}^c})(D_y - x)_{\mathcal{K}^c\mathcal{K}^c},$$

we see that this matrix is invertible, and that

$$((D_y + W)_{\mathcal{K}^c\mathcal{K}^c} - x)^{-1} = (D_y - x)^{-1}_{\mathcal{K}^c\mathcal{K}^c} \Theta(x). \quad (3.28)$$

We next bound the Schur remainder

$$\mathcal{S}(x) = W_{\mathcal{K}\mathcal{K}^c}((D_y + W)_{\mathcal{K}^c\mathcal{K}^c} - x)^{-1}W_{\mathcal{K}^c\mathcal{K}}$$

entrywise. Every entry of $W_{\mathcal{K}^c\mathcal{K}}$ is bounded by w_0 , so

$$|(\Theta(x)W_{\mathcal{K}^c\mathcal{K}})_{ji'}| \leq \|\Theta(x)\|_{\infty \rightarrow \infty} w_0 \leq 2w_0, \quad j \in \mathcal{K}^c, i' \in \mathcal{K}.$$

By (3.28), for $i, i' \in \mathcal{K}$ we have

$$|\mathcal{S}(x)_{ii'}| = \left| \sum_{j \in \mathcal{K}^c} \frac{W_{ij}}{y_j - x} (\Theta(x)W_{\mathcal{K}^c\mathcal{K}})_{ji'} \right| \leq 2w_0^2 \sum_{j \in \mathcal{K}^c} \frac{1}{|y_j - x|} \leq 2w_0^2\mathcal{L}.$$

Bounding the operator norms by the Frobenius norms, we obtain

$$\|\mathcal{S}(x)\| \leq |\mathcal{K}| \cdot 2w_0^2\mathcal{L} \leq \chi, \quad \|W_{\mathcal{K}\mathcal{K}}\| \leq |\mathcal{K}|w_0 \leq k_0w_0.$$

By (3.26), we have $\|W_{\mathcal{K}\mathcal{K}}\| + \chi < \vartheta$, so the hypotheses of Lemma 3.7 hold. It gives (3.27) and the location of the eigenvalues, with i_0 replaced by the number i'_0 of eigenvalues of $D_y + W$ smaller than $E_- - \vartheta$. Moreover, $\mathcal{R}_{\mathcal{K}} \cap \mathbb{R}$ contains exactly $|\mathcal{K}|$ eigenvalues of $D_y + W$.

It remains to show that $i'_0 = i_0$. The preceding argument applies to every cluster $\mathcal{K}'$ with $(\mathcal{K}')^c$ nonempty, and the case $\mathcal{K}' = [N]$ was treated above, so each interval $\mathcal{R}_{\mathcal{K}'} \cap \mathbb{R}$ with $\mathcal{K}' \in \mathfrak{K}$ contains exactly $|\mathcal{K}'|$ eigenvalues of $D_y + W$.

Let $\mathcal{K}', \mathcal{K}'' \in \mathfrak{K}$ with $\max \mathcal{K}' < \min \mathcal{K}''$. Then $\max_{i \in \mathcal{K}'} y_i \leq \min_{i \in \mathcal{K}''} y_i$, and since the two intervals are disjoint, every point of $\mathcal{R}_{\mathcal{K}'} \cap \mathbb{R}$ is smaller than every point of $\mathcal{R}_{\mathcal{K}''} \cap \mathbb{R}$. As the cluster sizes add up to N , every eigenvalue of $D_y + W$ belongs to one of these intervals. Therefore the eigenvalues smaller than $E_- - \vartheta$ are those in the intervals of the clusters $\mathcal{K}'$ with $\max \mathcal{K}' < \min \mathcal{K}$, and their number is $\min \mathcal{K} - 1 = i_0$. This completes the proof. $\square$

In Section 6 and Section 9, the clusters are formed by joining consecutive points at distance smaller than a threshold, and the number of points in intervals of length N^{-1} is controlled by condition (iii) of Definition 3.1. (In Definition 7.1, consecutive eigenvalues are joined when their distance is at most the threshold, while in Corollary 3.9 below, consecutive points are joined when their distance is less than the threshold. Each section uses one of the two conventions throughout, and Lemma 3.8 applies to both.) The next corollary records the parameters of Lemma 3.8 in this situation. Figure 3.1 illustrates its proof.

Corollary 3.9. *Let $N \geq 2$, $\varepsilon > 0$, and $q > 0$ satisfy $2N^\varepsilon q < N^{-1}$. Let $y_1 \leq \dots \leq y_N$ be real numbers in $[-4, 4]$ such that $\#\{i \in [N] : y_i \in [E, E + N^{-1}]\} \leq N^\varepsilon$ for all $E \in \mathbb{R}$. Let $\mathfrak{K}$ be the family of clusters whose elements are the maximal sets $\mathcal{K}$ of consecutive integers such that $y_{i+1} - y_i < q$ for all i with $i, i+1 \in \mathcal{K}$. Then every $\mathcal{K} \in \mathfrak{K}$ satisfies $|\mathcal{K}| \leq N^\varepsilon$, and the hypotheses of Lemma 3.8 on y and $\mathfrak{K}$ hold with*

$$k_0 = N^\varepsilon, \quad \vartheta = q/3, \quad \mathcal{L} = 12N^\varepsilon(q^{-1} + N \log N). \quad (3.29)$$

Let W be an $N \times N$ Hermitian matrix with $\max_{i,j} |W_{ij}| \leq w_0$, where

$$N^\varepsilon w_0(q^{-1} + N \log N) \leq 1/48. \quad (3.30)$$

Then for every $\mathcal{K} = \{i_0 + 1, \dots, i_0 + |\mathcal{K}|\} \in \mathfrak{K}$ and every $i \in [|\mathcal{K}|]$,

$$|\lambda_{i_0+i}(D_y + W) - \lambda_i((D_y + W)_{\mathcal{K}\mathcal{K}})| \leq 24N^{2\varepsilon}w_0^2(q^{-1} + N \log N), \quad (3.31)$$

and $\min_{j \in \mathcal{K}} y_j - q/32 \leq \lambda_{i_0+i}(D_y + W) \leq \max_{j \in \mathcal{K}} y_j + q/32$. In particular, if $i \in [N-1]$ and i and $i+1$ belong to different clusters, then $\lambda_{i+1}(D_y + W) - \lambda_i(D_y + W) \geq q/2$.

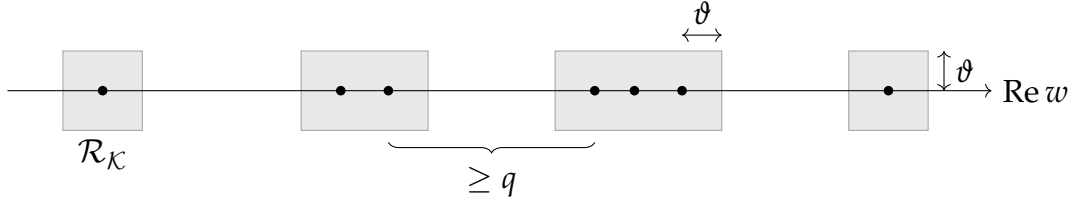


Figure 3.1. The points y_i (dots), the clusters obtained with the threshold q , and the rectangles $\mathcal{R}_K$ (shaded) with margin $\vartheta = q/3$, as in Corollary 3.9. Consecutive points in different clusters are at distance at least q , so the rectangles are disjoint, and every point outside a cluster is at distance at least $2q/3$ from the rectangle of that cluster.

Proof. We first bound the size of the clusters. Let $\mathcal{K} = \{i_0 + 1, \dots, i_0 + |\mathcal{K}|\} \in \mathfrak{K}$, and suppose that $|\mathcal{K}| > N^\varepsilon$. Set $i_1 = i_0 + \lfloor N^\varepsilon \rfloor + 1$, so that $i_1 \in \mathcal{K}$. Consecutive points in $\mathcal{K}$ are at distance smaller than q , so

$$y_{i_1} - y_{i_0+1} < (i_1 - i_0 - 1)q \leq N^\varepsilon q < N^{-1}.$$

Thus the interval $[y_{i_0+1}, y_{i_0+1} + N^{-1}]$ contains the $\lfloor N^\varepsilon \rfloor + 1 > N^\varepsilon$ points $y_{i_0+1}, \dots, y_{i_1}$, which contradicts the assumption. Hence $|\mathcal{K}| \leq N^\varepsilon = k_0$.

We next bound the distances from a rectangle to the points outside its cluster. Let $\mathcal{K} \in \mathfrak{K}$, set $E_- = \min_{i \in \mathcal{K}} y_i$ and $E_+ = \max_{i \in \mathcal{K}} y_i$, and fix $j \notin \mathcal{K}$. By the definition of $\mathfrak{K}$, we have $y_{i+1} - y_i \geq q$ whenever i and $i+1$ belong to different clusters. For $j < \min \mathcal{K}$, the labels $\min \mathcal{K} - 1$ and $\min \mathcal{K}$ belong to different clusters, and similarly for $j > \max \mathcal{K}$ and the labels $\max \mathcal{K}$ and $\max \mathcal{K} + 1$. Since $E_- = y_{\min \mathcal{K}}$ and $E_+ = y_{\max \mathcal{K}}$, we obtain

$$y_j \leq y_{\min \mathcal{K} - 1} \leq E_- - q \quad (j < \min \mathcal{K}), \quad y_j \geq y_{\max \mathcal{K} + 1} \geq E_+ + q \quad (j > \max \mathcal{K}).$$

Consequently, for every $w \in \mathcal{R}_K$,

$$|y_j - w| \geq |y_j - \operatorname{Re} w| \geq q - q/3 = 2q/3 \quad \text{for all } j \notin \mathcal{K}. \quad (3.32)$$

We next check that the rectangles are disjoint. Recall that $\vartheta = q/3$. Let $\mathcal{K}' \in \mathfrak{K}$ with $\min \mathcal{K}' > \max \mathcal{K}$. Then $\min_{i \in \mathcal{K}'} y_i \geq E_+ + q$, so every point of $\mathcal{R}_{\mathcal{K}'}$ has real part at least $E_+ + 2q/3$, while every point of $\mathcal{R}_K$ has real part at most $E_+ + q/3$. Hence the rectangles are pairwise disjoint.

We now bound the harmonic sum. Fix $\mathcal{K} \in \mathfrak{K}$ and $w \in \mathcal{R}_K$, and set $x = \operatorname{Re} w$. We have $|x| < 5$, since $|y_i| \leq 4$ for all i and $q < N^{-1} \leq 1$, and therefore $|y_j - x| < 9$ for all j . For each integer $n \geq 0$, let

$$\mathcal{M}_n = \{j \in [N] \setminus \mathcal{K} : n/N \leq |y_j - x| < (n+1)/N\}.$$

The points y_j with $j \in \mathcal{M}_n$ belong to the union of the two intervals $[x - (n+1)/N, x - n/N]$ and $[x + n/N, x + (n+1)/N]$ of length N^{-1} , so $|\mathcal{M}_n| \leq 2N^\varepsilon$ by assumption.

The labels $j \in \mathcal{M}_0$ satisfy $|y_j - x| \geq 2q/3$ by (3.32), and the labels $j \in \mathcal{M}_n$ with $n \geq 1$ satisfy $|y_j - x| \geq n/N$. Further, $\mathcal{M}_n$ is empty for $n \geq 9N$. Therefore

$$\sum_{j \notin \mathcal{K}} \frac{1}{|y_j - w|} \leq \sum_{n \geq 0} \sum_{j \in \mathcal{M}_n} \frac{1}{|y_j - x|} \leq 2N^\varepsilon \cdot \frac{3}{2q} + \sum_{n=1}^{9N} 2N^\varepsilon \cdot \frac{N}{n} \leq \frac{3N^\varepsilon}{q} + 12N^{1+\varepsilon} \log N \leq \mathcal{L}.$$

In the third inequality, we used the bound

$$\sum_{n=1}^{9N} \frac{1}{n} \leq 1 + \log(9N) \leq 6 \log N \quad (N \geq 2).$$

Hence the hypotheses of Lemma 3.8 on y and $\mathfrak{K}$ hold with the parameters (3.29).

We now verify (3.26). By (3.29) and (3.30), we have

$$\begin{aligned} w_0 \mathcal{L} &= 12N^\varepsilon w_0 (q^{-1} + N \log N) \leq \frac{1}{4}, & k_0 w_0 &= q \cdot N^\varepsilon w_0 q^{-1} \leq \frac{q}{48}, \\ 2k_0 w_0^2 \mathcal{L} &= 2(k_0 w_0)(w_0 \mathcal{L}) \leq \frac{q}{96}, \end{aligned}$$

and hence

$$k_0 w_0 + 2k_0 w_0^2 \mathcal{L} \leq \frac{q}{32} < \vartheta.$$

Therefore Lemma 3.8 applies. By (3.29), we have

$$2k_0 w_0^2 \mathcal{L} = 24N^{2\varepsilon} w_0^2 (q^{-1} + N \log N),$$

so the lemma gives (3.31) and the location of the eigenvalues within distance $q/32$ of $[\min_{j \in \mathcal{K}} y_j, \max_{j \in \mathcal{K}} y_j]$.

Finally, if i and $i+1$ belong to different clusters, then $y_{i+1} - y_i \geq q$, and the location of the eigenvalues gives

$$\lambda_{i+1}(D_y + W) - \lambda_i(D_y + W) \geq (y_{i+1} - q/32) - (y_i + q/32) \geq q - q/16 \geq q/2.$$

This completes the proof. $\square$

3.3 A Gaussian kernel estimate

We will use the following functions throughout the paper. For every integer $p \geq 1$, we define $f_p: \mathbb{R} \rightarrow (0, 1]$ by

$$f_p(y) = (1 + y^2)^{-p}, \quad y \in \mathbb{R}. \quad (3.33)$$

In Section 6, the gap of a 2×2 block of a matrix with a small Gaussian component is the norm of a Gaussian vector in $\mathbb{R}^2$ with nonzero mean. The following lemma computes the expectation of a function of this norm at a small scale δ , for functions bounded in

absolute value by f_p , uniformly in the mean and with an error that decays in the mean. The case $n = 3$ is used for Hermitian matrices in Section 9. For an integer $n \geq 1$ and $v > 0$, we let

$$\mathbf{g}_v(z) = (2\pi v)^{-n/2} \exp(-|z|^2/(2v)), \quad z \in \mathbb{R}^n, \quad (3.34)$$

denote the density of the centered Gaussian distribution on $\mathbb{R}^n$ with covariance $v\text{Id}_n$; the dimension n will be clear from the context.

Lemma 3.10. *Let $n \in \{2, 3\}$, and let $p > (n+1)/2$ be an integer. There exists $C > 0$, depending only on n and p , such that the following holds. Let $x \geq 0$, $\varrho \in [1/2, 2]$, $v, v_0 \in [1/8, 8]$, and $\delta \in (0, 1]$, and let G_x be a Gaussian vector in $\mathbb{R}^n$ with mean $(\varrho x, 0) \in \mathbb{R} \times \mathbb{R}^{n-1}$ and covariance $v\text{Id}_n$. Then for every Borel function $f: [0, \infty) \rightarrow \mathbb{R}$ such that $|f| \leq f_p$,*

$$\left| \delta^{-n} \mathbb{E}f(|G_x|/\delta) - \mathbf{g}_{v_0}(x, 0) \int_{\mathbb{R}^n} f(|y|) \, dy \right| \leq C f_p(x) (\delta + |\varrho - 1| + |v - v_0|). \quad (3.35)$$

In particular, $\mathbb{E}f_p(|G_x|/\delta) \leq C \delta^n f_p(x)$.

Proof. Throughout the proof, C denotes a constant that may depend on n and p , and we write $\hat{x} = (x, 0)$, with $0 \in \mathbb{R}^{n-1}$. We will use the bounds

$$\mathbf{g}_v(\varrho \hat{x}) \leq C e^{-x^2/64}, \quad |\nabla \mathbf{g}_v(z)| = \frac{|z|}{v} \mathbf{g}_v(z) \leq C e^{-|z|^2/32} \quad (z \in \mathbb{R}^n), \quad (3.36)$$

which hold for $\varrho \in [1/2, 2]$ and $v \in [1/8, 8]$, together with the fact that $e^{-cx^2} \leq C_c f_p(x)$ for every $c > 0$. As $2p - 1 > n$, we also have

$$\int_{\mathbb{R}^n} f_p(|y|) \, dy + \int_{\mathbb{R}^n} |y| f_p(|y|) \, dy \leq C. \quad (3.37)$$

(The assumption $p > (n+1)/2$, which is equivalent to $2p - n > 1$, is used here and again in the bound for the second term below.)

Recall that G_x has mean $\varrho \hat{x}$ and covariance $v\text{Id}_n$, so that its density is $z \mapsto \mathbf{g}_v(z - \varrho \hat{x})$. Changing variables $z = \delta y$ in the integral defining $\mathbb{E}f(|G_x|/\delta)$, we have

$$\delta^{-n} \mathbb{E}f(|G_x|/\delta) = \delta^{-n} \int_{\mathbb{R}^n} f(|z|/\delta) \mathbf{g}_v(z - \varrho \hat{x}) \, dz = \int_{\mathbb{R}^n} f(|y|) \mathbf{g}_v(\delta y - \varrho \hat{x}) \, dy.$$

Using $\mathbf{g}_v(-\varrho \hat{x}) = \mathbf{g}_v(\varrho \hat{x})$ and $|f| \leq f_p$, we find that the left side of (3.35) is bounded by

$$\begin{aligned} & \int_{|y| \leq \delta^{-1}} f_p(|y|) |\mathbf{g}_v(\delta y - \varrho \hat{x}) - \mathbf{g}_v(-\varrho \hat{x})| \, dy + \int_{|y| > \delta^{-1}} f_p(|y|) (\mathbf{g}_v(\delta y - \varrho \hat{x}) + \mathbf{g}_v(\varrho \hat{x})) \, dy \\ & + |\mathbf{g}_v(\varrho \hat{x}) - \mathbf{g}_{v_0}(\hat{x})| \int_{\mathbb{R}^n} f_p(|y|) \, dy. \end{aligned} \quad (3.38)$$

We bound the three terms in turn.

We begin with the first term. Let $|y| \leq \delta^{-1}$, and let $[-\varrho\hat{x}, \delta y - \varrho\hat{x}]$ denote the segment between these two points. Every point z' of this segment satisfies

$$|z'| \geq \max(\varrho x - 1, 0) \geq \max(x/2 - 1, 0), \quad \text{and hence} \quad |z'|^2 \geq \frac{x^2}{8} - 1,$$

where we used $(a - b)^2 \geq a^2/2 - b^2$. By the mean value theorem and (3.36),

$$|\mathfrak{g}_v(\delta y - \varrho\hat{x}) - \mathfrak{g}_v(-\varrho\hat{x})| \leq \delta|y| \sup_{z' \in [-\varrho\hat{x}, \delta y - \varrho\hat{x}]} |\nabla \mathfrak{g}_v(z')| \leq C\delta|y|e^{-x^2/256}.$$

Integrating against $f_p(|y|)$ and using (3.37), we find that the first term in (3.38) is bounded by

$$C\delta e^{-x^2/256} \int_{\mathbb{R}^n} |y| f_p(|y|) dy \leq C\delta e^{-x^2/256} \leq C\delta f_p(x).$$

The second term requires more care, since $\mathfrak{g}_v(\delta y - \varrho\hat{x})$ is of order one when δy is close to $\varrho\hat{x}$; there, the decay in x comes from the factor $f_p(|y|)$. The part with $\mathfrak{g}_v(\varrho\hat{x})$ is simpler. By (3.36) and $f_p(|y|) \leq |y|^{-2p}$, we have

$$\mathfrak{g}_v(\varrho\hat{x}) \int_{|y| > \delta^{-1}} f_p(|y|) dy \leq C e^{-x^2/64} \int_{|y| > \delta^{-1}} |y|^{-2p} dy \leq C e^{-x^2/64} \delta^{2p-n} \leq C\delta f_p(x),$$

where we used $2p - n > 1$ and $\delta \leq 1$ in the last step. For the part with $\mathfrak{g}_v(\delta y - \varrho\hat{x})$, we change variables back to $z = \delta y$ and use that

$$f_p(|z|/\delta) \leq (\delta/|z|)^{2p} \leq 2^p \delta^{2p} (1 + |z|^2)^{-p} \quad \text{for } |z| > 1.$$

This gives

$$\begin{aligned} \int_{|y| > \delta^{-1}} f_p(|y|) \mathfrak{g}_v(\delta y - \varrho\hat{x}) dy &= \delta^{-n} \int_{|z| > 1} f_p(|z|/\delta) \mathfrak{g}_v(z - \varrho\hat{x}) dz \\ &\leq 2^p \delta^{2p-n} \int_{\mathbb{R}^n} \frac{\mathfrak{g}_v(z - \varrho\hat{x})}{(1 + |z|^2)^p} dz. \end{aligned}$$

To bound the last integral uniformly in x , we use the inequality

$$1 + |a|^2 \leq 1 + (|a - b| + |b|)^2 \leq 2(1 + |a - b|^2)(1 + |b|^2), \quad a, b \in \mathbb{R}^n,$$

with $a = \varrho\hat{x}$ and $b = z$, together with $1 + \varrho^2 x^2 \geq (1 + x^2)/4$. We obtain

$$(1 + |z|^2)^{-p} \leq 2^p (1 + |z - \varrho\hat{x}|^2)^p (1 + \varrho^2 x^2)^{-p} \leq 8^p (1 + |z - \varrho\hat{x}|^2)^p f_p(x),$$

and therefore

$$\int_{\mathbb{R}^n} \frac{\mathfrak{g}_v(z - \varrho\hat{x})}{(1 + |z|^2)^p} dz \leq 8^p f_p(x) \int_{\mathbb{R}^n} (1 + |z'|^2)^p \mathfrak{g}_v(z') dz' \leq C f_p(x).$$

Using $\delta^{2p-n} \leq \delta$, we conclude that the second term in (3.38) is bounded by $C\delta f_p(x)$.

Next, we bound the third term. We consider

$$\mathfrak{g}_{v'}(\varrho' \hat{x}) = (2\pi v')^{-n/2} \exp\left(-\frac{\varrho'^2 x^2}{2v'}\right)$$

as a function of (ϱ', v') in the rectangle $[1/2, 2] \times [1/8, 8]$, with partial derivatives

$$\partial_{\varrho'} \mathfrak{g}_{v'}(\varrho' \hat{x}) = -\frac{\varrho' x^2}{v'} \mathfrak{g}_{v'}(\varrho' \hat{x}), \quad \partial_{v'} \mathfrak{g}_{v'}(\varrho' \hat{x}) = \left(-\frac{n}{2v'} + \frac{\varrho'^2 x^2}{2v'^2}\right) \mathfrak{g}_{v'}(\varrho' \hat{x}).$$

Using (3.36), we see that on this rectangle

$$|\partial_{\varrho'} \mathfrak{g}_{v'}(\varrho' \hat{x})| + |\partial_{v'} \mathfrak{g}_{v'}(\varrho' \hat{x})| \leq C(1 + x^2) e^{-x^2/64} \leq C f_p(x).$$

The rectangle is convex, so the mean value theorem along the segment from $(1, v_0)$ to (ϱ, v) gives

$$|\mathfrak{g}_v(\varrho \hat{x}) - \mathfrak{g}_{v_0}(\hat{x})| \leq C f_p(x) (|\varrho - 1| + |v - v_0|).$$

By (3.37), the third term in (3.38) is therefore bounded by $C f_p(x) (|\varrho - 1| + |v - v_0|)$.

Combining the bounds for the three terms in (3.38), we obtain (3.35). For the final claim, we apply (3.35) with $f = f_p$, and we obtain

$$\delta^{-n} \mathbb{E} f_p(|G_x|/\delta) \leq \mathfrak{g}_{v_0}(\hat{x}) \int_{\mathbb{R}^n} f_p(|y|) dy + C f_p(x) (\delta + |\varrho - 1| + |v - v_0|) \leq C f_p(x).$$

In the last step, we used (3.37), the bound $\mathfrak{g}_{v_0}(\hat{x}) \leq C f_p(x)$, which holds as $v_0 \in [1/8, 8]$, and $\delta + |\varrho - 1| + |v - v_0| \leq 10$. $\square$

Reduction to three propositions

In this section, we prove Theorem 1.2 assuming three propositions. The first, Proposition 4.1, counts pairs of eigenvalues at a scale slightly above the critical scale $N^{-3/2}$; we prove it in Section 6. The second, Proposition 4.4, is a local limit theorem for the linearized increments produced by resampling a small number of entries of H , which we prove in Section 8. The third, Proposition 4.5, states that the small gaps of the resampled matrix are close to the linearized gaps; its proof is given in Section 7. We state the propositions in Section 4.1 and Section 4.2, and we deduce Theorem 1.2 from them in Section 4.3.

Throughout the rest of the paper, we use the scales

$$d = N^{-3/2}, \quad \tau = N^{1/100}, \quad \ell = d\tau^{1/2} = N^{-299/200}, \quad s = N^{-7/5}. \quad (4.1)$$

The smallest gaps in Theorem 1.2 are of order d . Recall the scales $\ell_\alpha = N^{-1-\alpha}$ and $M_\alpha = N^{1-2\alpha}$ from (3.15). We set $\alpha_* = 99/200$, so that $\ell = \ell_{\alpha_*}$ and $M_{\alpha_*} = N^3\ell^2 = \tau$. Heuristically, a bulk interval contains of order τ pairs of eigenvalues at distance at most ℓ from each other. The scale s is the threshold below which two consecutive eigenvalues are treated together in Proposition 4.5. These scales satisfy

$$\frac{d}{\ell} = N^{-1/200}, \quad \frac{\ell}{s} = N^{-19/200}, \quad Ns = N^{-2/5}. \quad (4.2)$$

Recall that $g_i = \lambda_{i+1} - \lambda_i$ and $E_i = (\lambda_i + \lambda_{i+1})/2$ for $i \in [N-1]$.

4.1 Pair counts at the scale ℓ

Recall the functions $f_p(y) = (1+y^2)^{-p}$ from (3.33). We set $p_* = 61$. This value, like the lower bound $\alpha \geq 1/10$ in Proposition 6.7, is chosen for the bound (6.47) in the proof of Proposition 6.7, which requires $2p(\alpha' - \kappa) > 1 + 2\alpha'$ for all integers $p \geq p_*$, where $\kappa = \alpha - \alpha'/6$; this holds because $\alpha' - \kappa > \alpha/6 \geq 1/60$ and $\alpha' < 1/2$.

Proposition 4.1. *(i) For every continuous function $f: [0, \infty) \times (-2, 2) \rightarrow \mathbb{R}$ with compact support, we have*

$$\frac{1}{\tau} \sum_{1 \leq i < j \leq N} f\left(\frac{\lambda_j - \lambda_i}{\ell}, \frac{\lambda_i + \lambda_j}{2}\right) \longrightarrow \int_{-2}^2 \int_0^\infty c(E) r f(r, E) dr dE \quad (4.3)$$

in probability.

(ii) For every bulk interval I , there exists a constant $C_I > 0$ such that

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\sum_{\substack{1 \leq i < j \leq N \\ (\lambda_i + \lambda_j)/2 \in I}} f_{p_*} \left(\frac{\lambda_j - \lambda_i}{\ell} \right) \leq C_I \tau \right) = 1. \quad (4.4)$$

Compared to Theorem 1.3 with $\alpha = \alpha_*$, the sum in the first part runs over all pairs $i < j$ instead of consecutive eigenvalues, and the test function may depend on the energy. The second part controls pairs at all distances, with a weight that decays only polynomially in $(\lambda_j - \lambda_i)/\ell$. We will use it to bound the number of pairs with $g_i \leq R\ell$ by $O_R(\tau)$, and to control the pairs with $g_i \gg \ell$, which the first part does not see.

4.2 The resampling and the linearized increments

The resampling. Set

$$m = m_N = 2 \lfloor \tau/8 \rfloor, \quad \text{so that} \quad \lim_{N \rightarrow \infty} \frac{4m}{\tau} = 1. \quad (4.5)$$

Let $(a_1, b_1), \dots, (a_m, b_m)$ be independent random pairs, each uniformly distributed on the set $\{(a, b) \in [N]^2 : a < b\}$. Further, let $Y_1, \dots, Y_m$ be independent copies of X . We assume that H , the pairs (a_l, b_l) , and the variables Y_l are mutually independent. We call (a_l, b_l) the l -th *selected position*, and we call $(a_l, b_l, Y_l)_{l \in [m]}$ the *resampling variables*. Define real symmetric matrices $H^{(0)}, \dots, H^{(m)}$ by $H^{(0)} = H$ and

$$H^{(l)} = H^{(l-1)} + (N^{-1/2} Y_l - H_{a_l b_l}^{(l-1)}) V^{a_l b_l}, \quad l \in [m], \quad (4.6)$$

where V^{ab} is as in (3.1). Thus $H^{(l)}$ is obtained from $H^{(l-1)}$ by setting its (a_l, b_l) and (b_l, a_l) entries equal to $N^{-1/2} Y_l$. We set

$$\tilde{H} = H^{(m)}, \quad \Delta = \tilde{H} - H, \quad (4.7)$$

and we call $\tilde{H}$ the *resampled matrix*. We also write $\tilde{\lambda}_i = \lambda_i(\tilde{H})$ for $i \in [N]$, and $\tilde{g}_i = \tilde{\lambda}_{i+1} - \tilde{\lambda}_i$ and $\tilde{E}_i = (\tilde{\lambda}_i + \tilde{\lambda}_{i+1})/2$ for $i \in [N-1]$. Since $a_l < b_l$, the resampling leaves the diagonal entries of H unchanged. A position may be selected more than once; in that case, the last replacement determines the final entry in that position. (The fact that m is even is used in Section 8.)

Lemma 4.2. *The resampled matrix $\tilde{H}$ has the same law as H .*

Proof. Let $\sigma = ((a'_l, b'_l))_{l \in [m]}$ be a deterministic sequence of pairs of integers with $1 \leq a'_l < b'_l \leq N$, and define $H_\sigma^{(l)}$ by (4.6) with (a'_l, b'_l) in place of (a_l, b_l) and $H_\sigma^{(0)} = H$. We show by induction on l that $H_\sigma^{(l)}$ has the law of H , the case $l = 0$ being trivial.

Suppose that this holds for $l - 1$. Then the entries of $H_\sigma^{(l-1)}$ on and above the diagonal are independent with the laws of the corresponding entries of H , and they are independent of Y_l , since $H_\sigma^{(l-1)}$ is a function of H and $Y_1, \dots, Y_{l-1}$. The matrix $H_\sigma^{(l)}$ is obtained by replacing the entry at (a'_l, b'_l) by $N^{-1/2}Y_l$, which has the law of $H_{a'_l b'_l}$, so the claim holds for l .

The selected positions are independent of H and of $(Y_l)_{l \in [m]}$, and $\tilde{H} = H_\sigma^{(m)}$ on the event that $(a_l, b_l) = (a'_l, b'_l)$ for all $l \in [m]$. Therefore, for every Borel set $\mathfrak{B}$ of real symmetric $N \times N$ matrices, we have

$$\mathbb{P}(\tilde{H} \in \mathfrak{B}) = \sum_{\sigma} \mathbb{P}((a_l, b_l)_{l \in [m]} = \sigma) \mathbb{P}(H_\sigma^{(m)} \in \mathfrak{B}) = \mathbb{P}(H \in \mathfrak{B}),$$

where the sum runs over all sequences σ as above, and the second equality follows from the claim. $\square$

Conditioning on H . Recall that the eigenvectors $u_1, \dots, u_N$ are measurable functions of H . Since the resampling variables are independent of H , Fubini's theorem shows that integrating over the resampling variables with H held fixed gives a version of the conditional expectation given H , and we always use this version. An event $\mathcal{A}$ that is measurable with respect to H has the form $\{H \in \mathfrak{A}\}$ for a Borel set $\mathfrak{A}$ of real symmetric matrices. A statement such as “on $\mathcal{A}$, we have $\mathbb{P}(\mathcal{B} \mid H) \geq 1 - o(1)$ ” means that, for every $A \in \mathfrak{A}$, the probability of $\mathcal{B}$ computed with H replaced by A is bounded below by $1 - o(1)$. In Proposition 4.4 and Proposition 4.5 below, the error terms are deterministic quantities in the sense of our conventions. They may depend on fixed parameters, such as k and the test functions, but not on A , on the resampling variables, or on the eigenvalue labels involved.

Linearized increments. For $i \in [N - 1]$ and a real symmetric $N \times N$ matrix B , define

$$L_i(B) = \left(\frac{u_i^T B u_i - u_{i+1}^T B u_{i+1}}{2}, u_i^T B u_{i+1} \right) \in \mathbb{R}^2. \quad (4.8)$$

The map L_i is linear, and $L_i(H) = ((\lambda_i - \lambda_{i+1})/2, 0)$. We set

$$z_i = d^{-1} L_i(H) = \left(-\frac{g_i}{2d}, 0 \right), \quad \text{so that} \quad |z_i| = \frac{g_i}{2d}. \quad (4.9)$$

The map L_i is relevant because of an elementary identity for 2×2 matrices. For $y_1, y_2, y_3 \in \mathbb{R}$, the formula for the roots of the characteristic polynomial gives

$$\lambda_2 \begin{pmatrix} y_1 & y_3 \\ y_3 & y_2 \end{pmatrix} - \lambda_1 \begin{pmatrix} y_1 & y_3 \\ y_3 & y_2 \end{pmatrix} = ((y_1 - y_2)^2 + 4y_3^2)^{1/2} = 2 \left| \left(\frac{y_1 - y_2}{2}, y_3 \right) \right|. \quad (4.10)$$

Applying (4.10) to the compression of $H + B$ to the span of u_i and u_{i+1} , we find that the 2×2 matrix $(u_j^T (H + B) u_{j'})_{j, j' \in \{i, i+1\}}$ has eigenvalue gap $2|L_i(H + B)|$.

For $1 \leq a < b \leq N$ and $i \in [N-1]$, define

$$D_{ab}^{(i)} = NL_i(V^{ab}) = N(u_i(a)u_i(b) - u_{i+1}(a)u_{i+1}(b), u_i(a)u_{i+1}(b) + u_i(b)u_{i+1}(a)) \in \mathbb{R}^2, \quad (4.11)$$

where the second equality follows from (3.1). For $i \in [N-1]$ and $l \in [m]$, define

$$J_{i,l} = (Y_l - X_{a_l b_l})D_{a_l b_l}^{(i)}, \quad S_i = \sum_{l=1}^m J_{i,l}. \quad (4.12)$$

We call S_i the *linearized increment of the pair* $(i, i+1)$. Note that $X_{a_l b_l} = N^{1/2}H_{a_l b_l}$ is the entry of the original matrix H , also when the position (a_l, b_l) was selected before the l -th step. Conditionally on H , the triples (a_l, b_l, Y_l) , $l \in [m]$, are independent and identically distributed, so S_i is a sum of m conditionally independent and identically distributed random vectors.

For several pairs, we use the following notation. For an integer $k \geq 1$ and $\mathbf{i} = (i_1, \dots, i_k) \in [N-1]^k$, let

$$\mathbf{z}_{\mathbf{i}} = (z_{i_1}, \dots, z_{i_k}), \quad D_{ab}^{(\mathbf{i})} = (D_{ab}^{(i_1)}, \dots, D_{ab}^{(i_k)}), \quad S_{\mathbf{i}} = (S_{i_1}, \dots, S_{i_k}), \quad (4.13)$$

which are vectors in $\mathbb{R}^{2k}$. Then $S_{\mathbf{i}} = \sum_{l=1}^m (Y_l - X_{a_l b_l})D_{a_l b_l}^{(\mathbf{i})}$. The vectors $S_{i_1}, \dots, S_{i_k}$ are built from the same resampling variables, so they are not conditionally independent given H . The next lemma explains the name of S_i .

Lemma 4.3. *On the event that the selected positions $(a_1, b_1), \dots, (a_m, b_m)$ are distinct, we have*

$$\Delta = N^{-1/2} \sum_{l=1}^m (Y_l - X_{a_l b_l})V^{a_l b_l} \quad \text{and} \quad d^{-1}L_i(\Delta) = S_i \quad \text{for all } i \in [N-1]. \quad (4.14)$$

In particular, on this event, the 2×2 matrix $(u_j^T \tilde{H} u_{j'})_{j, j' \in \{i, i+1\}}$ has eigenvalue gap $2d|z_i + S_i|$ for every $i \in [N-1]$.

Proof. Suppose that the selected positions are distinct. Then for every $l \in [m]$, the entry of $H^{(l-1)}$ in position (a_l, b_l) is not changed in the first $l-1$ steps. Using (4.6), we deduce that

$$H_{a_l b_l}^{(l-1)} = H_{a_l b_l} = N^{-1/2}X_{a_l b_l}, \quad H^{(l)} - H^{(l-1)} = N^{-1/2}(Y_l - X_{a_l b_l})V^{a_l b_l}.$$

Summing over $l \in [m]$ gives the first identity in (4.14). Using the linearity of L_i , (4.11), and (4.9), we then obtain

$$L_i(\Delta) = N^{-1/2} \sum_{l=1}^m (Y_l - X_{a_l b_l})L_i(V^{a_l b_l}) = N^{-3/2} \sum_{l=1}^m J_{i,l} = dS_i, \\ L_i(\tilde{H}) = L_i(H) + L_i(\Delta) = d(z_i + S_i).$$

The last claim now follows from the consequence of (4.10) stated above, applied with $B = \Delta$. $\square$

We call $2d|z_i + S_i|$ the *linearized gap* of the pair $(i, i + 1)$. It is the gap that first-order perturbation theory predicts for the pair $(i, i + 1)$ when the interaction with the other eigenvalues is neglected.

A local limit theorem. We now state the local limit theorem for the linearized increments. The heuristic behind its normalization is explained in Section 2.2, and the conditional covariance of the summands of S_i , which is close to 4Id_{2k} , is computed in Lemma 8.4. The proposition resolves the conditional law of S_i on the unit scale, which is much smaller than the scale $\tau^{1/2}$ of its fluctuations.

Proposition 4.4. *Fix an integer $k \geq 1$. There exist events $\mathcal{A}_{N,k}$, measurable with respect to H , and a constant $c > 0$ such that $\lim_{N \rightarrow \infty} \mathbb{P}(\mathcal{A}_{N,k}) = 1$ and the following statements hold.*

- (i) *Let $\Psi \subset C_c^\infty(\mathbb{R}^{2k})$ be a family of functions whose supports are contained in a common compact set, and such that every partial derivative of every order is bounded uniformly over Ψ . Then there exist numbers $\epsilon_N = \epsilon_N(k, \Psi) \geq 0$ such that $\lim_{N \rightarrow \infty} \epsilon_N = 0$ and, on $\mathcal{A}_{N,k}$,*

$$\sup_{\mathbf{i}, x, \psi} \left| \tau^k \mathbb{E}[\psi(S_{\mathbf{i}} + x) \mid H] - \frac{e^{-|x|^2/(2\tau)}}{(2\pi)^k} \int_{\mathbb{R}^{2k}} \psi(y) \, dy \right| \leq \epsilon_N, \quad (4.15)$$

where the supremum is over all $\psi \in \Psi$, all $x \in \mathbb{R}^{2k}$, and all $\mathbf{i} = (i_1, \dots, i_k) \in [N-1]^k$ such that $i_{j+1} \geq i_j + 2$ for all $j \in [k-1]$.

- (ii) *Suppose that $k = 1$. For every $x_0 > 0$, there exists $C_{x_0} > 0$ such that on $\mathcal{A}_{N,1}$, we have*

$$\mathbb{P}(2|z_i + S_i| \leq x_0 \mid H) \leq C_{x_0} \tau^{-1} e^{-cg_i/\ell} \quad (4.16)$$

for all $i \in [N-1]$ such that $g_i \geq \ell$.

The uniformity in $\mathbf{i}$ and x in the first part is essential, since in Section 4.3 we apply Proposition 4.4 with $x = z_i$ to tuples $\mathbf{i}$ of pairs with small gaps, which are selected according to H . The second part shows that a pair with $g_i \gg \ell$, for which $|z_i| \gg \tau^{1/2}$, has a linearized gap of order d with probability much smaller than τ^{-1} .

Linearization of the small gaps. Finally, we compare the gaps of $\tilde{H}$ with the linearized gaps. By Proposition 3.4 and a union bound over i , the event that $\lambda_{i+2} - \lambda_i > 4s$ for all $i \in [N-2]$ has probability at least $1 - CN(Ns)^3 = 1 - CN^{-1/5}$. On this event, every pair $(i, i + 1)$ with $g_i \leq s$ is separated from the other eigenvalues by more than $3s$. To state the error terms in the comparison, we define

$$w_N = \tau N^{-3/2+3/100}, \quad \mathcal{L}_N = 2s^{-1} \log N, \quad \delta_N = 8w_N^2 \mathcal{L}_N. \quad (4.17)$$

These quantities are also used in Section 7, where w_N bounds the entries of Δ in the eigenbasis of H , $\mathcal{L}_N$ bounds the sums in (3.25), and $\delta_N/2$ is the resulting error in (3.27). By (4.1), we have

$$w_N = N^{-73/50}, \quad \frac{\delta_N}{d} = 16N^{-1/50} \log N, \quad (4.18)$$

so w_N and δ_N/d tend to zero.

Proposition 4.5. *There exist events $\mathcal{A}_N$, measurable with respect to H , and events $\mathcal{B}_N$, measurable with respect to H and the resampling variables, such that $\lim_{N \rightarrow \infty} \mathbb{P}(\mathcal{A}_N) = 1$ and the following statements hold for all sufficiently large N .*

- (i) *On $\mathcal{A}_N$, we have $\lambda_{i+2} - \lambda_i > 4s$ for all $i \in [N-2]$, and $\mathbb{P}(\mathcal{B}_N \mid H) \geq 1 - o(1)$.*
- (ii) *On $\mathcal{A}_N \cap \mathcal{B}_N$, we have for all $i \in [N-1]$ that*

$$|d^{-1}\tilde{g}_i - 2|z_i + S_i|| \leq \frac{\delta_N}{d} \quad \text{and} \quad |\tilde{E}_i - E_i| \leq w_N + \frac{\delta_N}{2} \quad \text{if } g_i \leq s, \quad (4.19)$$

$$\tilde{g}_i > s/2 \quad \text{if } g_i > s. \quad (4.20)$$

Thus every gap of $\tilde{H}$ of order d is, up to an error $\delta_N = o(d)$, the linearized gap of a pair with $g_i \leq s$. We prove Proposition 4.5 in Section 7 by applying Lemma 3.8 in the eigenbasis of H . Note that $\mathbb{P}(\mathcal{B}_N^c \mid H)$ is only $o(1)$ on $\mathcal{A}_N$, so in the proof of Theorem 1.2 we use it only to bound differences of bounded functionals, and we never multiply it by a number of pairs.

4.3 Proof of the main theorem

The heuristic behind the proof is explained in Section 2.2, and we proceed as follows. We introduce a point process $\hat{\Xi}_N$ of rescaled linearized gaps and its truncation $\hat{\Xi}_{N,R}$ to pairs with $g_i \leq R\ell$. In Lemma 4.9, which is the main step, we compute the limit of the Laplace functional of $\hat{\Xi}_{N,R}$ for smooth test functions, using the Bonferroni inequalities and Proposition 4.4. The truncation is needed because the errors in Proposition 4.4 are summed over j -tuples of pairs, and Proposition 4.1 bounds the number of pairs with $g_i \leq R\ell$ by $O_R(\tau)$. In Lemma 4.10 and Lemma 4.11, we extend this limit to continuous test functions and remove the truncation using the second part of Proposition 4.4. In the proof of Theorem 1.2, we pass to the gaps of $\tilde{H}$ using Proposition 4.5 and to H using Lemma 4.2, and we deduce the law of the smallest gap from the void probabilities of the Poisson limit, which are the subject of Lemma 4.12.

Setup. Let C_c^+ denote the set of continuous functions $f: [0, \infty) \times (-2, 2) \rightarrow [0, \infty)$ with compact support. For $f \in C_c^+$, set

$$\mathsf{L}(f) = \int (1 - e^{-f}) \, d\nu = \int_{-2}^2 \int_0^\infty c(E)r(1 - e^{-f(r,E)}) \, dr \, dE. \quad (4.21)$$

For $R > 0$, define the measure $\nu_R = (1 - e^{-R^2/8})\nu$ and set

$$\mathsf{L}_R(f) = \int (1 - e^{-f}) \, d\nu_R = (1 - e^{-R^2/8})\mathsf{L}(f). \quad (4.22)$$

Let Ξ be a Poisson point process on $[0, \infty) \times (-2, 2)$ with intensity ν . Then $\mathbb{E}e^{-\Xi(f)} = e^{-\mathcal{L}(f)}$ for all $f \in C_c^+$; see [42, Chapter 3].

Until the proof of Theorem 1.2, I denotes an arbitrary bulk interval; in that proof, it will be chosen according to the test function. Let I' be the closed interval with the same midpoint as I and with length

$$|I'| = |I| + \text{dist}(I, \{-2, 2\}), \quad (4.23)$$

where $|I|$ denotes the length of I . Then I' is a bulk interval and $I \subset \text{int}(I')$, where int denotes the interior. Define the sets of eigenvalue labels

$$\mathcal{C} = \{i \in [N-1] : g_i \leq s, E_i \in I'\}, \quad \mathcal{C}_R = \{i \in \mathcal{C} : g_i \leq R\ell\} \quad (R > 0), \quad (4.24)$$

and the point processes

$$\widehat{\Xi}_N = \sum_{i \in \mathcal{C}} \delta_{(2|z_i + S_i|, E_i)}, \quad \widehat{\Xi}_{N,R} = \sum_{i \in \mathcal{C}_R} \delta_{(2|z_i + S_i|, E_i)} \quad (4.25)$$

on $[0, \infty) \times (-2, 2)$.

Let $\mathcal{A}_N$ and $\mathcal{B}_N$ be the events of Proposition 4.5, and for every integer $k \geq 1$, let $\mathcal{A}_{N,k}$ be the events of Proposition 4.4. Let $C_{I'}$ be the constant of the second part of Proposition 4.1 for the interval I' . Define

$$\mathcal{A}_N^* = \mathcal{A}_N \cap \mathcal{A}_{N,1} \cap \left\{ \sum_{\substack{1 \leq i < j \leq N \\ (\lambda_i + \lambda_j)/2 \in I'}} f_{p_*} \left(\frac{\lambda_j - \lambda_i}{\ell} \right) \leq C_{I'} \tau \right\}. \quad (4.26)$$

All objects introduced in this paragraph, except $\widehat{\Xi}_N$, $\widehat{\Xi}_{N,R}$, and $\mathcal{B}_N$, are measurable with respect to H .

Lemma 4.6. *We have $\lim_{N \rightarrow \infty} \mathbb{P}(\mathcal{A}_N^*) = 1$. Further, for every $R' > 0$, there exists $N_0 > 0$, depending on R' , such that for $N \geq N_0$ the following statements hold on $\mathcal{A}_N^*$.*

- (i) *Distinct elements of $\mathcal{C}$ differ by at least 2.*
- (ii) *We have $\sum_{i \in \mathcal{C}} f_{p_*}(g_i/\ell) \leq C_{I'} \tau$, and $|\mathcal{C}_R| \leq C_{I'}(1 + R^2)^{p_*} \tau$ for all $R > 0$.*
- (iii) *For every function $Q: [0, \infty) \times (-2, 2) \rightarrow \mathbb{R}$ such that $Q(y, E) = 0$ for $y \geq R'$, we have*

$$\sum_{1 \leq i < j \leq N} Q\left(\frac{\lambda_j - \lambda_i}{\ell}, \frac{\lambda_i + \lambda_j}{2}\right) = \sum_{i=1}^{N-1} Q\left(\frac{g_i}{\ell}, E_i\right). \quad (4.27)$$

Proof. The first claim follows from Proposition 4.5, Proposition 4.4, the second part of Proposition 4.1, and a union bound. Now suppose that $\mathcal{A}_N^*$ holds. The first part of Proposition 4.5 gives

$$\lambda_{i+2} - \lambda_i > 4s \quad \text{for all } i \in [N-2]. \quad (4.28)$$

To prove (i), suppose that $i, i+1 \in \mathcal{C}$ for some i . Then

$$\lambda_{i+2} - \lambda_i = g_i + g_{i+1} \leq 2s,$$

which contradicts (4.28).

To prove (ii), note that for every $i \in \mathcal{C}$, the pair $(i, i+1)$ appears in the sum in (4.26), since its midpoint is $E_i \in I'$. All terms in this sum are nonnegative, so

$$\sum_{i \in \mathcal{C}} f_{p_*} \left(\frac{g_i}{\ell} \right) \leq C_{I'} \tau.$$

The bound on $|\mathcal{C}_R|$ follows, because $f_{p_*}(g_i/\ell) \geq (1 + R^2)^{-p_*}$ for $i \in \mathcal{C}_R$.

Finally, to prove (iii), observe that for $j \geq i+2$,

$$\lambda_j - \lambda_i \geq \lambda_{i+2} - \lambda_i > 4s \geq R' \ell,$$

where we used (4.28) in the second inequality, and the last inequality holds for large N by (4.2). Hence the terms with $j \geq i+2$ on the left side of (4.27) vanish, and the remaining terms, with $j = i+1$, are the terms on the right side. $\square$

The Bonferroni inequalities. For a finite set $\mathcal{T}$, real numbers $y = (y_i)_{i \in \mathcal{T}}$, and an integer $j \geq 0$, let

$$P_j(y) = \sum_{\substack{\mathcal{T}' \subset \mathcal{T} \\ |\mathcal{T}'|=j}} \prod_{i \in \mathcal{T}'} y_i \quad (4.29)$$

denote the j -th elementary symmetric polynomial of y . In particular, $P_0(y) = 1$, and $P_j(y) = 0$ for $j > |\mathcal{T}|$.

Lemma 4.7. *Let $\mathcal{T}$ be a finite set, and let $\mathbf{w} = (w_i)_{i \in \mathcal{T}} \in [0, 1]^{\mathcal{T}}$. Then for every integer $n \geq 0$,*

$$\sum_{j=0}^{2n+1} (-1)^j P_j(\mathbf{w}) \leq \prod_{i \in \mathcal{T}} (1 - w_i) \leq \sum_{j=0}^{2n} (-1)^j P_j(\mathbf{w}). \quad (4.30)$$

Proof. Let $(b_i)_{i \in \mathcal{T}}$ be independent random variables such that $\mathbb{P}(b_i = 1) = w_i$ and $\mathbb{P}(b_i = 0) = 1 - w_i$. We set

$$M = \sum_{i \in \mathcal{T}} b_i.$$

Expanding the products and using independence, we find

$$P_j(\mathbf{w}) = \mathbb{E} \sum_{\substack{\mathcal{T}' \subset \mathcal{T} \\ |\mathcal{T}'|=j}} \prod_{i \in \mathcal{T}'} b_i = \mathbb{E} \binom{M}{j}, \quad \prod_{i \in \mathcal{T}} (1 - w_i) = \mathbb{E} \prod_{i \in \mathcal{T}} (1 - b_i) = \mathbb{P}(M = 0).$$

For integers $M_0 \geq 1$ and $n_0 \geq 0$, Pascal's rule gives the telescoping identity

$$\sum_{j=0}^{n_0} (-1)^j \binom{M_0}{j} = (-1)^{n_0} \binom{M_0-1}{n_0},$$

whose right side is nonnegative for even n_0 and nonpositive for odd n_0 , while for $M_0 = 0$ the left side equals 1. Applying this with $M_0 = M$ and $n_0 \in \{2n, 2n+1\}$ and taking expectations, we obtain (4.30). (This is the classical Bonferroni inequality for the events $\{b_i = 1\}$.) $\square$

Counts of small linearized gaps.

Lemma 4.8. *Let $x_0 > 0$. There exist constants $C, c > 0$ and $N_0 > 0$, depending on I and x_0 , such that for $N \geq N_0$, on $\mathcal{A}_N^*$, we have*

$$\sum_{\substack{i \in \mathcal{C} \\ g_i \geq R\ell}} \mathbb{P}(2|z_i + S_i| \leq x_0 \mid H) \leq C \sup_{y \geq R} ((1+y^2)^{p_*} e^{-cy}) \quad \text{for all } R \geq 0. \quad (4.31)$$

In particular, for $N \geq N_0$, on $\mathcal{A}_N^*$, we have

$$\mathbb{E}[\widehat{\Xi}_N([0, x_0] \times I') \mid H] \leq C', \quad (4.32)$$

$$\mathbb{E}[(\widehat{\Xi}_N - \widehat{\Xi}_{N,R})([0, x_0] \times I') \mid H] \leq C \sup_{y \geq R} ((1+y^2)^{p_*} e^{-cy}) \quad \text{for all } R > 0, \quad (4.33)$$

where $C' = C \sup_{y \geq 0} ((1+y^2)^{p_*} e^{-cy})$.

Proof. Let $c > 0$ be the constant of Proposition 4.4 for $k = 1$. Fix $\psi_0 \in C_c^\infty(\mathbb{R}^2)$ such that $0 \leq \psi_0 \leq 1$ and $\psi_0(y) = 1$ for $|y| \leq x_0/2$. Suppose that N is large and that $\mathcal{A}_N^*$ holds. Let $i \in \mathcal{C}$. Applying the first part of Proposition 4.4 with $k = 1$, the family $\Psi = \{\psi_0\}$, and $x = z_i$ on the event $\mathcal{A}_{N,1} \supset \mathcal{A}_N^*$, we obtain

$$\mathbb{P}(2|z_i + S_i| \leq x_0 \mid H) \leq \mathbb{E}[\psi_0(S_i + z_i) \mid H] \leq \frac{1}{\tau} \left(\frac{1}{2\pi} \int_{\mathbb{R}^2} \psi_0(y) dy + \epsilon_N \right) \leq \frac{C}{\tau},$$

where $\epsilon_N = \epsilon_N(1, \{\psi_0\})$ is the sequence of Proposition 4.4, which is bounded.

For $g_i < \ell$, we have $e^{-cg_i/\ell} \geq e^{-c}$, while for $g_i \geq \ell$, the second part of Proposition 4.4 applies. After enlarging C , we therefore find that

$$\mathbb{P}(2|z_i + S_i| \leq x_0 \mid H) \leq C\tau^{-1}e^{-cg_i/\ell} \quad \text{for all } i \in \mathcal{C}.$$

The sum over the pairs with $g_i \geq R\ell$ can then be bounded using the second part of Lemma 4.6. For $R \geq 0$, we obtain

$$\begin{aligned} \sum_{\substack{i \in \mathcal{C} \\ g_i \geq R\ell}} \mathbb{P}(2|z_i + S_i| \leq x_0 \mid H) &\leq \frac{C}{\tau} \sup_{y \geq R} ((1+y^2)^{p_*} e^{-cy}) \sum_{i \in \mathcal{C}} f_{p_*}\left(\frac{g_i}{\ell}\right) \\ &\leq CC_{I'} \sup_{y \geq R} ((1+y^2)^{p_*} e^{-cy}), \end{aligned}$$

which is (4.31).

Observe that every point of $\widehat{\Xi}_N$ has its second coordinate in I' . Using the fact that $\mathcal{C}$ is measurable with respect to H , we therefore find

$$\begin{aligned}\mathbb{E}[\widehat{\Xi}_N([0, x_0] \times I') \mid H] &= \sum_{i \in \mathcal{C}} \mathbb{P}(2|z_i + S_i| \leq x_0 \mid H), \\ \mathbb{E}[(\widehat{\Xi}_N - \widehat{\Xi}_{N,R})([0, x_0] \times I') \mid H] &= \sum_{\substack{i \in \mathcal{C} \\ g_i > R\ell}} \mathbb{P}(2|z_i + S_i| \leq x_0 \mid H).\end{aligned}$$

Hence (4.31) implies (4.32) and (4.33). $\square$

Smooth test functions. The local limit theorem applies only to smooth test functions, but the function $y \mapsto f(2|y|, E)$ on $\mathbb{R}^2$ is in general not smooth at $y = 0$, even for smooth f . For this reason, we first consider test functions of the form $f(r, E) = h(r^2, E)$ with h smooth.

Lemma 4.9. *Let $R > 0$, and let $h \in C_c^\infty(\mathbb{R}^2)$ be a nonnegative function with $\text{supp } h \subset \mathbb{R} \times \text{int}(I')$. Define $f(r, E) = h(r^2, E)$ for $(r, E) \in [0, \infty) \times (-2, 2)$. Then $f \in C_c^+$, and*

$$\mathbb{E}[e^{-\widehat{\Xi}_{N,R}(f)} \mid H] \longrightarrow e^{-L_R(f)} \quad (4.34)$$

in probability.

Proof. Fix $u_h > 0$ and a compact interval $I_h \subset \text{int}(I')$ such that $\text{supp } h \subset [-u_h, u_h] \times I_h$. Then f is continuous and nonnegative, and $\text{supp } f \subset [0, u_h^{1/2}] \times I_h$, so $f \in C_c^+$.

We begin by introducing planar test functions. For $E \in (-2, 2)$ and $y \in \mathbb{R}^2$, define

$$\psi_E(y) = 1 - \exp(-h(4|y|^2, E)), \quad A_h(E) = \int_{\mathbb{R}^2} \psi_E(y) \, dy. \quad (4.35)$$

Since $f(2|y|, E) = h(4|y|^2, E)$, we have

$$1 - e^{-f(2|y|, E)} = \psi_E(y) \quad \text{for all } y \in \mathbb{R}^2 \text{ and } E \in (-2, 2). \quad (4.36)$$

We record three properties of ψ_E and A_h . First, $0 \leq \psi_E \leq 1$, and $\psi_E(y) = 0$ if $|y| > u_h^{1/2}/2$ or $E \notin I_h$. Second, $\psi_E \in C_c^\infty(\mathbb{R}^2)$, and the chain rule shows that every partial derivative of ψ_E is bounded uniformly in $E \in (-2, 2)$. Third, A_h is continuous by dominated convergence, it vanishes outside I_h , and

$$A_h(E) = \frac{\pi}{2} \int_0^\infty r(1 - e^{-f(r, E)}) \, dr, \quad (4.37)$$

which follows from (4.36) by using polar coordinates in $\mathbb{R}^2$ with radius $|y| = r/2$.

We now apply the Bonferroni inequalities. For $i \in \mathcal{C}_R$, define

$$w_i = \psi_{E_i}(z_i + S_i) \in [0, 1], \quad (4.38)$$

and let $w = (w_i)_{i \in \mathcal{C}_R}$. Using (4.36), we find

$$e^{-\hat{\Xi}_{N,R}(f)} = \prod_{i \in \mathcal{C}_R} e^{-f(2|z_i + S_i|, E_i)} = \prod_{i \in \mathcal{C}_R} (1 - w_i).$$

Since $\mathcal{C}_R$ is measurable with respect to H , Lemma 4.7 with $\mathcal{T} = \mathcal{C}_R$ implies that for every integer $n \geq 0$,

$$\sum_{j=0}^{2n+1} (-1)^j \mathbb{E}[P_j(w) \mid H] \leq \mathbb{E}[e^{-\hat{\Xi}_{N,R}(f)} \mid H] \leq \sum_{j=0}^{2n} (-1)^j \mathbb{E}[P_j(w) \mid H]. \quad (4.39)$$

We claim that for every integer $j \geq 1$,

$$\mathbb{E}[P_j(w) \mid H] \longrightarrow \frac{L_R(f)^j}{j!} \quad \text{in probability.} \quad (4.40)$$

We first show that (4.40) implies the lemma. Let $\omega > 0$. The alternating series $\sum_{j \geq 0} (-1)^j L_R(f)^j / j!$ converges to $e^{-L_R(f)}$, so there exists $n \geq 0$ such that

$$\left| \sum_{j=0}^{2n} \frac{(-1)^j L_R(f)^j}{j!} - e^{-L_R(f)} \right| \leq \omega, \quad \left| \sum_{j=0}^{2n+1} \frac{(-1)^j L_R(f)^j}{j!} - e^{-L_R(f)} \right| \leq \omega.$$

Then (4.40) implies that the event

$$\left| \mathbb{E}[P_j(w) \mid H] - \frac{L_R(f)^j}{j!} \right| \leq \frac{\omega}{2n+2} \quad \text{for all } j \in [2n+1]$$

has probability tending to one. On this event, (4.39) and the choice of n give

$$\left| \mathbb{E}[e^{-\hat{\Xi}_{N,R}(f)} \mid H] - e^{-L_R(f)} \right| \leq 2\omega.$$

The number $\omega > 0$ was arbitrary, so (4.34) follows.

It remains to prove (4.40). Fix $j \geq 1$. For $i \in \mathcal{C}_R$, define

$$\bar{w}_i = \frac{1}{2\pi\tau} \exp\left(-\frac{g_i^2}{8\ell^2}\right) A_h(E_i), \quad (4.41)$$

and let $\bar{w} = (\bar{w}_i)_{i \in \mathcal{C}_R}$. We compare $\mathbb{E}[P_j(w) \mid H]$ with $P_j(\bar{w})$, then $P_j(\bar{w})$ with $(\sum_{i \in \mathcal{C}_R} \bar{w}_i)^j / j!$, and finally we show that $\sum_{i \in \mathcal{C}_R} \bar{w}_i$ converges to $L_R(f)$ in probability.

For the first comparison, let Ψ_j be the family of all functions of the form $\psi_{E'_1} \otimes \cdots \otimes \psi_{E'_j}$ with $E'_1, \dots, E'_j \in (-2, 2)$, where

$$(\psi_{E'_1} \otimes \cdots \otimes \psi_{E'_j})(y_1, \dots, y_j) = \psi_{E'_1}(y_1) \cdots \psi_{E'_j}(y_j), \quad y_1, \dots, y_j \in \mathbb{R}^2.$$

By the first two properties of ψ_E , the members of Ψ_j are supported in a product of j disks of radius $u_h^{1/2}/2$, and each of their partial derivatives is bounded uniformly over Ψ_j . Thus Ψ_j satisfies the hypotheses of the first part of Proposition 4.4 with $k = j$. Let N be large enough that Lemma 4.6 applies with $R' = R + 1$, and suppose that $\mathcal{A}_N^* \cap \mathcal{A}_{N,j}$ holds.

Let $\mathcal{T}' \subset \mathcal{C}_R$ with $|\mathcal{T}'| = j$. We list the elements of $\mathcal{T}'$ in increasing order as $\mathbf{i} = (i_1, \dots, i_j)$. The first part of Lemma 4.6 shows that consecutive entries of $\mathbf{i}$ differ by at least 2. We set

$$\psi = \psi_{E_{i_1}} \otimes \cdots \otimes \psi_{E_{i_j}} \in \Psi_j.$$

Then we have

$$\prod_{i \in \mathcal{T}'} w_i = \psi(S_{\mathbf{i}} + z_{\mathbf{i}}), \quad \int_{\mathbb{R}^{2j}} \psi(y) dy = \prod_{i \in \mathcal{T}'} A_h(E_i), \quad \frac{|z_{\mathbf{i}}|^2}{2\tau} = \sum_{i \in \mathcal{T}'} \frac{g_i^2}{8d^2\tau} = \sum_{i \in \mathcal{T}'} \frac{g_i^2}{8\ell^2},$$

where we used (4.9) and $\ell^2 = d^2\tau$ in the last two equalities. Therefore the first part of Proposition 4.4, applied with $k = j$ and $x = z_{\mathbf{i}}$, gives

$$\left| \mathbb{E} \left[\prod_{i \in \mathcal{T}'} w_i \mid H \right] - \prod_{i \in \mathcal{T}'} \bar{w}_i \right| \leq \epsilon_N \tau^{-j},$$

where $\epsilon_N = \epsilon_N(j, \Psi_j)$ is the sequence of Proposition 4.4, which does not depend on $\mathcal{T}'$.

We now sum over the subsets $\mathcal{T}' \subset \mathcal{C}_R$ with $|\mathcal{T}'| = j$. By the second part of Lemma 4.6, their number is bounded by

$$\frac{|\mathcal{C}_R|^j}{j!} \leq \frac{(K_R \tau)^j}{j!}, \quad \text{where} \quad K_R = C_{I'}(1 + R^2)^{p*}.$$

Hence, on $\mathcal{A}_N^* \cap \mathcal{A}_{N,j}$,

$$|\mathbb{E}[P_j(w) \mid H] - P_j(\bar{w})| \leq \frac{(K_R \tau)^j}{j!} \epsilon_N \tau^{-j} = \frac{K_R^j \epsilon_N}{j!}. \quad (4.42)$$

Next, we compare $P_j(\bar{w})$ with $(\sum_{i \in \mathcal{C}_R} \bar{w}_i)^j / j!$. We may assume that $j \geq 2$, since the two quantities coincide for $j = 1$. We expand $(\sum_{i \in \mathcal{C}_R} \bar{w}_i)^j$ as a sum over ordered j -tuples of elements of $\mathcal{C}_R$. The tuples with distinct entries contribute $j!P_j(\bar{w})$. Every other tuple has equal entries in at least two positions, and for each of the $\binom{j}{2}$ choices of two positions, the tuples with equal entries in these positions contribute

$$\left(\sum_{i \in \mathcal{C}_R} \bar{w}_i^2 \right) \left(\sum_{i \in \mathcal{C}_R} \bar{w}_i \right)^{j-2} \leq \left(\max_{i \in \mathcal{C}_R} \bar{w}_i \right) \left(\sum_{i \in \mathcal{C}_R} \bar{w}_i \right)^{j-1}.$$

Let $\|A_h\|_\infty = \sup_E A_h(E)$, which is finite by the first property of ψ_E . Then

$$\max_{i \in \mathcal{C}_R} \bar{w}_i \leq \frac{\|A_h\|_\infty}{2\pi\tau}, \quad \sum_{i \in \mathcal{C}_R} \bar{w}_i \leq \frac{K_R \|A_h\|_\infty}{2\pi},$$

where the second bound follows from the second part of Lemma 4.6. We conclude that

$$0 \leq \left(\sum_{i \in \mathcal{C}_R} \bar{w}_i \right)^j - j! P_j(\bar{w}) \leq \binom{j}{2} \frac{\|A_h\|_\infty}{2\pi\tau} \left(\frac{K_R \|A_h\|_\infty}{2\pi} \right)^{j-1} = O(\tau^{-1}). \quad (4.43)$$

Finally, we show that

$$\sum_{i \in \mathcal{C}_R} \bar{w}_i \longrightarrow \mathsf{L}_R(f) \quad \text{in probability.} \quad (4.44)$$

For $\omega \in (0, 1)$, define functions on $[0, \infty) \times (-2, 2)$ by

$$\begin{aligned} Q_\omega^+(y, E) &= \frac{1}{2\pi} \min \left\{ 1, \frac{(R + \omega - y)_+}{\omega} \right\} e^{-y^2/8} A_h(E), \\ Q_\omega^-(y, E) &= \frac{1}{2\pi} \min \left\{ 1, \frac{(R - y)_+}{\omega} \right\} e^{-y^2/8} A_h(E). \end{aligned}$$

They are continuous with support in $[0, R + 1] \times I_h$, and they satisfy

$$Q_\omega^-(y, E) \leq \frac{1}{2\pi} \mathbf{1}\{y \leq R\} e^{-y^2/8} A_h(E) \leq Q_\omega^+(y, E).$$

Note that $R\ell \leq s$ for large N by (4.2). Since $A_h(E_i) = 0$ unless $E_i \in I_h \subset I'$, we then have

$$\sum_{i \in \mathcal{C}_R} \bar{w}_i = \frac{1}{\tau} \sum_{i=1}^{N-1} \frac{1}{2\pi} \mathbf{1}\{g_i \leq R\ell\} \exp \left(-\frac{g_i^2}{8\ell^2} \right) A_h(E_i).$$

Using (4.27) with $R' = R + 1$, we deduce that on $\mathcal{A}_N^*$, for large N ,

$$\frac{1}{\tau} \sum_{1 \leq i < j' \leq N} Q_\omega^-\left(\frac{\lambda_{j'} - \lambda_i}{\ell}, \frac{\lambda_i + \lambda_{j'}}{2}\right) \leq \sum_{i \in \mathcal{C}_R} \bar{w}_i \leq \frac{1}{\tau} \sum_{1 \leq i < j' \leq N} Q_\omega^+\left(\frac{\lambda_{j'} - \lambda_i}{\ell}, \frac{\lambda_i + \lambda_{j'}}{2}\right).$$

The first part of Proposition 4.1 implies that the left and right sides converge in probability to

$$\int_{-2}^2 \int_0^\infty c(E) y Q_\omega^-(y, E) dy dE \quad \text{and} \quad \int_{-2}^2 \int_0^\infty c(E) y Q_\omega^+(y, E) dy dE,$$

respectively. As $\omega \rightarrow 0$, dominated convergence shows that these integrals converge to

$$\frac{1}{2\pi} \int_0^R y e^{-y^2/8} dy \int_{-2}^2 c(E) A_h(E) dE = \frac{1}{2\pi} \cdot 4(1 - e^{-R^2/8}) \cdot \frac{\pi}{2} \mathsf{L}(f) = \mathsf{L}_R(f),$$

where in the first equality we used (4.37) and (4.21). Now let $\omega' > 0$, and choose $\omega \in (0, 1)$ such that each of these integrals differs from $\mathsf{L}_R(f)$ by at most ω' . Since $\mathbb{P}(\mathcal{A}_N^*) \rightarrow 1$, the two-sided bound on $\sum_{i \in \mathcal{C}_R} \bar{w}_i$ above and the convergence of its outer sides imply that

$$\left| \sum_{i \in \mathcal{C}_R} \bar{w}_i - \mathsf{L}_R(f) \right| \leq 2\omega'$$

with probability tending to one. The number $\omega' > 0$ was arbitrary, so (4.44) follows.

We now combine the three steps. The convergence (4.44) and the continuous mapping theorem give

$$\frac{1}{j!} \left(\sum_{i \in \mathcal{C}_R} \bar{w}_i \right)^j \longrightarrow \frac{\mathbb{L}_R(f)^j}{j!} \quad \text{in probability.}$$

The estimates (4.42) and (4.43), whose right sides tend to zero, hold on $\mathcal{A}_N^* \cap \mathcal{A}_{N,j}$, and $\mathbb{P}(\mathcal{A}_N^* \cap \mathcal{A}_{N,j}) \rightarrow 1$ by Lemma 4.6 and Proposition 4.4. Combining these facts with the last display, we obtain (4.40). This completes the proof. $\square$

Continuous test functions and the truncation. In the next two lemmas and in the proof of Theorem 1.2, we use the elementary inequality

$$|e^{-y_1} - e^{-y_2}| \leq \min\{1, |y_1 - y_2|\}, \quad y_1, y_2 \geq 0. \quad (4.45)$$

We first extend Lemma 4.9 to continuous test functions.

Lemma 4.10. *Let $R > 0$ and $x_0 > 0$, and let $f \in C_c^+$ vanish outside $[0, x_0] \times I$. Then*

$$\lim_{N \rightarrow \infty} \mathbb{E} e^{-\hat{\Xi}_{N,R}(f)} = e^{-\mathbb{L}_R(f)}. \quad (4.46)$$

Proof. To apply Lemma 4.9, we approximate f by test functions of the form $h(r^2, E)$. Define $\bar{h}: \mathbb{R}^2 \rightarrow [0, \infty)$ by $\bar{h}(u, E) = 0$ for $E \notin (-2, 2)$ and

$$\bar{h}(u, E) = \min\{1, (1+u)_+\} f((\max\{u, 0\})^{1/2}, E) \quad \text{for } u \in \mathbb{R}, E \in (-2, 2).$$

Then $\bar{h}$ is continuous, and

$$\text{supp } \bar{h} \subset [-1, x_0^2] \times I, \quad \bar{h}(r^2, E) = f(r, E) \quad \text{for } r \geq 0, E \in (-2, 2).$$

For $\omega > 0$, let h_ω be the convolution of $\bar{h}$ with a smooth nonnegative function of integral one supported in the disk of radius ω . Then $h_\omega \in C_c^\infty(\mathbb{R}^2)$ is nonnegative, and $\|h_\omega - \bar{h}\|_\infty \rightarrow 0$ as $\omega \rightarrow 0$, since $\bar{h}$ is uniformly continuous. For $\omega < 1$ smaller than the distance from I to $\mathbb{R} \setminus I'$, we have $\text{supp } h_\omega \subset [-2, x_0^2 + 1] \times \text{int}(I')$. Set

$$f_\omega(r, E) = h_\omega(r^2, E), \quad (r, E) \in [0, \infty) \times (-2, 2).$$

Then f and f_ω vanish outside $[0, x_0 + 1] \times I'$, and

$$\sup_{(r, E) \in [0, \infty) \times (-2, 2)} |f_\omega(r, E) - f(r, E)| \leq \|h_\omega - \bar{h}\|_\infty.$$

We now compare the Laplace functionals of $\hat{\Xi}_{N,R}$ at f and at f_ω . Using (4.45), we have

$$\begin{aligned} |\mathbb{E} e^{-\hat{\Xi}_{N,R}(f)} - \mathbb{E} e^{-\hat{\Xi}_{N,R}(f_\omega)}| &\leq \mathbb{E} \min\{1, \|h_\omega - \bar{h}\|_\infty \hat{\Xi}_{N,R}([0, x_0 + 1] \times I')\} \\ &\leq \mathbb{P}((\mathcal{A}_N^*)^c) + \|h_\omega - \bar{h}\|_\infty \mathbb{E} \left[\mathbf{1}_{\mathcal{A}_N^*} \mathbb{E} [\hat{\Xi}_N([0, x_0 + 1] \times I') \mid H] \right] \\ &\leq o(1) + C \|h_\omega - \bar{h}\|_\infty. \end{aligned}$$

In the last inequality, we used (4.32) with $x_0 + 1$ in place of x_0 , and $\widehat{\Xi}_{N,R} \leq \widehat{\Xi}_N$. Next, Lemma 4.9 and the bounded convergence theorem give

$$\lim_{N \rightarrow \infty} \mathbb{E} e^{-\widehat{\Xi}_{N,R}(f_\omega)} = e^{-\mathbb{L}_R(f_\omega)}.$$

Moreover,

$$|\mathbb{L}_R(f_\omega) - \mathbb{L}_R(f)| \leq \int |f_\omega - f| d\nu_R \leq \|h_\omega - \bar{h}\|_\infty \nu([0, x_0 + 1] \times I').$$

Combining these estimates, we obtain

$$\limsup_{N \rightarrow \infty} |\mathbb{E} e^{-\widehat{\Xi}_{N,R}(f)} - e^{-\mathbb{L}_R(f)}| \leq (C + \nu([0, x_0 + 1] \times I')) \|h_\omega - \bar{h}\|_\infty,$$

and letting $\omega \rightarrow 0$ proves (4.46). $\square$

Next, we remove the truncation, using the bound (4.33), which relies on the second part of Proposition 4.4.

Lemma 4.11. *Let $x_0 > 0$, and let $f \in C_c^+$ vanish outside $[0, x_0] \times I$. Then*

$$\lim_{N \rightarrow \infty} \mathbb{E} e^{-\widehat{\Xi}_N(f)} = e^{-\mathbb{L}(f)}. \quad (4.47)$$

Proof. Let $C, c > 0$ be the constants of Lemma 4.8 for the window x_0 , and let $R > 0$. Then

$$\begin{aligned} |\mathbb{E} e^{-\widehat{\Xi}_N(f)} - \mathbb{E} e^{-\widehat{\Xi}_{N,R}(f)}| &\leq \mathbb{E} \min \{1, \|f\|_\infty (\widehat{\Xi}_N - \widehat{\Xi}_{N,R})([0, x_0] \times I')\} \\ &\leq \mathbb{P}((\mathcal{A}_N^*)^c) + C \|f\|_\infty \sup_{y \geq R} ((1 + y^2)^{p_*} e^{-cy}), \end{aligned}$$

where we used (4.45) and the fact that f vanishes outside $[0, x_0] \times I$ in the first inequality, and (4.33) in the second. Further, (4.22) gives

$$|e^{-\mathbb{L}_R(f)} - e^{-\mathbb{L}(f)}| \leq \mathbb{L}(f) - \mathbb{L}_R(f) = e^{-R^2/8} \mathbb{L}(f).$$

Together with Lemma 4.10, these two bounds yield

$$\limsup_{N \rightarrow \infty} |\mathbb{E} e^{-\widehat{\Xi}_N(f)} - e^{-\mathbb{L}(f)}| \leq C \|f\|_\infty \sup_{y \geq R} ((1 + y^2)^{p_*} e^{-cy}) + e^{-R^2/8} \mathbb{L}(f).$$

Letting $R \rightarrow \infty$, we obtain (4.47). $\square$

Void probabilities. We will deduce (1.6) from the convergence of Ξ_N by means of the following standard consequence of vague convergence.

Lemma 4.12. *Let $\mathfrak{S} \subset [0, \infty) \times (-2, 2)$ be a Borel set whose closure in $[0, \infty) \times (-2, 2)$ is compact and whose boundary $\partial\mathfrak{S}$ in this space satisfies $\nu(\partial\mathfrak{S}) = 0$. Let $(\mu_N)_{N \geq 2}$ be random locally finite counting measures on $[0, \infty) \times (-2, 2)$ that converge in distribution to Ξ with respect to the vague topology. Then*

$$\lim_{N \rightarrow \infty} \mathbb{P}(\mu_N(\mathfrak{S}) = 0) = e^{-\nu(\mathfrak{S})}.$$

Proof. As Ξ has intensity ν , the assumption $\nu(\partial\mathfrak{S}) = 0$ implies that $\Xi(\partial\mathfrak{S}) = 0$ almost surely. Since $\mathfrak{S}$ is relatively compact, the map $\mu \mapsto \mu(\mathfrak{S})$ is continuous, for the vague topology, at every locally finite measure μ with $\mu(\partial\mathfrak{S}) = 0$ (see [42, Chapter 4]). The continuous mapping theorem therefore shows that $\mu_N(\mathfrak{S})$ converges in distribution to $\Xi(\mathfrak{S})$, which is a Poisson random variable with mean $\nu(\mathfrak{S})$. These random variables take integer values, so

$$\lim_{N \rightarrow \infty} \mathbb{P}(\mu_N(\mathfrak{S}) = 0) = \mathbb{P}(\Xi(\mathfrak{S}) = 0) = e^{-\nu(\mathfrak{S})}.$$

□

Proof of Theorem 1.2. Let $f \in C_c^+$. Since the support of f is compact, f vanishes outside $[0, x_0] \times I$ for some $x_0 > 0$ and some bulk interval I . Throughout the proof, I' , $\mathcal{C}$, $\mathcal{C}_R$, $\widehat{\Xi}_N$, $\widehat{\Xi}_{N,R}$, and $\mathcal{A}_N^*$ are defined as in the setup above, with this choice of I . We extend f by zero to $[0, \infty) \times \mathbb{R}$, and for $\omega > 0$ we define

$$\text{osc}_f(\omega) = \sup \{ |f(r, E) - f(r', E')| : (r, E), (r', E') \in [0, \infty) \times \mathbb{R}, |(r, E) - (r', E')| \leq \omega \}. \quad (4.48)$$

We will show that

$$\lim_{N \rightarrow \infty} \mathbb{E} e^{-\Xi_N(f)} = e^{-\mathbb{L}(f)}. \quad (4.49)$$

By Lemma 4.11, the limit (4.47) holds for the process $\widehat{\Xi}_N$ of linearized gaps. It remains to pass from $\widehat{\Xi}_N$ to Ξ_N .

Passage to the resampled matrix and to H . Define

$$\widetilde{\Xi}_N = \sum_{i=1}^{N-1} \delta_{(d^{-1}\tilde{g}_i, \tilde{\lambda}_i)},$$

regarded as a measure on $[0, \infty) \times (-2, 2)$ in the same way as Ξ_N , so that

$$\widetilde{\Xi}_N(f) = \sum_i f(d^{-1}\tilde{g}_i, \tilde{\lambda}_i).$$

The eigenvalues are continuous functions of the matrix entries, so $\Xi_N(f)$ is a measurable function of H , and $\tilde{\Xi}_N(f)$ is the same function of $\tilde{H}$. Hence Lemma 4.2 gives

$$\mathbb{E}e^{-\Xi_N(f)} = \mathbb{E}e^{-\tilde{\Xi}_N(f)}. \quad (4.50)$$

We now compare $\tilde{\Xi}_N(f)$ with $\hat{\Xi}_N(f)$. Note that $\hat{\Xi}_N$ records the midpoints E_i in the spectrum of H , whereas $\tilde{\Xi}_N$ records the left endpoints $\tilde{\lambda}_i$ in the spectrum of $\tilde{H}$. By (4.18), we have $\delta_N \leq d$ for large N .

Suppose that N is large and that $\mathcal{A}_N^* \cap \mathcal{B}_N$ holds, so that the conclusions of the second part of Proposition 4.5 hold. Let $i \in [N-1]$ be such that $f(d^{-1}\tilde{g}_i, \tilde{\lambda}_i) \neq 0$. Then $\tilde{g}_i \leq x_0 d$ and $\tilde{\lambda}_i \in I$. We have $x_0 d < s/2$ for large N by (4.2), so (4.20) implies that $g_i \leq s$. Then (4.19) gives

$$|E_i - \tilde{\lambda}_i| \leq |E_i - \tilde{E}_i| + \frac{\tilde{g}_i}{2} \leq w_N + \frac{\delta_N}{2} + \frac{x_0 d}{2}, \quad 2|z_i + S_i| \leq d^{-1}\tilde{g}_i + \frac{\delta_N}{d} \leq x_0 + 1.$$

In particular, $E_i \in I'$ for large N , so $i \in \mathcal{C}$.

We have shown that every nonzero term of $\tilde{\Xi}_N(f)$ comes from some $i \in \mathcal{C}$ with $2|z_i + S_i| \leq x_0 + 1$. The same holds for the nonzero terms of $\hat{\Xi}_N(f)$, so

$$\tilde{\Xi}_N(f) - \hat{\Xi}_N(f) = \sum_{\substack{i \in \mathcal{C} \\ 2|z_i + S_i| \leq x_0 + 1}} \left(f(d^{-1}\tilde{g}_i, \tilde{\lambda}_i) - f(2|z_i + S_i|, E_i) \right).$$

For every i in this sum, (4.19) gives

$$\begin{aligned} \tilde{g}_i &\leq d \left(2|z_i + S_i| + \frac{\delta_N}{d} \right) \leq (x_0 + 2)d, \\ |d^{-1}\tilde{g}_i - 2|z_i + S_i|| + |\tilde{\lambda}_i - E_i| &\leq \frac{\delta_N}{d} + \frac{\tilde{g}_i}{2} + |\tilde{E}_i - E_i| \leq \frac{\delta_N}{d} + \frac{(x_0 + 2)d}{2} + w_N + \frac{\delta_N}{2}. \end{aligned}$$

We denote the right side of the second line by ς_N . It is deterministic, and it tends to zero by (4.18). Using the definition (4.48), we conclude that on $\mathcal{A}_N^* \cap \mathcal{B}_N$,

$$|\tilde{\Xi}_N(f) - \hat{\Xi}_N(f)| \leq \text{osc}_f(\varsigma_N) \hat{\Xi}_N([0, x_0 + 1] \times I'),$$

and $\text{osc}_f(\varsigma_N) = o(1)$ because f is uniformly continuous.

We can now bound the difference of the Laplace functionals. Using the first part of Proposition 4.5 and (4.32) with $x_0 + 1$ in place of x_0 , we find

$$\begin{aligned} |\mathbb{E}e^{-\tilde{\Xi}_N(f)} - \mathbb{E}e^{-\hat{\Xi}_N(f)}| &\leq \mathbb{P}((\mathcal{A}_N^*)^c) + \mathbb{E}[\mathbf{1}_{\mathcal{A}_N^*} \mathbb{P}(\mathcal{B}_N^c | H)] \\ &\quad + \text{osc}_f(\varsigma_N) \mathbb{E}[\mathbf{1}_{\mathcal{A}_N^*} \mathbb{E}[\hat{\Xi}_N([0, x_0 + 1] \times I') | H]] \\ &\leq o(1) + o(1) + C \text{osc}_f(\varsigma_N) = o(1). \end{aligned}$$

Here the conditional probability of $\mathcal{B}_N^c$ enters only through the bound $|e^{-\tilde{\Xi}_N(f)} - e^{-\tilde{\Xi}_N(f)}| \leq 1$. Combining the last estimate with (4.47) and (4.50), we obtain (4.49).

Vague convergence. The test function $f \in C_c^+$ was arbitrary, so (4.49) holds for all $f \in C_c^+$, and $e^{-L(f)} = \mathbb{E}e^{-\Xi(f)}$. The space $[0, \infty) \times (-2, 2)$ is locally compact, separable, and metrizable, and the measures Ξ_N and ν are locally finite on it. By the Laplace functional criterion for convergence in distribution of random measures (see [42, Chapter 4]), the convergence of $\mathbb{E}e^{-\Xi_N(f)}$ to $\mathbb{E}e^{-\Xi(f)}$ for all $f \in C_c^+$ implies that Ξ_N converges in distribution to Ξ for the vague topology. This is the first claim of Theorem 1.2.

The smallest gap. In this step, I denotes a nonempty interval with closure $\bar{I} \subset (-2, 2)$, as in Theorem 1.2, and we fix $x \geq 0$. Set $\mathfrak{S} = [0, x] \times I$. The definition of Ξ_N gives

$$\left\{ N^{3/2} \min_{\lambda_i \in I} g_i > x \right\} = \{ \Xi_N(\mathfrak{S}) = 0 \},$$

where the minimum is over $i \in [N-1]$ such that $\lambda_i \in I$, and an empty minimum is $+\infty$.

The closure of $\mathfrak{S}$ in $[0, \infty) \times (-2, 2)$ is the compact set $[0, x] \times \bar{I}$, and the interior of $\mathfrak{S}$ contains $[0, x) \times \text{int}(I)$. The difference of these two sets is contained in $(\{x\} \times \bar{I}) \cup ([0, x] \times (\bar{I} \setminus \text{int}(I)))$, which has Lebesgue measure zero, since $\bar{I} \setminus \text{int}(I)$ consists of at most two points. The measure ν has a density with respect to the Lebesgue measure, so $\nu(\partial \mathfrak{S}) = 0$ and

$$\nu(\mathfrak{S}) = \int_I \int_0^x c(E) y \, dy \, dE = \frac{\pi^2 x^2}{12} \int_I \rho_{\text{sc}}(E)^3 \, dE, \quad (4.51)$$

where we used the definition of $c(E)$ in the last equality. By the first claim of Theorem 1.2, we may apply Lemma 4.12 with $\mu_N = \Xi_N$, which gives

$$\lim_{N \rightarrow \infty} \mathbb{P}(\Xi_N(\mathfrak{S}) = 0) = e^{-\nu(\mathfrak{S})}.$$

Together with (4.51), this proves the second claim of Theorem 1.2. □

A comparison estimate for pair statistics

In this section, we compare Laplace transforms of pair statistics of H with those of the Ornstein–Uhlenbeck interpolation H_t defined in (3.2). In Section 6, pair statistics of H_t are computed for suitable small times t using the Gaussian component of H_t , and Proposition 5.9 below transfers these computations to H . The pair statistics count pairs of eigenvalues at distance of order $\ell_\alpha = N^{-1-\alpha}$, weighted by a smooth function of their midpoint, and the observables we compare are exponentials of nonpositive multiples of such statistics.

Our argument follows the continuity estimate of Bourgade and Yau [11, Lemma A.1] (see also [23, Chapter 15]), which is a continuous version of the Lindeberg replacement strategy [15, 70] along the Ornstein–Uhlenbeck interpolation. It requires bounds on the derivatives of the observable in one matrix entry, up to order three, or four when $\mathbb{E}X^3 = 0$. Obtaining these bounds is the main difficulty. Derivatives of eigenvalues of order two and higher contain inverse gaps $(\lambda_i - \lambda_j)^{-1}$, and bounding the resulting sums would require control of pairs of nearby eigenvalues, which is what the comparison is meant to establish. We therefore avoid eigenvalue perturbation theory. Instead, we show that the pair statistic extends to a holomorphic function of the matrix entries whose real part is nonnegative as long as the spectrum stays in a thin rectangle of height of order ℓ_α around $[-4, 4]$. The observable then has modulus at most one on a complex disk in each entry, whose radius is slightly smaller than $N^{-\alpha}$, and Cauchy’s estimates bound its derivatives by powers of the inverse radius, uniformly in the normalization of the statistic. These bounds suffice because the radius is much larger than the typical size $N^{-1/2}$ of an entry.

Section 5.1 introduces the smoothed energy weights. In Section 5.2, we show that a complex perturbation of one entry of a regular matrix keeps its spectrum in the thin rectangle; this is the only place in this section where we use the conditions of Definition 3.1. Section 5.3 constructs the holomorphic extension, and Section 5.4 proves the comparison estimate. The proof uses the observable only through the two properties listed in Remark 5.10.

5.1 Smoothed energy weights

Recall the scale $\ell_\alpha = N^{-1-\alpha}$ from (3.15). For $\alpha \in (0, 1/2)$, a real symmetric $N \times N$ matrix A , and functions $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ and $f: [0, \infty) \rightarrow \mathbb{R}$, we define the pair statistic

$$\mathcal{X}_\alpha(A; \varphi, f) = \sum_{1 \leq i < j \leq N} \varphi\left(\frac{\lambda_i(A) + \lambda_j(A)}{2}\right) f\left(\frac{\lambda_j(A) - \lambda_i(A)}{\ell_\alpha}\right). \quad (5.1)$$

Throughout the rest of this section, we fix $\alpha \in (0, 1/2)$. Recall the functions f_p from (3.33). For an integer $p \geq 1$, we extend f_p to $\mathbb{C} \setminus \{i, -i\}$ by the same formula, $f_p(z) = (1 + z^2)^{-p}$.

We set $h_N = N^{-1/10}$, and for a nonnegative $\varphi \in C_c^\infty((-2, 2))$, we define the smoothed energy weight

$$\varphi_N(w) = \frac{1}{\sqrt{2\pi} h_N} \int_{\mathbb{R}} \varphi(y) \exp\left(-\frac{(w - y)^2}{2h_N^2}\right) dy, \quad w \in \mathbb{C}. \quad (5.2)$$

We use this notation for every nonnegative function in $C_c^\infty((-2, 2))$; for example, $\varphi_{j,N}$ in Definition 5.4 below denotes the function defined by (5.2) with φ_j in place of φ . Differentiating under the integral sign, we see that φ_N is an entire function. For real x , the number $\varphi_N(x)$ is the average of φ with respect to the Gaussian density with mean x and variance h_N^2 . For $z \in \mathbb{C} \setminus \{0\}$, we let $\arg z \in (-\pi, \pi]$ denote the principal argument of z .

Part (3) of the following lemma controls the argument of φ_N near the real axis. We will use it, together with part (1), in Section 5.3, and we will use parts (1), (2), (4), and (5) in Section 6 and Section 9.

Lemma 5.1. *Let $\varphi \in C_c^\infty((-2, 2))$ be nonnegative. Then the following statements hold for all $N \geq 1$.*

(1) *For all $x \in \mathbb{R}$, we have $0 \leq \varphi_N(x) \leq \|\varphi\|_\infty$. If φ is not identically zero, then $\varphi_N(x) > 0$ for all $x \in \mathbb{R}$.*

(2) *For all $x \in \mathbb{R}$ with $|x| \leq 5$ and all $w \in \mathbb{C}$ with $|w - x| \leq 1$, we have*

$$|\varphi_N(w) - \varphi_N(x)| \leq (\exp(8|w - x|h_N^{-2}) - 1)\varphi_N(x). \quad (5.3)$$

If moreover $w \in \mathbb{R}$, then

$$\exp(-8|w - x|h_N^{-2})\varphi_N(x) \leq \varphi_N(w) \leq \exp(8|w - x|h_N^{-2})\varphi_N(x). \quad (5.4)$$

(3) *For all $w \in \mathbb{C}$ such that $(|\operatorname{Re} w| + 2)|\operatorname{Im} w| \leq h_N^2$, either $\varphi_N(w) = 0$ or*

$$|\arg \varphi_N(w)| \leq (|\operatorname{Re} w| + 2)|\operatorname{Im} w|h_N^{-2}. \quad (5.5)$$

(4) *For all $x \in \mathbb{R}$, we have*

$$\varphi_N(x) \leq 2\|\varphi\|_\infty \exp(-\operatorname{dist}(x, \operatorname{supp} \varphi)^2 / (2h_N^2)). \quad (5.6)$$

(5) We have $\sup_{x \in \mathbb{R}} |\varphi_N(x) - \varphi(x)| \leq \|\varphi'\|_\infty h_N$.

Proof. The first claim in part (1) holds since $\varphi_N(x)$ is the average of φ against a probability density. The second holds since this density is positive everywhere and φ is continuous, nonnegative, and not identically zero.

We now prove part (2). Fix $x \in \mathbb{R}$ with $|x| \leq 5$ and $w \in \mathbb{C}$ with $|w - x| \leq 1$. For $y \in \text{supp } \varphi$, we have $|x - y| \leq 7$, and expanding $(w - y)^2$ around $x - y$ gives

$$\exp\left(-\frac{(w - y)^2}{2h_N^2}\right) = \exp\left(-\frac{(x - y)^2}{2h_N^2}\right) \exp\left(-\frac{2(x - y)(w - x) + (w - x)^2}{2h_N^2}\right). \quad (5.7)$$

The exponent in the second factor satisfies

$$\frac{|2(x - y)(w - x) + (w - x)^2|}{2h_N^2} \leq \frac{14|w - x| + |w - x|^2}{2h_N^2} \leq 8|w - x|h_N^{-2}. \quad (5.8)$$

Using $|e^z - 1| \leq e^{|z|} - 1$ for $z \in \mathbb{C}$ and integrating (5.7) against $\varphi(y) dy$, we find

$$\begin{aligned} |\varphi_N(w) - \varphi_N(x)| &\leq \frac{1}{\sqrt{2\pi} h_N} \int_{\mathbb{R}} \varphi(y) \exp\left(-\frac{(x - y)^2}{2h_N^2}\right) (e^{8|w - x|h_N^{-2}} - 1) dy \\ &= (e^{8|w - x|h_N^{-2}} - 1) \varphi_N(x), \end{aligned}$$

which is (5.3).

When w is real, the exponent in the second factor of (5.7) is real, so by (5.8), we have

$$e^{-8|w - x|h_N^{-2}} \leq \exp\left(-\frac{2(x - y)(w - x) + (w - x)^2}{2h_N^2}\right) \leq e^{8|w - x|h_N^{-2}}, \quad y \in \text{supp } \varphi.$$

Integrating (5.7) against $\varphi(y) dy$ gives (5.4).

To prove part (3), we expand

$$(w - y)^2 = (\text{Re } w - y)^2 - (\text{Im } w)^2 + 2i \text{Im } w (\text{Re } w - y),$$

which gives

$$\begin{aligned} \varphi_N(w) &= \frac{1}{\sqrt{2\pi} h_N} \int_{\mathbb{R}} \varphi(y) \exp\left(-\frac{(\text{Re } w - y)^2 - (\text{Im } w)^2}{2h_N^2}\right) \\ &\quad \times \exp\left(-\frac{i \text{Im } w (\text{Re } w - y)}{h_N^2}\right) dy. \end{aligned}$$

The first two factors in the integrand are nonnegative. For $y \in \text{supp } \varphi$, the phase of the last factor has absolute value at most

$$\beta = (|\text{Re } w| + 2)|\text{Im } w|h_N^{-2} \leq 1.$$

Since $\beta < \pi/2$, the closed sector of half-angle β is an intersection of two closed half-planes,

$$\{0\} \cup \{z \neq 0 : |\arg z| \leq \beta\} = \{z : \operatorname{Re}(e^{i(\pi/2-\beta)}z) \geq 0\} \cap \{z : \operatorname{Re}(e^{-i(\pi/2-\beta)}z) \geq 0\}.$$

Each of these half-planes is preserved by integration against a nonnegative weight, so $\varphi_N(w)$ belongs to the sector, as claimed in (5.5).

For parts (4) and (5), we change variables in (5.2) and write, for $x \in \mathbb{R}$,

$$\varphi_N(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \varphi(x + h_N y) e^{-y^2/2} dy.$$

The integrand vanishes unless $h_N|y| \geq \operatorname{dist}(x, \operatorname{supp} \varphi)$, so (5.6) follows from the Gaussian tail bound

$$\int_{|y| \geq y_0} e^{-y^2/2} dy \leq 2\sqrt{2\pi} e^{-y_0^2/2}, \quad y_0 \geq 0.$$

Further, the mean value theorem gives

$$|\varphi_N(x) - \varphi(x)| \leq \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |\varphi(x + h_N y) - \varphi(x)| e^{-y^2/2} dy \leq \left(\frac{2}{\pi}\right)^{1/2} \|\varphi'\|_{\infty} h_N \leq \|\varphi'\|_{\infty} h_N,$$

which is part (5). $\square$

5.2 Spectra under a perturbation of one entry

For an integer $p \geq 1$, we set

$$\omega_p = \frac{1}{1000(p+1)}, \quad c_p = \frac{\omega_p}{100}, \quad (5.9)$$

and we define the open rectangle

$$\mathfrak{R}_{\alpha,p} = \{w \in \mathbb{C} : |\operatorname{Re} w| < 4, |\operatorname{Im} w| < \omega_p \ell_{\alpha}\}. \quad (5.10)$$

Recall the set $\mathcal{G}_{\varepsilon}$ of regular matrices from Definition 3.1 and the matrices V^{ab} from (3.1). In the next lemma, the size $8c_p N^{-\alpha-4\varepsilon}$ of the perturbation is much larger than the typical size $N^{-1/2}$ of an entry of H_t when $\alpha + 4\varepsilon < 1/2$.

Lemma 5.2. *Let $\varepsilon > 0$, and let $p \geq 1$ be an integer. There exists $N_0 = N_0(\alpha, p)$ such that the following holds for all $N \geq N_0$. Let $A \in \mathcal{G}_{\varepsilon}$, let $a, b \in [N]$ with $a \leq b$, and let $\zeta \in \mathbb{C}$ with $|\zeta| \leq 8c_p N^{-\alpha-4\varepsilon}$. Then $\operatorname{Spec}(A + \zeta V^{ab}) \subset \mathfrak{R}_{\alpha,p}$.*

Proof. We fix a spectral decomposition

$$A = \sum_{i=1}^N \lambda_i(A) v_i v_i^T,$$

where $(v_i)_{i \in [N]}$ is an orthonormal basis of $\mathbb{R}^N$. Recall from the definition of $\mathcal{G}_\varepsilon$ that $\|A\| \leq 3$ and $\|v_i\|_\infty \leq N^{-1/2+\varepsilon}$ for all $i \in [N]$.

We write $\partial\mathfrak{R}_{\alpha,p}$ for the boundary of $\mathfrak{R}_{\alpha,p}$. It consists of two horizontal sides on the lines $\operatorname{Im} w = \pm\omega_p\ell_\alpha$ and two vertical sides on the lines $\operatorname{Re} w = \pm 4$.

We begin by bounding the entries of the resolvent $(A - w)^{-1}$ for $w \in \partial\mathfrak{R}_{\alpha,p}$. Let w lie on a horizontal side, so that $|\operatorname{Re} w| \leq 4$ and $|\operatorname{Im} w| = \omega_p\ell_\alpha$. For $a', b' \in \{a, b\}$, the spectral decomposition gives

$$|((A - w)^{-1})_{a'b'}| \leq \sum_{i=1}^N \frac{|v_i(a')| |v_i(b')|}{|\lambda_i(A) - w|} \leq N^{-1+2\varepsilon} \sum_{i=1}^N \frac{1}{|\lambda_i(A) - w|}. \quad (5.11)$$

To bound the last sum, we partition the real line into the intervals

$$\mathcal{I}_n = [\operatorname{Re} w + n/N, \operatorname{Re} w + (n+1)/N), \quad n \in \mathbb{Z}.$$

Every eigenvalue of A belongs to some $\mathcal{I}_n$ with $|n| \leq 7N$, since $\lambda_i(A) \in [-3, 3]$ and $|\operatorname{Re} w| \leq 4$. Moreover, the definition of $\mathcal{G}_\varepsilon$ implies that each $\mathcal{I}_n$ contains at most N^ε eigenvalues of A , counted with multiplicity. For $\lambda_i(A) \in \mathcal{I}_n$, we have

$$|\lambda_i(A) - w| \geq \begin{cases} \omega_p\ell_\alpha & \text{if } n \in \{-1, 0\}, \\ n/N & \text{if } n \geq 1, \\ (|n| - 1)/N & \text{if } n \leq -2, \end{cases} \quad (5.12)$$

where the first bound uses $|\operatorname{Im} w| = \omega_p\ell_\alpha$, and the other two use $|\lambda_i(A) - w| \geq |\lambda_i(A) - \operatorname{Re} w|$. Summing over n , we obtain

$$\sum_{i=1}^N \frac{1}{|\lambda_i(A) - w|} \leq \frac{2N^\varepsilon}{\omega_p\ell_\alpha} + 2N^\varepsilon \sum_{n=1}^{7N} \frac{N}{n} \leq \frac{2N^{1+\alpha+\varepsilon}}{\omega_p} + 2N^{1+\varepsilon}(1 + \log(7N)).$$

We now choose $N_0 = N_0(\alpha, p)$ such that

$$2\omega_p(1 + \log(7N)) \leq N^\alpha \quad \text{for all } N \geq N_0.$$

For such N , the right side is bounded by $3N^{1+\alpha+\varepsilon}/\omega_p$, and inserting this bound into (5.11), we obtain

$$|((A - w)^{-1})_{a'b'}| \leq \frac{3N^{\alpha+3\varepsilon}}{\omega_p}, \quad a', b' \in \{a, b\}. \quad (5.13)$$

It remains to treat the vertical sides. There we have $|\operatorname{Re} w| = 4$, so $\operatorname{dist}(w, \operatorname{Spec} A) \geq 1$ since $\|A\| \leq 3$. Hence $\|(A - w)^{-1}\| \leq 1$ there, and (5.13) holds on all of $\partial\mathfrak{R}_{\alpha,p}$.

Next, we reduce the characteristic polynomial of $A + \zeta V^{ab}$ to a determinant of size at most two. Recall Sylvester's determinant identity

$$\det(\operatorname{Id}_N + Q_1 Q_2) = \det(\operatorname{Id}_2 + Q_2 Q_1), \quad Q_1 \in \mathbb{C}^{N \times 2}, \quad Q_2 \in \mathbb{C}^{2 \times N}.$$

For $a < b$, we have

$$\zeta V^{ab} = Q_1 Q_2, \quad Q_1 = (e_a, e_b), \quad Q_2 = \zeta(e_b, e_a)^T.$$

Factoring out $A - w$ and applying the identity, we find, for $w \notin \text{Spec } A$ and $x \in [0, 1]$,

$$\det(A + x\zeta V^{ab} - w) = \det(A - w) \det \left(\text{Id}_2 + x\zeta \begin{pmatrix} ((A - w)^{-1})_{ba} & ((A - w)^{-1})_{bb} \\ ((A - w)^{-1})_{aa} & ((A - w)^{-1})_{ab} \end{pmatrix} \right). \quad (5.14)$$

For $a = b$, the same reasoning with $Q_1 = e_a$ and $Q_2 = \zeta e_a^T$ gives

$$\det(A + x\zeta V^{aa} - w) = \det(A - w) (1 + x\zeta ((A - w)^{-1})_{aa}). \quad (5.15)$$

Now let $w \in \partial \mathfrak{R}_{\alpha, p}$ and $x \in [0, 1]$. Observe that $\partial \mathfrak{R}_{\alpha, p}$ does not meet $[-3, 3]$, so $w \notin \text{Spec } A$. Using (5.13) and our assumption on ζ , we see that every entry of the 2×2 matrix in (5.14), multiplied by $x\zeta$, has modulus bounded by

$$8c_p N^{-\alpha-4\epsilon} \cdot \frac{3N^{\alpha+3\epsilon}}{\omega_p} = \frac{24c_p}{\omega_p} N^{-\epsilon} \leq \frac{1}{4},$$

where we used $c_p = \omega_p/100$. For a 2×2 matrix Q' whose entries have modulus at most $1/4$, we have

$$|\det(\text{Id}_2 + Q')| = |1 + \text{Tr } Q' + \det Q'| \geq 1 - \frac{1}{2} - \frac{1}{8} > 0.$$

In the case $a = b$, the last factor in (5.15) has modulus at least $3/4$. We conclude that

$$\det(A + x\zeta V^{ab} - w) \neq 0 \quad \text{for all } w \in \partial \mathfrak{R}_{\alpha, p} \text{ and } x \in [0, 1]. \quad (5.16)$$

Finally, we use the argument principle. For $x \in [0, 1]$, we set

$$\chi_x(w) = \det(w \text{Id}_N - A - x\zeta V^{ab}), \quad w \in \mathbb{C}.$$

This is a monic polynomial of degree N in w , whose coefficients are polynomials in x . Combining (5.16) with the compactness of $[0, 1] \times \partial \mathfrak{R}_{\alpha, p}$, we see that $|\chi_x(w)|$ is bounded below by a positive constant for (x, w) in this set. Hence, with $\partial \mathfrak{R}_{\alpha, p}$ oriented counter-clockwise, the integral

$$\frac{1}{2\pi i} \oint_{\partial \mathfrak{R}_{\alpha, p}} \frac{\partial_w \chi_x(w)}{\chi_x(w)} dw,$$

which counts the zeros of χ_x in $\mathfrak{R}_{\alpha, p}$ with multiplicity, is a continuous function of $x \in [0, 1]$. It takes integer values, so it is constant.

For $x = 0$, the zeros of χ_0 are the eigenvalues of A , which belong to $[-3, 3] \subset \mathfrak{R}_{\alpha, p}$, so the integral is N for all x . Therefore all N zeros of χ_1 , which are the eigenvalues of $A + \zeta V^{ab}$ counted with algebraic multiplicity, lie in $\mathfrak{R}_{\alpha, p}$. This completes the proof. $\square$

5.3 A holomorphic extension of the pair statistic

We will need the following elementary fact.

Lemma 5.3. *Let $\mathcal{O} \subset \mathbb{C}$ be open. Then the set of matrices $A \in \mathbb{C}^{N \times N}$ such that $\text{Spec } A \subset \mathcal{O}$ is open in $\mathbb{C}^{N \times N}$.*

Proof. Let $(A_n)_{n \geq 1}$ be a sequence of matrices converging to A such that each A_n has an eigenvalue $w_n \notin \mathcal{O}$. The sequence (w_n) is bounded, since $|w_n| \leq \|A_n\|$, and after passing to a subsequence, we may assume that $w_n \rightarrow w$. Then $w \notin \mathcal{O}$, since $\mathbb{C} \setminus \mathcal{O}$ is closed, and

$$\det(A - w) = \lim_{n \rightarrow \infty} \det(A_n - w_n) = 0.$$

Hence the set of matrices with an eigenvalue outside $\mathcal{O}$ is closed. $\square$

We now introduce the tensor products used to define the extension. We identify $\mathbb{C}^N \otimes \mathbb{C}^N$ with $\mathbb{C}^{N^2}$, and we order the basis $(e_i \otimes e_j)_{i,j \in [N]}$ lexicographically in (i, j) . For $A \in \mathbb{C}^{N \times N}$, we define

$$\mathfrak{D}(A) = A \otimes \text{Id} - \text{Id} \otimes A, \quad \mathfrak{M}(A) = \frac{1}{2}(A \otimes \text{Id} + \text{Id} \otimes A). \quad (5.17)$$

By Schur's theorem, there exist a unitary matrix O and an upper triangular matrix T such that $A = OTO^*$, where the diagonal entries $T_{ii} = w_i$ are the eigenvalues of A , listed with algebraic multiplicity (in any prescribed order). Then

$$\begin{aligned} (O \otimes O)^* \mathfrak{D}(A) (O \otimes O) &= T \otimes \text{Id} - \text{Id} \otimes T, \\ (O \otimes O)^* \mathfrak{M}(A) (O \otimes O) &= \frac{1}{2}(T \otimes \text{Id} + \text{Id} \otimes T). \end{aligned} \quad (5.18)$$

The matrices $T \otimes \text{Id}$ and $\text{Id} \otimes T$ are upper triangular with respect to the lexicographic order, since

$$(T \otimes \text{Id})(e_i \otimes e_j) = \sum_{i' \leq i} T_{i'i} e_{i'} \otimes e_j, \quad (\text{Id} \otimes T)(e_i \otimes e_j) = \sum_{j' \leq j} T_{j'j} e_i \otimes e_{j'},$$

and their diagonal entries at $e_i \otimes e_j$ are w_i and w_j , respectively. Hence the right sides of (5.18) are upper triangular, with diagonal entries $w_i - w_j$ and $(w_i + w_j)/2$ at $e_i \otimes e_j$.

For a square matrix M , we define

$$\varphi_N(M) = \sum_{n=0}^{\infty} \frac{\varphi_N^{(n)}(0)}{n!} M^n, \quad (5.19)$$

where the series converges absolutely since φ_N is entire. Every power of an upper triangular matrix is upper triangular, with the corresponding powers of the diagonal entries

on its diagonal. Hence, for an upper triangular matrix M , the matrix $\varphi_N(M)$ is upper triangular with diagonal entries $\varphi_N(M_{ii})$. Similarly, the inverse of an invertible upper triangular matrix is upper triangular, with the reciprocal diagonal entries. Moreover, for every square matrix M and every unitary matrix Q of the same size, we have

$$\varphi_N(Q^*MQ) = Q^*\varphi_N(M)Q. \quad (5.20)$$

Suppose now that every eigenvalue w_i of A satisfies $|\operatorname{Im} w_i| < \omega_p \ell_\alpha$. By (5.18), the matrix $\operatorname{Id} + \ell_\alpha^{-2} \mathfrak{D}(A)^2$ is similar to an upper triangular matrix with diagonal entries $1 + ((w_i - w_j)/\ell_\alpha)^2$. For $z = (w_i - w_j)/\ell_\alpha$, we have $|\operatorname{Im} z| < 2\omega_p$, and

$$\operatorname{Re}(1 + z^2) = 1 + (\operatorname{Re} z)^2 - (\operatorname{Im} z)^2 \geq 1 - 4\omega_p^2 > 0.$$

Therefore $\operatorname{Id} + \ell_\alpha^{-2} \mathfrak{D}(A)^2$ is invertible, which justifies the following definition.

Definition 5.4. For an integer $p \geq 1$, we define

$$\mathcal{U}_{\alpha,p} = \{A \in \mathbb{C}^{N \times N} : |\operatorname{Im} w| < \omega_p \ell_\alpha \text{ for every eigenvalue } w \text{ of } A\}. \quad (5.21)$$

For a nonnegative $\varphi \in C_c^\infty((-2, 2))$, an integer $p \geq 1$, and $A \in \mathcal{U}_{\alpha,p}$, we define

$$\mathcal{X}_\alpha(A; \varphi_N, f_p) = \frac{1}{2} \operatorname{Tr} \left[\varphi_N(\mathfrak{M}(A)) (\operatorname{Id} + \ell_\alpha^{-2} \mathfrak{D}(A)^2)^{-p} \right] - \frac{1}{2} \operatorname{Tr} \varphi_N(A). \quad (5.22)$$

Let $\mathcal{J}$ be a finite nonempty set. For each $j \in \mathcal{J}$, let $\varphi_j \in C_c^\infty((-2, 2))$ be nonnegative, let $p_j \geq 1$ be an integer, and let $\theta_j \geq 0$. Set $p = \max_{j \in \mathcal{J}} p_j$. We define the observable

$$F(A) = \exp \left(- \sum_{j \in \mathcal{J}} \theta_j \mathcal{X}_\alpha(A; \varphi_{j,N}, f_{p_j}) \right), \quad A \in \mathcal{U}_{\alpha,p}. \quad (5.23)$$

Since $\omega_p \leq \omega_{p_j}$, we have $\mathcal{U}_{\alpha,p} \subset \mathcal{U}_{\alpha,p_j}$ for all $j \in \mathcal{J}$, so (5.23) is well defined. By Lemma 5.3, the set $\mathcal{U}_{\alpha,p}$ is open. It contains every real symmetric matrix and every matrix with spectrum in $\mathfrak{R}_{\alpha,p}$. The observable F depends on N and on the data $(\varphi_j, p_j, \theta_j)_{j \in \mathcal{J}}$, which we suppress from the notation. The parameters θ_j may depend on N , and all bounds below are uniform in them.

Lemma 5.5. Let $p \geq 1$ be an integer, and let $\varphi \in C_c^\infty((-2, 2))$ be nonnegative. Then the following statements hold for all $N \geq 1$.

- (1) The function $A \mapsto \mathcal{X}_\alpha(A; \varphi_N, f_p)$ is holomorphic in the entries of A on $\mathcal{U}_{\alpha,p}$. If $A \in \mathcal{U}_{\alpha,p}$ has eigenvalues $w_1, \dots, w_N$, listed with algebraic multiplicity in any order, then

$$\mathcal{X}_\alpha(A; \varphi_N, f_p) = \sum_{1 \leq i < j \leq N} \varphi_N \left(\frac{w_i + w_j}{2} \right) f_p \left(\frac{w_j - w_i}{\ell_\alpha} \right). \quad (5.24)$$

In particular, (5.22) agrees with (5.1) for every real symmetric matrix A .

- (2) Let $A \in \mathcal{U}_{\alpha,p}$, and suppose that $(|\operatorname{Re} w| + 2)|\operatorname{Im} w'| \leq h_N^2/200$ for all eigenvalues w and w' of A . Then every summand on the right side of (5.24) is either zero or has argument of absolute value at most $1/100$. In particular, $\operatorname{Re} \mathcal{X}_\alpha(A; \varphi_N, f_p) \geq 0$.
- (3) If $\operatorname{Spec} A \subset \mathfrak{R}_{\alpha,p}$, then $A \in \mathcal{U}_{\alpha,p}$, and A satisfies the hypothesis of part (2).
- (4) Let F be as in Definition 5.4, and suppose that $p = \max_{j \in \mathcal{J}} p_j$. Then F is holomorphic in the entries of A on $\mathcal{U}_{\alpha,p}$, and $|F(A)| \leq 1$ for every $A \in \mathcal{U}_{\alpha,p}$ satisfying the hypothesis of part (2). In particular, $|F(A)| \leq 1$ if $\operatorname{Spec} A \subset \mathfrak{R}_{\alpha,p}$ or if A is real symmetric.

Proof. We begin with holomorphy. The entries of $\mathfrak{M}(A)^n$ are polynomials in the entries of A , and $\|\mathfrak{M}(A)^n\| \leq \|A\|^n$. Since φ_N is entire, the series defining $\varphi_N(\mathfrak{M}(A))$ therefore converges uniformly on bounded subsets of $\mathbb{C}^{N \times N}$, and the entries of $\varphi_N(\mathfrak{M}(A))$ are holomorphic functions of the entries of A , as locally uniform limits of polynomials. The same holds for $\varphi_N(A)$. By Cramer's rule, the entries of $(\operatorname{Id} + \ell_\alpha^{-2} \mathfrak{D}(A)^2)^{-1}$ are quotients of polynomials in the entries of A , whose common denominator $\det(\operatorname{Id} + \ell_\alpha^{-2} \mathfrak{D}(A)^2)$ does not vanish on $\mathcal{U}_{\alpha,p}$. Hence the right side of (5.22) is holomorphic on $\mathcal{U}_{\alpha,p}$.

Next, we prove (5.24). Let $A = OTO^*$ as above, with $T_{ii} = w_i$. Using (5.18) and (5.20) with $Q = O \otimes O$, we find that

$$\begin{aligned} (O \otimes O)^* \varphi_N(\mathfrak{M}(A)) (\operatorname{Id} + \ell_\alpha^{-2} \mathfrak{D}(A)^2)^{-p} (O \otimes O) \\ = \varphi_N\left(\frac{1}{2}(T \otimes \operatorname{Id} + \operatorname{Id} \otimes T)\right) \left(\operatorname{Id} + \ell_\alpha^{-2}(T \otimes \operatorname{Id} - \operatorname{Id} \otimes T)^2\right)^{-p}. \end{aligned}$$

By the facts on triangular matrices recalled after (5.19), the right side is a product of upper triangular matrices. The diagonal entry of this product at $e_i \otimes e_j$ is

$$\varphi_N\left(\frac{w_i + w_j}{2}\right) f_p\left(\frac{w_i - w_j}{\ell_\alpha}\right).$$

Similarly, $O^* \varphi_N(A) O = \varphi_N(T)$ is upper triangular with diagonal entries $\varphi_N(w_i)$. Taking traces, we obtain

$$\begin{aligned} \mathcal{X}_\alpha(A; \varphi_N, f_p) &= \frac{1}{2} \sum_{i,j=1}^N \varphi_N\left(\frac{w_i + w_j}{2}\right) f_p\left(\frac{w_i - w_j}{\ell_\alpha}\right) - \frac{1}{2} \sum_{i=1}^N \varphi_N(w_i) \\ &= \sum_{1 \leq i < j \leq N} \varphi_N\left(\frac{w_i + w_j}{2}\right) f_p\left(\frac{w_j - w_i}{\ell_\alpha}\right), \end{aligned}$$

where we used that $f_p(0) = 1$, so that the terms with $i = j$ cancel the second sum, and that the summand is symmetric in (i, j) , since $f_p(-z) = f_p(z)$. For a real symmetric A , we may take $w_i = \lambda_i(A)$, and (5.24) becomes (5.1).

To prove part (2), we first bound the argument of f_p . Let $z \in \mathbb{C}$ with $|\operatorname{Im} z| \leq 1/2$. Then

$$\begin{aligned} \operatorname{Re}(1 + z^2) &= 1 + (\operatorname{Re} z)^2 - (\operatorname{Im} z)^2 \geq \frac{3}{4}(1 + (\operatorname{Re} z)^2) > 0, \\ |\operatorname{Im}(1 + z^2)| &= 2|\operatorname{Re} z| |\operatorname{Im} z| \leq (1 + (\operatorname{Re} z)^2) |\operatorname{Im} z|. \end{aligned} \tag{5.25}$$

The real part of $1 + z^2$ is positive, so (5.25) and the inequality $|\arctan y| \leq |y|$ give

$$|\arg(1 + z^2)| \leq \frac{|\operatorname{Im}(1 + z^2)|}{\operatorname{Re}(1 + z^2)} \leq \frac{4}{3} |\operatorname{Im} z|. \quad (5.26)$$

For such z with the additional property $(4/3)p|\operatorname{Im} z| < \pi$, it follows that $\arg f_p(z) = -p \arg(1 + z^2)$, and therefore

$$|\arg f_p(z)| \leq \frac{4}{3} p |\operatorname{Im} z|. \quad (5.27)$$

Now let A be as in part (2), fix $i < j$, and set $z = (w_j - w_i)/\ell_\alpha$. The assumption $A \in \mathcal{U}_{\alpha,p}$ gives

$$|\operatorname{Im} z| < 2\omega_p \leq \frac{1}{1000},$$

and in particular $(4/3)p|\operatorname{Im} z| < \pi$. By (5.27), we have

$$|\arg f_p(z)| \leq \frac{8p\omega_p}{3} = \frac{8p}{3000(p+1)} < \frac{1}{375}.$$

For the midpoint $(w_i + w_j)/2$, we have

$$\left| \operatorname{Re} \frac{w_i + w_j}{2} \right| \leq \max(|\operatorname{Re} w_i|, |\operatorname{Re} w_j|), \quad \left| \operatorname{Im} \frac{w_i + w_j}{2} \right| \leq \max(|\operatorname{Im} w_i|, |\operatorname{Im} w_j|),$$

so that our assumption gives

$$\left(\left| \operatorname{Re} \frac{w_i + w_j}{2} \right| + 2 \right) \left| \operatorname{Im} \frac{w_i + w_j}{2} \right| \leq \max_{w, w' \in \{w_i, w_j\}} (|\operatorname{Re} w| + 2) |\operatorname{Im} w'| \leq \frac{h_N^2}{200}.$$

Applying part (3) of Lemma 5.1, we find that either $\varphi_N((w_i + w_j)/2) = 0$ or $|\arg \varphi_N((w_i + w_j)/2)| \leq 1/200$. In the latter case, we have

$$\left| \arg \left(\varphi_N \left(\frac{w_i + w_j}{2} \right) f_p(z) \right) \right| \leq \left| \arg \varphi_N \left(\frac{w_i + w_j}{2} \right) \right| + |\arg f_p(z)| \leq \frac{1}{200} + \frac{1}{375} < \frac{1}{100},$$

where the first inequality holds because the middle expression is smaller than π . Every nonzero complex number whose argument has absolute value at most $1/100$ has positive real part, so the real part of the sum in (5.24) is nonnegative.

For part (3), let $\operatorname{Spec} A \subset \mathfrak{R}_{\alpha,p}$. Then $A \in \mathcal{U}_{\alpha,p}$, and all eigenvalues w and w' of A satisfy

$$(|\operatorname{Re} w| + 2) |\operatorname{Im} w'| < 6\omega_p \ell_\alpha \leq \frac{6}{2000} N^{-1} \leq \frac{N^{-1/5}}{200} = \frac{h_N^2}{200}.$$

It remains to prove part (4). The function F is holomorphic on $\mathcal{U}_{\alpha,p}$ by part (1), since $\mathcal{U}_{\alpha,p} \subset \mathcal{U}_{\alpha,p_j}$ for all $j \in \mathcal{J}$. Let $A \in \mathcal{U}_{\alpha,p}$ satisfy the hypothesis of part (2). Applying part

(2) with p_j and φ_j in place of p and φ , we find that $\operatorname{Re} \mathcal{X}_\alpha(A; \varphi_{j,N}, f_{p_j}) \geq 0$ for every $j \in \mathcal{J}$. Using $\theta_j \geq 0$, we obtain

$$|F(A)| = \exp \left(- \sum_{j \in \mathcal{J}} \theta_j \operatorname{Re} \mathcal{X}_\alpha(A; \varphi_{j,N}, f_{p_j}) \right) \leq 1.$$

The last claim follows from part (3) and, for real symmetric A , from the fact that $\mathcal{X}_\alpha(A; \varphi_{j,N}, f_{p_j}) \geq 0$, which holds by (5.1), part (1) of Lemma 5.1, and the positivity of f_{p_j} on $\mathbb{R}$. $\square$

For a function $\mathcal{Q}$ defined on an open set of $N \times N$ matrices containing the real symmetric matrices, a real symmetric matrix B , indices $a \leq b$, and an integer $n \geq 1$, we write

$$\partial_{ab}^n \mathcal{Q}(B) = \frac{d^n}{dx^n} \mathcal{Q}(B + xV^{ab}) \Big|_{x=0} \quad (5.28)$$

whenever this derivative exists, and $\partial_{ab} = \partial_{ab}^1$. In the real coordinates x_{ab} obtained by writing a real symmetric matrix as $\sum_{a \leq b} x_{ab} V^{ab}$, the operator ∂_{ab} is the partial derivative with respect to x_{ab} . The next lemma gives a crude bound on the derivatives of F that holds at every real symmetric matrix, without any regularity assumption. We will use it to verify the hypotheses of the derivative formula in Lemma 5.7 below and to control contributions of events of very small probability.

Lemma 5.6. *Let F be as in Definition 5.4, and let $p = \max_{j \in \mathcal{J}} p_j$. For every real symmetric matrix B and all $a \leq b$, the function $x \mapsto F(B + xV^{ab})$ is smooth on $\mathbb{R}$. Moreover, for every integer $n \geq 1$, we have*

$$|\partial_{ab}^n F(B)| \leq n! \left(\frac{2}{\omega_p} \right)^n \left(\frac{1 + \|B\|}{\ell_\alpha h_N^2} \right)^n. \quad (5.29)$$

Proof. Fix B and $a \leq b$, and set

$$r_B = \frac{\omega_p \ell_\alpha h_N^2}{2(1 + \|B\|)}.$$

We first locate the eigenvalues of $B + \zeta V^{ab}$ for $\zeta \in \mathbb{C}$. Let $w \in \mathbb{C}$ satisfy $\operatorname{dist}(w, \operatorname{Spec} B) > |\zeta|$. Since B is real symmetric and $\|V^{ab}\| = 1$, we have

$$\|\zeta(B - w)^{-1} V^{ab}\| \leq |\zeta| \operatorname{dist}(w, \operatorname{Spec} B)^{-1} < 1,$$

so the matrix

$$B + \zeta V^{ab} - w = (B - w)(\operatorname{Id} + \zeta(B - w)^{-1} V^{ab})$$

is invertible. Hence every eigenvalue w of $B + \zeta V^{ab}$ satisfies $\operatorname{dist}(w, \operatorname{Spec} B) \leq |\zeta|$, and in particular

$$|\operatorname{Im} w| \leq |\zeta|, \quad |\operatorname{Re} w| \leq \|B\| + |\zeta|. \quad (5.30)$$

The bound (5.30) shows that $B + \zeta V^{ab} \in \mathcal{U}_{\alpha,p}$ whenever $|\zeta| < \omega_p \ell_\alpha$. Hence, by part (4) of Lemma 5.5, the function $\zeta \mapsto F(B + \zeta V^{ab})$ is holomorphic on the disk $\{|\zeta| < \omega_p \ell_\alpha\}$. Applying this with $B + x_0 V^{ab}$ in place of B for each $x_0 \in \mathbb{R}$, we see that $x \mapsto F(B + x V^{ab})$ is real analytic, and in particular smooth, on $\mathbb{R}$.

Now let $|\zeta| \leq r_B$, and let w and w' be eigenvalues of $B + \zeta V^{ab}$. Using (5.30), $r_B \leq 1$, and $r_B < \omega_p \ell_\alpha$, we find that $B + \zeta V^{ab} \in \mathcal{U}_{\alpha,p}$ and that

$$(|\operatorname{Re} w| + 2)|\operatorname{Im} w'| \leq (\|B\| + 3)r_B \leq 3(1 + \|B\|)r_B = \frac{3}{2}\omega_p \ell_\alpha h_N^2 \leq \frac{h_N^2}{200}.$$

Hence $B + \zeta V^{ab}$ satisfies the hypothesis of part (2) of Lemma 5.5, and part (4) of that lemma gives $|F(B + \zeta V^{ab})| \leq 1$ for $|\zeta| \leq r_B$. Cauchy's estimate on the circle of radius r_B about the origin gives

$$|\partial_{ab}^n F(B)| \leq n! r_B^{-n},$$

which is (5.29). □

5.4 Comparison along the Ornstein–Uhlenbeck interpolation

For $a \leq b$ and $t \in [0, 1]$, we let $x_{ab}(t) = (H_t)_{ab}$ denote the (a, b) coordinate of H_t , so that

$$H_t = \sum_{a \leq b} x_{ab}(t) V^{ab}, \quad x_{ab}(t) = e^{-t/2} H_{ab} + (1 - e^{-t})^{1/2} G_{ab}, \quad (5.31)$$

and we let $H_t^{[ab]} = H_t - x_{ab}(t) V^{ab}$ denote the matrix obtained from H_t by setting its (a, b) and (b, a) entries to zero. We also set $v_{ab}^G = \mathbb{E} G_{ab}^2$, so that $v_{ab}^G = 1/N$ for $a < b$ and $v_{aa}^G = 2/N$. The variables $(H_{ab})_{a \leq b}$ and $(G_{ab})_{a \leq b}$ are independent, so for each t , the coordinates $(x_{ab}(t))_{a \leq b}$ are independent. In particular, $x_{ab}(t)$ is independent of $H_t^{[ab]}$.

We record some properties of the coordinates. Recall from Definition 1.1 that $H_{ab} = N^{-1/2} X_{ab}$ for $a < b$ and $H_{aa} = N^{-1/2} Z_a$, where $\mathbb{P}(|X| > y) + \mathbb{P}(|Z| > y) \leq C_0 e^{-c_0 y^2}$ for $y \geq 0$. Since G_{ab} is a centered Gaussian variable with variance at most $2/N$, we have, for all $t \in [0, 1]$, $a \leq b$, and $y \geq 0$,

$$\mathbb{P}(|x_{ab}(t)| > y) \leq \mathbb{P}(|H_{ab}| > y/2) + \mathbb{P}(|G_{ab}| > y/2) \leq C_0 e^{-c_0 N y^2/4} + 2e^{-N y^2/16}. \quad (5.32)$$

In particular, for every integer $n \geq 1$, there exists $C_n > 0$ such that

$$\mathbb{E}|x_{ab}(t)|^n \leq C_n N^{-n/2} \quad (5.33)$$

for all $t \in [0, 1]$ and $a \leq b$. Using the independence of H_{ab} and G_{ab} , the identities $\mathbb{E} H_{ab} = \mathbb{E} G_{ab} = \mathbb{E} G_{ab}^3 = 0$, $\mathbb{E} H_{ab}^2 = 1/N$ for $a < b$, and $\mathbb{E} H_{aa}^2 = \sigma_Z^2/N$, we also find

$$\mathbb{E} x_{ab}(t)^2 = v_{ab}^G, \quad \mathbb{E} x_{ab}(t)^3 = \frac{e^{-3t/2} \mathbb{E} X^3}{N^{3/2}} \quad (a < b), \quad \mathbb{E} x_{aa}(t)^2 = v_{aa}^G + \frac{e^{-t}(\sigma_Z^2 - 2)}{N}. \quad (5.34)$$

The following lemma expresses $\mathbb{E}Q(H_t) - \mathbb{E}Q(H)$ in terms of the derivatives of Q in single coordinates.

Lemma 5.7. *Let Q be a bounded, twice continuously differentiable, complex-valued function on the space of real symmetric $N \times N$ matrices. Suppose that there exists $K_0 > 0$ such that*

$$|\partial_{ab}Q(A)| + |\partial_{ab}^2Q(A)| \leq K_0(1 + \|A\|)^2 \quad (5.35)$$

for every real symmetric matrix A and all $a \leq b$. Then for all $t \in [0, 1]$,

$$\mathbb{E}Q(H_t) - \mathbb{E}Q(H) = \frac{1}{2} \int_0^t \sum_{a \leq b} \mathbb{E} \left[v_{ab}^G \partial_{ab}^2 Q(H_{t'}) - x_{ab}(t') \partial_{ab} Q(H_{t'}) \right] dt'. \quad (5.36)$$

Proof. Fix realizations of H and G . The map $t' \mapsto H_{t'}$ is continuous on $[0, 1]$ and continuously differentiable on $(0, 1]$. Differentiating (5.31) and using the identity

$$e^{-t'}(1 - e^{-t'})^{-1/2} = (1 - e^{-t'})^{-1/2} - (1 - e^{-t'})^{1/2},$$

we find that

$$\frac{d}{dt'} x_{ab}(t') = -\frac{1}{2} x_{ab}(t') + \frac{1}{2} (1 - e^{-t'})^{-1/2} G_{ab}, \quad t' \in (0, 1].$$

The chain rule gives, for $t' \in (0, 1]$,

$$\frac{d}{dt'} Q(H_{t'}) = \sum_{a \leq b} \partial_{ab} Q(H_{t'}) \left(-\frac{1}{2} x_{ab}(t') + \frac{1}{2} (1 - e^{-t'})^{-1/2} G_{ab} \right). \quad (5.37)$$

We now justify integrating (5.37). For $t' \in (0, 1]$, we have

$$\|H_{t'}\| \leq \|H\| + \|G\|, \quad |x_{ab}(t')| \leq |H_{ab}| + |G_{ab}|, \quad 1 - e^{-t'} \geq t'/e.$$

Together with (5.35), these bounds give

$$\left| \frac{d}{dt'} Q(H_{t'}) \right| \leq K_0(1 + \|H\| + \|G\|)^2 (1 + (e/t')^{1/2}) \sum_{a \leq b} (|H_{ab}| + |G_{ab}|). \quad (5.38)$$

The entries of H and G have moments of all orders, and $\|H\|^2 \leq \sum_{a,b} H_{ab}^2$. Therefore the right side of (5.38) has finite expectation for every $t' \in (0, 1]$, and this expectation is integrable in t' over $(0, 1]$ (the factor $(t')^{-1/2}$ is integrable at $t' = 0$). Applying the fundamental theorem of calculus on $[t_0, t]$ with $0 < t_0 < t$ and letting $t_0 \rightarrow 0$, using the continuity of $t' \mapsto Q(H_{t'})$ at $t' = 0$, we obtain

$$Q(H_t) - Q(H) = \int_0^t \frac{d}{dt'} Q(H_{t'}) dt', \quad t \in [0, 1].$$

Taking expectations and using Fubini's theorem, we find

$$\mathbb{E}Q(H_t) - \mathbb{E}Q(H) = \int_0^t \mathbb{E} \left[\frac{d}{dt'} Q(H_{t'}) \right] dt'. \quad (5.39)$$

Next, we compute the expectation of (5.37) with Gaussian integration by parts. Fix $t' \in (0, 1]$ and $a \leq b$, and condition on H and on $(G_{a'b'})_{(a',b') \neq (a,b)}$. Then $H_{t'}^{[ab]}$ and H_{ab} are fixed, and

$$H_{t'} = H_{t'}^{[ab]} + (e^{-t'/2} H_{ab} + (1 - e^{-t'})^{1/2} G_{ab}) V^{ab}.$$

Hence $\partial_{ab} Q(H_{t'})$ is a continuously differentiable function of G_{ab} , with

$$\frac{\partial}{\partial G_{ab}} \partial_{ab} Q(H_{t'}) = (1 - e^{-t'})^{1/2} \partial_{ab}^2 Q(H_{t'}).$$

Moreover, (5.35) implies that this function and its derivative grow at most polynomially in G_{ab} .

The variable G_{ab} is centered Gaussian with variance v_{ab}^G and independent of the conditioning variables. We use the standard Gaussian integration by parts formula

$$\mathbb{E}[G_{ab} \psi(G_{ab})] = v_{ab}^G \mathbb{E}[\psi'(G_{ab})],$$

valid for every continuously differentiable $\psi: \mathbb{R} \rightarrow \mathbb{C}$ such that ψ and ψ' grow at most polynomially. Applying it conditionally, and integrating over the conditioning variables, we obtain

$$\mathbb{E}[G_{ab} \partial_{ab} Q(H_{t'})] = (1 - e^{-t'})^{1/2} v_{ab}^G \mathbb{E}[\partial_{ab}^2 Q(H_{t'})]. \quad (5.40)$$

Taking expectations in (5.37) and inserting (5.40), we see that the integrands on the right sides of (5.39) and (5.36) coincide, which proves (5.36). $\square$

We now explain how the integrand in (5.36) will be estimated for $Q = F$. Each summand has the form

$$\mathbb{E}[v_{ab}^G Y''(x_{ab}(t')) - x_{ab}(t') Y'(x_{ab}(t'))]$$

for a random function Y independent of $x_{ab}(t')$, namely the restriction of F to the line through $H_{t'}^{[ab]}$ in the direction V^{ab} . When $v_{ab}^G = \mathbb{E}x_{ab}(t')^2$, a Taylor expansion of Y' and Y'' around 0 shows that the terms of order zero and one cancel, and what remains is of order $\mathbb{E}|x_{ab}(t')|^3$ times the third derivative of Y . The next lemma, a truncated form of the cumulant expansion of [44] (see also [54, Proposition 3.1]), makes this precise. We state it with a cutoff, since good bounds on the derivatives of Y are available only on a bounded interval.

Lemma 5.8. *Let $\varkappa$ be a real random variable such that $\mathbb{E}\varkappa = 0$, $\mathbb{E}\varkappa^2 \leq 1$, and $\mathbb{E}\varkappa^4 < \infty$. Let $r \in (0, 1]$ and $K_0 \geq 0$, and let $Y: \mathbb{R} \rightarrow \mathbb{C}$ be four times continuously differentiable with*

$$|Y^{(n)}(x)| \leq n! r^{-n} \quad \text{for } |x| \leq r \text{ and } n \in \{1, 2, 3, 4\}, \quad (5.41)$$

$$|Y'(x)| + |Y''(x)| \leq K_0(1 + |x|)^2 \quad \text{for } x \in \mathbb{R}. \quad (5.42)$$

Set $\mathfrak{t} = (9r^{-3} + 2K_0)\mathbb{E}[(1 + |\varkappa|)^4 \mathbf{1}_{|\varkappa| > r}]$. Then

$$|\mathbb{E}[\varkappa Y'(\varkappa)] - \mathbb{E}[\varkappa^2] \mathbb{E}[Y''(\varkappa)]| \leq 9r^{-3} \mathbb{E}|\varkappa|^3 + \mathfrak{t}. \quad (5.43)$$

If moreover $\mathbb{E}\varkappa^3 = 0$, then

$$|\mathbb{E}[\varkappa Y'(\varkappa)] - \mathbb{E}[\varkappa^2] \mathbb{E}[Y''(\varkappa)]| \leq 16r^{-4} \mathbb{E}\varkappa^4 + \mathfrak{t}. \quad (5.44)$$

Finally, we have

$$|\mathbb{E}Y''(\varkappa)| \leq 2r^{-2} + \mathfrak{t}. \quad (5.45)$$

Proof. For notational convenience, we abbreviate $\mathbf{1}_- = \mathbf{1}_{|\varkappa| \leq r}$, $\mathbf{1}_+ = \mathbf{1}_{|\varkappa| > r}$, and $v = \mathbb{E}\varkappa^2$. We first prove (5.43). Taylor's theorem with integral remainder and (5.41) with $n = 3$ give, on the event $\{|\varkappa| \leq r\}$,

$$\begin{aligned} Y'(\varkappa) &= Y'(0) + \varkappa Y''(0) + \rho_1, & Y''(\varkappa) &= Y''(0) + \rho_2, \\ |\rho_1| &\leq 3r^{-3} \varkappa^2, & |\rho_2| &\leq 6r^{-3} |\varkappa|. \end{aligned}$$

Therefore

$$\mathbb{E}[(\varkappa Y'(\varkappa) - v Y''(\varkappa)) \mathbf{1}_-] = Y'(0) \mathbb{E}[\varkappa \mathbf{1}_-] + Y''(0) \mathbb{E}[(\varkappa^2 - v) \mathbf{1}_-] + \mathbb{E}[(\varkappa \rho_1 - v \rho_2) \mathbf{1}_-]. \quad (5.46)$$

The identities $\mathbb{E}\varkappa = 0$ and $\mathbb{E}[\varkappa^2 - v] = 0$ give

$$\mathbb{E}[\varkappa \mathbf{1}_-] = -\mathbb{E}[\varkappa \mathbf{1}_+], \quad \mathbb{E}[(\varkappa^2 - v) \mathbf{1}_-] = -\mathbb{E}[(\varkappa^2 - v) \mathbf{1}_+].$$

Using $|Y'(0)| \leq r^{-1}$, $|Y''(0)| \leq 2r^{-2}$, $v \leq 1$, and $r \leq 1$, we conclude that the first two terms on the right side of (5.46) are bounded in absolute value by

$$r^{-1} \mathbb{E}[|\varkappa| \mathbf{1}_+] + 2r^{-2} \mathbb{E}[(\varkappa^2 + 1) \mathbf{1}_+] \leq 3r^{-3} \mathbb{E}[(1 + |\varkappa|)^4 \mathbf{1}_+].$$

Using Hölder's inequality in the form $v \mathbb{E}|\varkappa| \leq \mathbb{E}|\varkappa|^3$, we find that the last term in (5.46) satisfies

$$|\mathbb{E}[(\varkappa \rho_1 - v \rho_2) \mathbf{1}_-]| \leq 3r^{-3} \mathbb{E}|\varkappa|^3 + 6r^{-3} v \mathbb{E}|\varkappa| \leq 9r^{-3} \mathbb{E}|\varkappa|^3.$$

To bound the contribution of the event $\{|\varkappa| > r\}$, we use (5.42), which gives

$$|\mathbb{E}[(\varkappa Y'(\varkappa) - v Y''(\varkappa)) \mathbf{1}_+]| \leq K_0 \mathbb{E}[(|\varkappa| + 1)(1 + |\varkappa|)^2 \mathbf{1}_+] \leq K_0 \mathbb{E}[(1 + |\varkappa|)^4 \mathbf{1}_+].$$

Adding the contributions of the two events, we obtain (5.43).

When $\mathbb{E}\mathcal{X}^3 = 0$, we expand one order further. On the event $\{|\mathcal{X}| \leq r\}$, (5.41) with $n = 4$ gives

$$Y'(\mathcal{X}) = Y'(0) + \mathcal{X}Y''(0) + \frac{\mathcal{X}^2}{2}Y'''(0) + \rho_3, \quad Y''(\mathcal{X}) = Y''(0) + \mathcal{X}Y'''(0) + \rho_4, \\ |\rho_3| \leq 4r^{-4}|\mathcal{X}|^3, \quad |\rho_4| \leq 12r^{-4}\mathcal{X}^2.$$

Compared with (5.46), there is an additional term involving $Y'''(0)$. We have $\mathbb{E}[\mathcal{X}^3/2 - v\mathcal{X}] = 0$ and $|Y'''(0)| \leq 6r^{-3}$, so this term satisfies

$$|Y'''(0)\mathbb{E}[(\mathcal{X}^3/2 - v\mathcal{X})\mathbf{1}_-]| = |Y'''(0)\mathbb{E}[(\mathcal{X}^3/2 - v\mathcal{X})\mathbf{1}_+]| \leq 6r^{-3}\mathbb{E}[(1 + |\mathcal{X}|)^4\mathbf{1}_+].$$

The remainder term is now bounded by

$$|\mathbb{E}[(\mathcal{X}\rho_3 - v\rho_4)\mathbf{1}_-]| \leq 4r^{-4}\mathbb{E}\mathcal{X}^4 + 12r^{-4}v^2 \leq 16r^{-4}\mathbb{E}\mathcal{X}^4,$$

where we used $v^2 \leq \mathbb{E}\mathcal{X}^4$. The other terms are bounded as before, and the tail contributions add up to at most

$$(3r^{-3} + 6r^{-3} + K_0)\mathbb{E}[(1 + |\mathcal{X}|)^4\mathbf{1}_+] \leq \mathfrak{t}.$$

This gives (5.44).

Finally, (5.41) with $n = 2$ and (5.42) give

$$|\mathbb{E}Y''(\mathcal{X})| \leq 2r^{-2} + K_0\mathbb{E}[(1 + |\mathcal{X}|)^2\mathbf{1}_+] \leq 2r^{-2} + \mathfrak{t},$$

which is (5.45). □

Proposition 5.9. *Let $\varepsilon > 0$ satisfy $\alpha + 5\varepsilon < 1/2$. Let F be as in Definition 5.4, with $p = \max_{j \in \mathcal{J}} p_j$, and set $r_N = c_p N^{-\alpha-4\varepsilon}$. There exists $C > 0$, and for every $D > 0$ there exists $C_D > 0$, such that for all $N \geq 2$ and $t \in [0, 1]$,*

$$|\mathbb{E}F(H_t) - \mathbb{E}F(H)| \leq Ct(N^{1/2}r_N^{-3} + r_N^{-2}) + C_D N^{-D}. \quad (5.47)$$

If $\mathbb{E}X^3 = 0$, then (5.47) holds with $N^{1/2}r_N^{-3}$ replaced by r_N^{-4} . The constant C depends only on α, ε, p , and the laws of X and Z , and C_D depends in addition on D . In particular, they do not depend on $\mathcal{J}$, on the weights φ_j , or on the parameters θ_j .

The proof has four steps. We first express $\mathbb{E}F(H_t) - \mathbb{E}F(H)$ through the derivative formula of Lemma 5.7. We then show that for every time $t' \in (0, 1]$, outside an event of very small probability, the function $\zeta \mapsto F(H_{t'}^{[ab]} + \zeta V^{ab})$ is bounded by one on a complex disk of radius $4r_N$, so that Cauchy's estimates bound its derivatives on $[-r_N, r_N]$. This is the main step, which rests on Lemma 5.2. Next, we Taylor expand in the coordinate $x_{ab}(t')$ using Lemma 5.8. Finally, we sum over the coordinates and integrate in time.

Proof. We abbreviate $\mathfrak{R} = \mathfrak{R}_{\alpha,p}$. Let $N_0 \geq 1$ be so large that the conclusion of Lemma 5.2 holds for $N \geq N_0$, and such that $N^{-1/2} \leq r_N$ and $\max(1, \sigma_Z^2, 2) \leq N$ for $N \geq N_0$; such an N_0 exists since $\alpha + 4\varepsilon < 1/2$. For $N < N_0$, the left side of (5.47) is bounded by $2 \leq 2N_0^D N^{-D}$, so we may assume that $N \geq N_0$. Our assumption $\alpha + 5\varepsilon < 1/2$ gives

$$Nr_N^2 = c_p^2 N^{1-2\alpha-8\varepsilon} \geq c_p^2 N^{2\varepsilon}, \quad \ell_\alpha h_N^2 = N^{-6/5-\alpha} \geq N^{-2}. \quad (5.48)$$

Step 1. The derivative formula. By part (4) of Lemma 5.5, the function F is bounded by one on real symmetric matrices. Further, F is holomorphic on the open set $\mathcal{U}_{\alpha,p}$, so the function $(x_{ab})_{a \leq b} \mapsto F(\sum_{a \leq b} x_{ab} V^{ab})$ extends holomorphically to an open subset of $\mathbb{C}^{N(N+1)/2}$ containing $\mathbb{R}^{N(N+1)/2}$. Hence F is real analytic, and in particular twice continuously differentiable, on the space of real symmetric matrices. Combining Lemma 5.6 with (5.48), we find that for every real symmetric A and all $a \leq b$,

$$|\partial_{ab} F(A)| \leq \frac{2}{\omega_p} N^2 (1 + \|A\|), \quad |\partial_{ab}^2 F(A)| \leq \frac{8}{\omega_p^2} N^4 (1 + \|A\|)^2. \quad (5.49)$$

Hence Lemma 5.7 applies with $\mathcal{Q} = F$. For $t' \in [0, 1]$ and $a \leq b$, we set

$$\mathcal{Q}_{ab}(t') = \mathbb{E} \left[v_{ab}^G \partial_{ab}^2 F(H_{t'}) - x_{ab}(t') \partial_{ab} F(H_{t'}) \right],$$

and (5.36) reads

$$\mathbb{E} F(H_t) - \mathbb{E} F(H) = \frac{1}{2} \int_0^t \sum_{a \leq b} \mathcal{Q}_{ab}(t') dt'. \quad (5.50)$$

Step 2. A holomorphic bound on a disk. Fix $t' \in (0, 1]$ and $a \leq b$. We define $\mathcal{E}_{ab}(t')$ to be the event that

$$\text{Spec}(H_{t'}^{[ab]} + \zeta V^{ab}) \subset \mathfrak{R} \quad \text{for all } \zeta \in \mathbb{C} \text{ with } |\zeta| \leq 4r_N. \quad (5.51)$$

The set of real symmetric matrices B such that $\text{Spec}(B + \zeta V^{ab}) \subset \mathfrak{R}$ for all $|\zeta| \leq 4r_N$ is open. Indeed, if $B_n \rightarrow B$ and $B_n + \zeta_n V^{ab}$ has an eigenvalue outside $\mathfrak{R}$ for some $|\zeta_n| \leq 4r_N$, then for a limit point ζ of (ζ_n) , the matrix $B + \zeta V^{ab}$ has an eigenvalue outside $\mathfrak{R}$ by Lemma 5.3 with $\mathcal{O} = \mathfrak{R}$. Hence $\mathcal{E}_{ab}(t')$ is an event determined by $H_{t'}^{[ab]}$.

Next, we bound the probability of the complement of $\mathcal{E}_{ab}(t')$. Suppose that $H_{t'} \in \mathcal{G}_\varepsilon$ and $|x_{ab}(t')| \leq r_N$. For $|\zeta| \leq 4r_N$, we have

$$H_{t'}^{[ab]} + \zeta V^{ab} = H_{t'} + (\zeta - x_{ab}(t')) V^{ab}, \quad |\zeta - x_{ab}(t')| \leq 5r_N \leq 8c_p N^{-\alpha-4\varepsilon},$$

so Lemma 5.2 shows that $\text{Spec}(H_{t'}^{[ab]} + \zeta V^{ab}) \subset \mathfrak{R}$. Therefore

$$\{H_{t'} \in \mathcal{G}_\varepsilon\} \cap \{|x_{ab}(t')| \leq r_N\} \subset \mathcal{E}_{ab}(t').$$

Using Proposition 3.2, (5.32), and (5.48), we obtain, for every $D > 0$,

$$\mathbb{P}(\mathcal{E}_{ab}(t')^c) \leq \mathbb{P}(H_{t'} \notin \mathcal{G}_\varepsilon) + \mathbb{P}(|x_{ab}(t')| > r_N) \leq C_D N^{-D}, \quad (5.52)$$

uniformly in $t' \in (0, 1]$ and $a \leq b$.

We now turn to the derivatives in the coordinate $x_{ab}(t')$. We define the random function

$$Y(x) = F(H_{t'}^{[ab]} + xV^{ab}), \quad x \in \mathbb{R}, \quad (5.53)$$

which is determined by $H_{t'}^{[ab]}$ and smooth by Lemma 5.6. Observe that (5.28) gives

$$\partial_{ab} F(H_{t'}) = Y'(x_{ab}(t')), \quad \partial_{ab}^2 F(H_{t'}) = Y''(x_{ab}(t')).$$

On the event $\mathcal{E}_{ab}(t')$, the set of $\zeta \in \mathbb{C}$ such that $H_{t'}^{[ab]} + \zeta V^{ab} \in \mathcal{U}_{\alpha,p}$ is open and contains the closed disk $\{|\zeta| \leq 4r_N\}$. Part (4) of Lemma 5.5 shows that $\zeta \mapsto F(H_{t'}^{[ab]} + \zeta V^{ab})$ is holomorphic on this open set and bounded by one in modulus on the closed disk. For real x with $|x| \leq r_N$, the closed disk of radius $3r_N$ about x is contained in $\{|\zeta| \leq 4r_N\}$, and Cauchy's estimates give

$$|Y^{(n)}(x)| \leq n! (3r_N)^{-n} \leq n! r_N^{-n} \quad \text{for } |x| \leq r_N \text{ and } n \geq 1, \text{ on the event } \mathcal{E}_{ab}(t'). \quad (5.54)$$

Without restriction to $\mathcal{E}_{ab}(t')$, the bound (5.49) gives

$$|Y'(x)| + |Y''(x)| \leq \frac{10}{\omega_p^2} N^4 (1 + \|H_{t'}^{[ab]}\|)^2 (1 + |x|)^2, \quad x \in \mathbb{R}, \quad (5.55)$$

where we used $\omega_p \leq 1$ and

$$1 + \|H_{t'}^{[ab]} + xV^{ab}\| \leq (1 + \|H_{t'}^{[ab]}\|)(1 + |x|).$$

Step 3. Taylor expansion in one coordinate. We keep t' and $a \leq b$ fixed and abbreviate $\varkappa = x_{ab}(t')$. The event $\mathcal{E}_{ab}(t')$ is determined by $H_{t'}^{[ab]}$, so we may split

$$\begin{aligned} \mathcal{Q}_{ab}(t') &= \mathbb{E} \left[\mathbf{1}_{\mathcal{E}_{ab}(t')} \mathbb{E} [v_{ab}^G Y''(\varkappa) - \varkappa Y'(\varkappa) \mid H_{t'}^{[ab]}] \right] \\ &\quad + \mathbb{E} \left[\mathbf{1}_{\mathcal{E}_{ab}(t')^c} (v_{ab}^G Y''(\varkappa) - \varkappa Y'(\varkappa)) \right]. \end{aligned} \quad (5.56)$$

We begin by bounding the second term. Using (5.49), we have

$$|v_{ab}^G Y''(\varkappa) - \varkappa Y'(\varkappa)| \leq CN^4 (1 + \|H_{t'}\|)^2 (N^{-1} + |\varkappa|).$$

Combining this bound with the Cauchy–Schwarz inequality, Proposition 3.2, (5.33), and (5.52), we find that the second term in (5.56) is bounded in absolute value by

$$CN^4 \left(\mathbb{E} [(1 + \|H_{t'}\|)^4 (N^{-1} + |\varkappa|)^2] \right)^{1/2} \mathbb{P}(\mathcal{E}_{ab}(t')^c)^{1/2} \leq C_D N^{-D}.$$

To bound the first term in (5.56), we apply Lemma 5.8 conditionally on $H_{t'}^{[ab]}$, with Y held fixed; this is possible since $\varkappa$ is independent of $H_{t'}^{[ab]}$. We have $\mathbb{E}\varkappa = 0$, and (5.34) and our choice of N_0 give

$$\mathbb{E}\varkappa^2 \leq \max(1, \sigma_Z^2, 2)N^{-1} \leq 1.$$

On the event $\mathcal{E}_{ab}(t')$, the function Y satisfies (5.41) with $r = r_N \leq 1$ by (5.54). It also satisfies (5.42) with

$$K_0 = \frac{10}{\omega_p^2} N^4 (1 + \|H_{t'}^{[ab]}\|)^2$$

by (5.55). By the Cauchy–Schwarz inequality, (5.32), (5.33), and (5.48), we have

$$\mathbb{E}[(1 + |\varkappa|)^4 \mathbf{1}_{|\varkappa| > r_N}] \leq (\mathbb{E}(1 + |\varkappa|)^8)^{1/2} \mathbb{P}(|\varkappa| > r_N)^{1/2} \leq C \exp(-cN^{2\varepsilon})$$

for some $c > 0$ depending on c_0 and p . Since $r_N^{-3} \leq N^{3/2}$, it follows that the tail term of Lemma 5.8 satisfies

$$\begin{aligned} \mathfrak{t} &= \left(9r_N^{-3} + \frac{20}{\omega_p^2} N^4 (1 + \|H_{t'}^{[ab]}\|)^2\right) \mathbb{E}[(1 + |\varkappa|)^4 \mathbf{1}_{|\varkappa| > r_N}] \\ &\leq C_D N^{-D} (1 + \|H_{t'}^{[ab]}\|)^2. \end{aligned} \tag{5.57}$$

We now decompose the integrand of the first term in (5.56) as

$$v_{ab}^G Y''(\varkappa) - \varkappa Y'(\varkappa) = (v_{ab}^G - \mathbb{E}\varkappa^2) Y''(\varkappa) + (\mathbb{E}[\varkappa^2] Y''(\varkappa) - \varkappa Y'(\varkappa)).$$

We bound the conditional expectations of the two terms on the right using (5.45) and (5.43), respectively. Then we take expectations, using (5.33) and the bound

$$\mathbb{E}(1 + \|H_{t'}^{[ab]}\|)^2 \leq C,$$

which follows from Proposition 3.2 and $\|H_{t'}^{[ab]}\| \leq \|H_{t'}\| + |\varkappa|$. We find that the first term in (5.56) is bounded in absolute value by

$$2|v_{ab}^G - \mathbb{E}\varkappa^2| r_N^{-2} + 9r_N^{-3} \mathbb{E}|\varkappa|^3 + C_D N^{-D} \leq 2|v_{ab}^G - \mathbb{E}\varkappa^2| r_N^{-2} + C N^{-3/2} r_N^{-3} + C_D N^{-D}.$$

When $\mathbb{E}X^3 = 0$ and $a < b$, we have $\mathbb{E}\varkappa^3 = 0$ by (5.34), and (5.44) gives the same bound with $C N^{-3/2} r_N^{-3}$ replaced by

$$16r_N^{-4} \mathbb{E}\varkappa^4 \leq C N^{-2} r_N^{-4}.$$

Recall from (5.34) that

$$v_{ab}^G = \mathbb{E}x_{ab}(t')^2 \quad (a < b), \quad |v_{aa}^G - \mathbb{E}x_{aa}(t')^2| \leq |\sigma_Z^2 - 2|N^{-1}.$$

Combining these facts with the bounds on the two terms in (5.56), we obtain

$$\begin{aligned} |\mathcal{Q}_{ab}(t')| &\leq CN^{-3/2}r_N^{-3} + C_D N^{-D}, \quad a < b, \\ |\mathcal{Q}_{aa}(t')| &\leq CN^{-1}r_N^{-2} + CN^{-3/2}r_N^{-3} + C_D N^{-D}, \end{aligned} \quad (5.58)$$

uniformly in $t' \in (0, 1]$. When $\mathbb{E}X^3 = 0$, the bound for $a < b$ holds with $CN^{-3/2}r_N^{-3}$ replaced by $CN^{-2}r_N^{-4}$.

Step 4. Summation of the generator. There are at most N^2 pairs $a < b$ and N diagonal coordinates. Our choice of N_0 gives $N^{-1/2}r_N^{-3} \leq r_N^{-2}$, so (5.58) implies that for every $t' \in (0, 1]$,

$$\begin{aligned} \sum_{a \leq b} |\mathcal{Q}_{ab}(t')| &\leq CN^{1/2}r_N^{-3} + Cr_N^{-2} + CN^{-1/2}r_N^{-3} + C_D N^{2-D} \\ &\leq C(N^{1/2}r_N^{-3} + r_N^{-2}) + C_D N^{2-D}. \end{aligned}$$

Inserting this bound into (5.50) and replacing D by $D + 2$, we obtain (5.47).

When $\mathbb{E}X^3 = 0$, the off-diagonal coordinates contribute at most Cr_N^{-4} instead of $CN^{1/2}r_N^{-3}$, and the same argument gives (5.47) with $N^{1/2}r_N^{-3}$ replaced by r_N^{-4} . The constants depend only on $\alpha, \varepsilon, p, D$, and the laws of X and Z , through Lemma 5.2, Proposition 3.2, (5.32), and (5.33). This completes the proof. $\square$

Remark 5.10. The proof of Proposition 5.9 uses the specific form of F only through two properties. First, F is bounded by one on real symmetric matrices, four times continuously differentiable, and satisfies

$$|\partial_{ab}F(A)| + |\partial_{ab}^2F(A)| \leq CN^C(1 + \|A\|)^2$$

for some $C > 0$, all real symmetric matrices A , and all $a \leq b$. Second, for all $t' \in [0, 1]$ and $a \leq b$, there exists an event determined by $H_{t'}^{[ab]}$, whose complement has probability at most $C_D N^{-D}$ for every $D > 0$, uniformly in t' and (a, b) , on which $\zeta \mapsto F(H_{t'}^{[ab]} + \zeta V^{ab})$ extends holomorphically to a neighborhood of the closed disk $\{|\zeta| \leq 4r_N\}$ and is bounded by one in modulus there. Every observable with these two properties satisfies (5.47) and its variant for $\mathbb{E}X^3 = 0$, with the same proof.

Pair statistics at the scales $N^{-1-\alpha}$

In this section, we prove Theorem 1.3 and Proposition 4.1. These results follow from a law of large numbers for the pair statistics of H at the scale $\ell_\alpha = N^{-1-\alpha}$, which we prove by induction on α . Recall the scales ℓ_α and $M_\alpha = N^{1-2\alpha} = N^3 \ell_\alpha^2$ from (3.15), the functions f_p from (3.33), the pair statistic $\mathcal{X}_\alpha(A; \varphi, f)$ from (5.1), and the smoothed energy weight φ_N of a nonnegative $\varphi \in C_c^\infty((-2, 2))$ from (5.2), where $h_N = N^{-1/10}$. We also recall that $p_* = 61$.

We use the same notation for ordered sequences. For $x = (x_1, \dots, x_N) \in \mathbb{R}^N$ with $x_1 \leq \dots \leq x_N$, we define $\mathcal{X}_\alpha(x; \varphi, f)$ by the formula (5.1), with $\lambda_i(A)$ replaced by x_i for all $i \in [N]$. Finally, we note that for all integers $p \geq 2$,

$$\int_0^\infty y f_p(y) dy = \frac{1}{2(p-1)}. \quad (6.1)$$

Definition 6.1. Let $\alpha \in (0, 1/2)$. We say that the statement $\mathcal{P}(\alpha)$ holds if for every nonnegative $\varphi \in C_c^\infty((-2, 2))$ and every integer $p \geq p_*$, we have

$$\frac{1}{M_\alpha} \mathcal{X}_\alpha(H; \varphi_N, f_p) \rightarrow \frac{1}{2(p-1)} \int_{-2}^2 \varphi(E) c(E) dE$$

in probability as $N \rightarrow \infty$.

The argument is outlined in Section 2.5. We prove $\mathcal{P}(\alpha)$ for $\alpha \in (0, 1/8)$ in Section 6.1, extend $\mathcal{P}(\alpha)$ to general test functions and unsmoothed energy weights in Section 6.2, show in Section 6.3 that $\mathcal{P}(\alpha)$ implies $\mathcal{P}(\alpha')$ for α' slightly larger than α , and iterate this step in Section 6.4 to deduce Theorem 1.3 and Proposition 4.1.

We will use the following two lemmas, the first in Section 6.1 and Section 6.2, and the second in Section 6.1 and Section 6.3. The first compares pair statistics with smoothed and unsmoothed energy weights.

Lemma 6.2. *Let $\alpha \in (0, 1/2)$, let $\varphi \in C_c^\infty((-2, 2))$ be nonnegative, and let $\delta > 0$ be such that every $E \in \mathbb{R}$ with $\text{dist}(E, \text{supp } \varphi) \leq \delta$ lies in $(-2, 2)$. Let $\chi \in C_c^\infty((-2, 2))$ satisfy $0 \leq \chi \leq 2$, and suppose that $\chi(E) = 2$ for all $E \in \mathbb{R}$ such that $\text{dist}(E, \text{supp } \varphi) \leq \delta$. Then there exist $C > 0$*

and $N_0 \geq 1$ such that for all $N \geq N_0$, all real symmetric $N \times N$ matrices A with $\|A\| \leq 3$, and all measurable $f: [0, \infty) \rightarrow \mathbb{R}$ with $|f| \leq f_{p*}$, we have

$$\begin{aligned} & |\mathcal{X}_\alpha(A; \varphi_N, f) - \mathcal{X}_\alpha(A; \varphi, f)| \\ & \leq Ch_N \min \{ \mathcal{X}_\alpha(A; \chi, f_{p*}), \mathcal{X}_\alpha(A; \chi_N, f_{p*}) \} + CN^2 \exp \left(- \frac{\delta^2}{4h_N^2} \right). \end{aligned} \quad (6.2)$$

Proof. We begin by recalling two consequences of parts (4) and (5) of Lemma 5.1. There exists $C > 0$ such that

$$\|\varphi_N - \varphi\|_\infty + \|\chi_N - \chi\|_\infty \leq Ch_N, \quad \sup_{E \in [-3, 3]: \text{dist}(E, \text{supp } \varphi) \geq \delta} \varphi_N(E) \leq C \exp \left(- \frac{\delta^2}{4h_N^2} \right). \quad (6.3)$$

Let N_0 be such that $Ch_N \leq 1$ for $N \geq N_0$. Then $\chi_N(E) \geq 1$ for all E such that $\text{dist}(E, \text{supp } \varphi) \leq \delta$.

Fix A and f as in the statement. For $i < j$, the midpoint $E = (\lambda_i(A) + \lambda_j(A))/2$ belongs to $[-3, 3]$, since $\|A\| \leq 3$, and we claim that

$$|\varphi_N(E) - \varphi(E)| \leq Ch_N \min \{ \chi(E), \chi_N(E) \} + C \exp \left(- \frac{\delta^2}{4h_N^2} \right). \quad (6.4)$$

When $\text{dist}(E, \text{supp } \varphi) \leq \delta$, this follows from the first bound in (6.3) and the fact that $\chi(E)$ and $\chi_N(E)$ are at least 1. Otherwise, we have $\varphi(E) = 0$, and (6.4) follows from the second bound in (6.3).

Next, we multiply (6.4) by $|f((\lambda_j(A) - \lambda_i(A))/\ell_\alpha)|$, which is bounded by $f_{p*}((\lambda_j(A) - \lambda_i(A))/\ell_\alpha) \leq 1$, and we sum over the at most N^2 pairs $i < j$. A sum of minima is bounded by the minimum of the sums, so we obtain (6.2). $\square$

The second lemma transfers a law of large numbers for the pair statistic from H_t to H .

Lemma 6.3. *Let $\alpha \in (0, 1/2)$, let $p \geq 1$ be an integer, let $\varphi \in C_c^\infty((-2, 2))$ be nonnegative, and let $L \geq 0$. Let $\varepsilon > 0$ satisfy $\alpha + 5\varepsilon < 1/2$, and let $(t_N)_{N \geq 1}$ be a sequence in $[0, 1]$ such that*

$$\lim_{N \rightarrow \infty} t_N N^{1/2+3\alpha+12\varepsilon} = 0.$$

If $M_\alpha^{-1} \mathcal{X}_\alpha(H_{t_N}; \varphi_N, f_p) \rightarrow L$ in probability, then $M_\alpha^{-1} \mathcal{X}_\alpha(H; \varphi_N, f_p) \rightarrow L$ in probability.

Proof. For every real symmetric matrix A , we have $\mathcal{X}_\alpha(A; \varphi_N, f_p) \geq 0$, since $\varphi_N \geq 0$ on $\mathbb{R}$ and $f_p > 0$. Fix $\theta > 0$. By assumption, the random variable $\exp(-\theta M_\alpha^{-1} \mathcal{X}_\alpha(H_{t_N}; \varphi_N, f_p))$ takes values in $(0, 1]$ and converges to $e^{-\theta L}$ in probability, so the bounded convergence theorem gives

$$\lim_{N \rightarrow \infty} \mathbb{E} \exp \left(- \theta M_\alpha^{-1} \mathcal{X}_\alpha(H_{t_N}; \varphi_N, f_p) \right) = e^{-\theta L}.$$

Let $r_N = c_p N^{-\alpha-4\varepsilon}$ be as in Proposition 5.9. Then

$$t_N(N^{1/2}r_N^{-3} + r_N^{-2}) = c_p^{-3}t_N N^{1/2+3\alpha+12\varepsilon} + c_p^{-2}t_N N^{2\alpha+8\varepsilon} \leq (c_p^{-3} + c_p^{-2})t_N N^{1/2+3\alpha+12\varepsilon},$$

which tends to zero. We apply Proposition 5.9 with the index set $\{1\}$, $\varphi_1 = \varphi$, $p_1 = p$, $\theta_1 = \theta/M_\alpha$, $D = 1$, and $t = t_N$. By Lemma 5.5, the observable F of Definition 5.4 is then given by

$$F(A) = \exp(-\theta M_\alpha^{-1} \mathcal{X}_\alpha(A; \varphi_N, f_p))$$

for every real symmetric matrix A , and we deduce that

$$\lim_{N \rightarrow \infty} \mathbb{E} \exp(-\theta M_\alpha^{-1} \mathcal{X}_\alpha(H; \varphi_N, f_p)) = e^{-\theta L} \quad \text{for all } \theta > 0. \quad (6.5)$$

We now apply (6.5) with $\theta = 1$ and $\theta = 2$, and we find

$$\mathbb{E} \left[(e^{-M_\alpha^{-1} \mathcal{X}_\alpha(H; \varphi_N, f_p)} - e^{-L})^2 \right] = \mathbb{E} e^{-2M_\alpha^{-1} \mathcal{X}_\alpha(H; \varphi_N, f_p)} - 2e^{-L} \mathbb{E} e^{-M_\alpha^{-1} \mathcal{X}_\alpha(H; \varphi_N, f_p)} + e^{-2L} \rightarrow 0.$$

Hence $\exp(-M_\alpha^{-1} \mathcal{X}_\alpha(H; \varphi_N, f_p)) \rightarrow e^{-L}$ in probability. The logarithm is continuous on $(0, \infty)$ and $e^{-L} > 0$, so this implies that $M_\alpha^{-1} \mathcal{X}_\alpha(H; \varphi_N, f_p) \rightarrow L$ in probability. $\square$

6.1 The base case

We begin with a deterministic lemma. It shows that the pair statistic changes by a small relative error when the bulk gaps are perturbed by much less than ℓ_α , provided that the points are close to the classical locations γ_i defined in Section 3.

Lemma 6.4. *Let $\alpha \in (0, 1/2)$ and $\varepsilon \in (0, 1/10]$, let $p \geq 1$ be an integer, and let $\varphi \in C_c^\infty((-2, 2))$ be nonnegative. Let $\delta \in (0, 1)$ satisfy $\text{supp } \varphi \subset [-2 + 3\delta, 2 - 3\delta]$, and set*

$$c_* = \int_{-2}^{-2+\delta/4} \rho_{\text{sc}}(E) dE, \quad \mathcal{J} = \{i \in [N] : c_* N \leq i \leq (1 - c_*)N\}.$$

Then for every $\omega > 0$ there exists N_0 such that the following holds for all $N \geq N_0$. Let $x, x' \in [-3, 3]^N$ be nondecreasing sequences such that

- (a) $\#\{i \in [N] : x_i \in [E, E + N^{-1}]\} \leq N^\varepsilon$ for all $E \in \mathbb{R}$;
- (b) $|x_i - \gamma_i| \leq N^{-1+\varepsilon}$ and $|x'_i - \gamma_i| \leq N^{-1+\varepsilon}$ for all $i \in \mathcal{J}$;
- (c) $|(x_{i+1} - x_i) - (x'_{i+1} - x'_i)| \leq N^{-2\varepsilon} \ell_\alpha$ for all i such that $i, i+1 \in \mathcal{J}$.

Then

$$|\mathcal{X}_\alpha(x; \varphi_N, f_p) - \mathcal{X}_\alpha(x'; \varphi_N, f_p)| \leq \omega(\mathcal{X}_\alpha(x'; \varphi_N, f_p) + M_\alpha). \quad (6.6)$$

Proof. We may assume that φ is not identically zero, since otherwise both statistics vanish. Throughout the proof, C denotes a constant depending on $\alpha, \varepsilon, p, \varphi$, and δ , and N_0 may

increase from line to line. Note that $0 < c_* < 1/2$, since $\delta/4 < 2$. Set $i_- = \lceil c_* N \rceil$ and $i_+ = \lfloor (1 - c_*)N \rfloor$. By the definition of the classical locations, we have

$$\int_{-2}^{\gamma_{i_-}} \rho_{\text{sc}}(E) \, dE = \frac{i_-}{N} \leq c_* + \frac{1}{N} < \int_{-2}^{-2+\delta/2} \rho_{\text{sc}}(E) \, dE$$

for $N \geq N_0$, where the last inequality holds because ρ_{sc} is positive on $(-2, 2)$. Using also the symmetry $\rho_{\text{sc}}(-E) = \rho_{\text{sc}}(E)$, we find that for $N \geq N_0$, we have $i_-, i_+ \in \mathcal{J}$ and

$$\gamma_{i_-} \leq -2 + \frac{\delta}{2}, \quad \gamma_{i_+} \geq 2 - \frac{\delta}{2}. \quad (6.7)$$

We first consider the pairs with a label outside $\mathcal{J}$. The following argument applies to x and to x' , and we write it for x . Using hypothesis (b), (6.7), and the monotonicity of x , we find that for $N \geq N_0$,

$$x_i \leq x_{i_-} \leq -2 + \frac{\delta}{2} + N^{-1+\varepsilon} \leq -2 + \delta \quad \text{for } i < c_* N, \quad x_i \geq 2 - \delta \quad \text{for } i > (1 - c_*)N. \quad (6.8)$$

Let $i < j$ be labels such that $i \notin \mathcal{J}$ or $j \notin \mathcal{J}$, and set $E = (x_i + x_j)/2$. Suppose first that $\text{dist}(E, \text{supp } \varphi) \leq \delta$, so that $E \in [-2 + 2\delta, 2 - 2\delta]$. The label i cannot exceed $(1 - c_*)N$, since otherwise $j > (1 - c_*)N$ as well, and (6.8) would give $E \geq 2 - \delta$. There remain the cases $i < c_* N$ and $i \in \mathcal{J}$, and in the second case we have $j > (1 - c_*)N$. Using (6.8) and $E \in [-2 + 2\delta, 2 - 2\delta]$, we find

$$\begin{aligned} x_j &= 2E - x_i \geq -2 + 3\delta \geq x_i + 2\delta \quad \text{if } i < c_* N, \\ x_i &= 2E - x_j \leq 2 - 3\delta \leq x_j - 2\delta \quad \text{if } i \in \mathcal{J}. \end{aligned}$$

Thus $x_j - x_i \geq 2\delta$, and hence

$$\varphi_N(E) f_p\left(\frac{x_j - x_i}{\ell_\alpha}\right) \leq \|\varphi\|_\infty \left(\frac{\ell_\alpha}{2\delta}\right)^{2p} \leq CN^{-2-2\alpha},$$

where we used the bound $\varphi_N \leq \|\varphi\|_\infty$ from part (1) of Lemma 5.1 and the inequality $2p(1 + \alpha) \geq 2 + 2\alpha$.

Suppose next that $\text{dist}(E, \text{supp } \varphi) > \delta$. Then part (4) of Lemma 5.1 gives

$$\varphi_N(E) f_p\left(\frac{x_j - x_i}{\ell_\alpha}\right) \leq \varphi_N(E) \leq C \exp\left(-\frac{\delta^2}{4h_N^2}\right) = C \exp\left(-\frac{\delta^2 N^{1/5}}{4}\right).$$

Summing over the at most N^2 pairs, we obtain

$$\sum_{\substack{1 \leq i < j \leq N \\ \{i, j\} \notin \mathcal{J}}} \varphi_N\left(\frac{x_i + x_j}{2}\right) f_p\left(\frac{x_j - x_i}{\ell_\alpha}\right) \leq CN^{-2\alpha} + CN^2 \exp\left(-\frac{\delta^2 N^{1/5}}{4}\right) \leq \frac{\omega}{4} M_\alpha \quad (6.9)$$

for $N \geq N_0$, and the same bound holds with x' in place of x .

We now compare the terms with $i, j \in \mathcal{J}$, and we begin with the gaps. Fix such labels $i < j$. The points $x_i, \dots, x_j$ belong to $[x_i, x_j]$, which is covered by $\lfloor N(x_j - x_i) \rfloor + 1$ intervals of length N^{-1} , so hypothesis (a) implies that

$$j - i + 1 \leq N^\varepsilon (N(x_j - x_i) + 1).$$

The set $\mathcal{J}$ consists of consecutive integers, so every label k with $i \leq k < j$ satisfies $k, k+1 \in \mathcal{J}$. Hypothesis (c) and a telescoping sum then give

$$\begin{aligned} |(x'_j - x'_i) - (x_j - x_i)| &\leq (j - i)N^{-2\varepsilon}\ell_\alpha \\ &\leq N^{-\varepsilon}(N(x_j - x_i) + 1)\ell_\alpha \leq N^{-\varepsilon}(x_j - x_i + \ell_\alpha), \end{aligned} \quad (6.10)$$

where in the last inequality we used that $N\ell_\alpha = N^{-\alpha} \leq 1$.

Next, observe that for $y \geq 0$,

$$\left| \frac{d}{dy} \log f_p(y) \right| = \frac{2py}{1+y^2} \leq \frac{4p}{1+y}, \quad (6.11)$$

since $y + y^2 \leq 2 + 2y^2$. We set

$$y = \frac{x_j - x_i}{\ell_\alpha}, \quad y' = \frac{x'_j - x'_i}{\ell_\alpha}.$$

By (6.10), we have $|y - y'| \leq N^{-\varepsilon}(1 + y)$, so every y'' between y and y' satisfies

$$1 + y'' \geq (1 - N^{-\varepsilon})(1 + y) \geq \frac{1 + y}{2}$$

for $N \geq N_0$. Integrating (6.11) along the segment between y and y' , we obtain

$$\left| \log f_p\left(\frac{x_j - x_i}{\ell_\alpha}\right) - \log f_p\left(\frac{x'_j - x'_i}{\ell_\alpha}\right) \right| \leq \frac{8p|y - y'|}{1 + y} \leq 8pN^{-\varepsilon}. \quad (6.12)$$

It remains to compare the energy weights. The midpoints $E = (x_i + x_j)/2$ and $E' = (x'_i + x'_j)/2$ belong to $[-3, 3]$, and hypothesis (b) gives

$$|E - E'| \leq 2N^{-1+\varepsilon} \leq 1.$$

By part (2) of Lemma 5.1, we have

$$|\varphi_N(E) - \varphi_N(E')| \leq (\exp(16N^{-1+\varepsilon}h_N^{-2}) - 1)\varphi_N(E') = (\exp(16N^{-4/5+\varepsilon}) - 1)\varphi_N(E'),$$

and the factor in parentheses tends to zero, as $\varepsilon \leq 1/10$. Combining this bound with (6.12), we find that for $N \geq N_0$ and all $i < j$ in $\mathcal{J}$,

$$\left| \varphi_N(E)f_p\left(\frac{x_j - x_i}{\ell_\alpha}\right) - \varphi_N(E')f_p\left(\frac{x'_j - x'_i}{\ell_\alpha}\right) \right| \leq \frac{\omega}{2}\varphi_N(E')f_p\left(\frac{x'_j - x'_i}{\ell_\alpha}\right).$$

We sum over these pairs and bound the resulting sum of nonnegative terms on the right by $\mathcal{X}_\alpha(x'; \varphi_N, f_p)$. Then the contributions of the pairs in $\mathcal{J}$ to $\mathcal{X}_\alpha(x; \varphi_N, f_p)$ and $\mathcal{X}_\alpha(x'; \varphi_N, f_p)$ differ by at most $(\omega/2)\mathcal{X}_\alpha(x'; \varphi_N, f_p)$. Together with (6.9) for x and for x' , this gives (6.6). $\square$

Proposition 6.5. *For every $\alpha \in (0, 1/8)$, the statement $\mathcal{P}(\alpha)$ holds.*

Proof. Fix $\alpha \in (0, 1/8)$, an integer $p \geq p_*$, and a nonnegative $\varphi \in C_c^\infty((-2, 2))$. We set

$$L = \frac{1}{2(p-1)} \int_{-2}^2 \varphi(E) c(E) dE, \quad T = N^{-3/4-\alpha}, \quad \varepsilon = \frac{1}{100} \left(\frac{1}{4} - 2\alpha \right). \quad (6.13)$$

Then $\varepsilon \in (0, 1/400]$, and a direct computation shows that

$$\begin{aligned} T &\geq N^{-1+\varepsilon}, & \alpha + 5\varepsilon &< \frac{1}{2}, \\ \frac{N^\varepsilon}{N^2 T} &= N^{-5/4+\alpha+\varepsilon} \leq N^{-2\varepsilon} \ell_\alpha, & T N^{1/2+3\alpha+12\varepsilon} &= N^{-1/4+2\alpha+12\varepsilon}. \end{aligned} \quad (6.14)$$

Indeed, the first inequality is equivalent to $\varepsilon \leq 1/4 - \alpha$, the third to $3\varepsilon \leq 1/4 - 2\alpha$, and the exponent in the last expression is $-(88/100)(1/4 - 2\alpha) < 0$. Using Lemma 6.3 with $t_N = T$, we see that it suffices to show that

$$\frac{1}{M_\alpha} \mathcal{X}_\alpha(H_T; \varphi_N, f_p) \rightarrow L \quad \text{in probability.} \quad (6.15)$$

We first consider the GOE matrix G . Let $\delta \in (0, 1)$ satisfy $\text{supp } \varphi \subset [-2 + 3\delta, 2 - 3\delta]$. We fix $\chi \in C_c^\infty((-2, 2))$ such that $0 \leq \chi \leq 2$ and $\chi = 2$ on $[-2 + 2\delta, 2 - 2\delta]$ (the interval $[-2 + 2\delta, 2 - 2\delta]$ contains every point at distance at most δ from $\text{supp } \varphi$, as required in Lemma 6.2).

The functions $(r, E) \mapsto \varphi(E) f_p(r)$ and $(r, E) \mapsto \chi(E) f_{p_*}(r)$ are continuous, their E -supports lie in bulk intervals, and they are bounded by a constant times $(1 + r^2)^{-2}$. Hence Proposition 3.6 and (6.1) show that

$$\frac{1}{M_\alpha} \mathcal{X}_\alpha(G; \varphi, f_p) \rightarrow L, \quad \frac{1}{M_\alpha} \mathcal{X}_\alpha(G; \chi, f_{p_*}) \rightarrow \frac{1}{2(p_*-1)} \int_{-2}^2 \chi(E) c(E) dE \quad (6.16)$$

in probability. Since $p \geq p_*$, we have $f_p \leq f_{p_*}$. Then Lemma 6.2, applied with $A = G$ and $f = f_p$, shows that on the event $\{\|G\| \leq 3\}$, for $N \geq N_0$,

$$\frac{1}{M_\alpha} |\mathcal{X}_\alpha(G; \varphi_N, f_p) - \mathcal{X}_\alpha(G; \varphi, f_p)| \leq \frac{Ch_N}{M_\alpha} \mathcal{X}_\alpha(G; \chi, f_{p_*}) + \frac{CN^2}{M_\alpha} \exp\left(-\frac{\delta^2 N^{1/5}}{4}\right).$$

By (6.16), the right side tends to zero in probability, and the event $\{\|G\| \leq 3\}$ has probability tending to one by Proposition 3.2. We conclude that

$$\frac{1}{M_\alpha} \mathcal{X}_\alpha(G; \varphi_N, f_p) \rightarrow L \quad \text{in probability.} \quad (6.17)$$

We now use the coupling. Let $(H_t^{\text{dyn}})_{t \geq 0}$ and $(\mu(t))_{t \geq 0}$ be as in Proposition 3.5, and set

$$\lambda(T) = (\lambda_1(H_T^{\text{dyn}}), \dots, \lambda_N(H_T^{\text{dyn}})), \quad \mu(T) = (\mu_1(T), \dots, \mu_N(T)).$$

Recall that H_T^{dyn} has the law of H_T , and that $\mu(T)$ has the law of the ordered eigenvalues of G .

Let c_* be as in Lemma 6.4, defined with our choice of δ . We consider the event on which $H_T^{\text{dyn}} \in \mathcal{G}_\varepsilon$, $\max_{i \in [N]} |\mu_i(T)| \leq 3$, and

$$\begin{aligned} & \max_{c_* N \leq i \leq (1-c_*)N} \max \{ |\lambda_i(H_T^{\text{dyn}}) - \gamma_i|, |\mu_i(T) - \gamma_i| \} \leq N^{-1+\varepsilon}, \\ & \max_{c_* N \leq i \leq (1-c_*)N} |(\lambda_{i+1}(H_T^{\text{dyn}}) - \lambda_i(H_T^{\text{dyn}})) - (\mu_{i+1}(T) - \mu_i(T))| \leq \frac{N^\varepsilon}{N^2 T}. \end{aligned} \quad (6.18)$$

By Proposition 3.2, applied to H_T and to G (whose ordered eigenvalues have the law of $\mu(T)$), the conditions other than the second inequality in (6.18) hold with overwhelming probability. We have $N^{-1+\varepsilon} \leq T < 1$ by (6.14), so Proposition 3.5 shows that the second inequality in (6.18) holds with probability at least $1 - N^{-1}$ for $N \geq N_0$. Hence this event has probability tending to one.

The sequences $x = \lambda(T)$ and $x' = \mu(T)$ satisfy the hypotheses (a)–(c) of Lemma 6.4 on this event, where for (c) we used the third relation in (6.14). Hence, for every $\varpi > 0$, with probability tending to one,

$$\begin{aligned} & \left| \frac{1}{M_\alpha} \mathcal{X}_\alpha(\lambda(T); \varphi_N, f_p) - L \right| \\ & \leq \left| \frac{1}{M_\alpha} \mathcal{X}_\alpha(\mu(T); \varphi_N, f_p) - L \right| + \varpi \left(\frac{1}{M_\alpha} \mathcal{X}_\alpha(\mu(T); \varphi_N, f_p) + 1 \right). \end{aligned}$$

By (6.17), we have $M_\alpha^{-1} \mathcal{X}_\alpha(\mu(T); \varphi_N, f_p) \rightarrow L$ in probability. Since $\varpi > 0$ is arbitrary, it follows that $M_\alpha^{-1} \mathcal{X}_\alpha(\lambda(T); \varphi_N, f_p) \rightarrow L$ in probability. The sequence $\lambda(T)$ has the law of the ordered eigenvalues of H_T , which gives (6.15). $\square$

6.2 General test functions

We now show that $\mathcal{P}(\alpha)$ extends to general test functions f and to unsmoothed energy weights, and that it gives an upper bound on the number of pairs at distance of order ℓ_α .

Lemma 6.6. *Let $\alpha \in (0, 1/2)$, and suppose that $\mathcal{P}(\alpha)$ holds. Let $\varphi \in C_c^\infty((-2, 2))$ be nonnegative, and let $f: [0, \infty) \rightarrow \mathbb{R}$ be a continuous function such that the limit $\lim_{y \rightarrow \infty} (1 + y^2)^{p_*} f(y)$ exists in $\mathbb{R}$. Then*

$$\frac{1}{M_\alpha} \mathcal{X}_\alpha(H; \varphi_N, f) \rightarrow \int_{-2}^2 \varphi(E) c(E) \, dE \int_0^\infty y f(y) \, dy \quad (6.19)$$

in probability, and the same holds with φ_N replaced by φ . Further, for every bulk interval I , there exists $C_I > 0$ such that

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\sum_{\substack{1 \leq i < j \leq N \\ (\lambda_i + \lambda_j)/2 \in I}} f_{p_*} \left(\frac{\lambda_j - \lambda_i}{\ell_\alpha} \right) \leq C_I M_\alpha \right) = 1. \quad (6.20)$$

Proof. We set

$$C_f = \sup_{y \geq 0} (1 + y^2)^{p_*} |f(y)|, \quad c_\varphi = \int_{-2}^2 \varphi(E) c(E) dE.$$

The constant C_f is finite by our assumption on f . (This assumption holds, for example, for continuous functions with compact support and for the function $y \mapsto e^{-y^2/8}$, which we use in Section 6.3.) Then $|f| \leq C_f f_{p_*}$, so the integral $\int_0^\infty y f(y) dy$ converges absolutely.

We begin with (6.19), and we first approximate f by a combination of the functions f_p . Define $\Theta_f: [0, 1] \rightarrow \mathbb{R}$ by

$$\Theta_f(x) = x^{-p_*} f((x^{-1} - 1)^{1/2}) \quad \text{for } x \in (0, 1], \quad \Theta_f(0) = \lim_{y \rightarrow \infty} (1 + y^2)^{p_*} f(y).$$

For $y \geq 0$ and $x = (1 + y^2)^{-1}$, we have $(x^{-1} - 1)^{1/2} = y$ and $x^{-p_*} = (1 + y^2)^{p_*}$. Hence Θ_f is continuous on $[0, 1]$, and

$$f(y) = f_{p_*}(y) \Theta_f((1 + y^2)^{-1}) \quad \text{for all } y \geq 0.$$

Let $\omega \in (0, 1)$. By the Weierstrass approximation theorem, there exist an integer $p_1 \geq p_*$ and real numbers $\beta_{p_*}, \dots, \beta_{p_1}$ such that

$$\sup_{x \in [0, 1]} \left| \Theta_f(x) - \sum_{p=p_*}^{p_1} \beta_p x^{p-p_*} \right| \leq \omega.$$

Note that $f_{p_*}(y)(1 + y^2)^{-(p-p_*)} = f_p(y)$. Therefore the function

$$\bar{f} = \sum_{p=p_*}^{p_1} \beta_p f_p$$

satisfies

$$|f(y) - \bar{f}(y)| \leq \omega f_{p_*}(y) \quad \text{for all } y \geq 0. \quad (6.21)$$

The statistic $\mathcal{X}_\alpha(H; \varphi_N, f)$ is linear in f . Applying $\mathcal{P}(\alpha)$ with each $p \in \{p_*, \dots, p_1\}$, and using (6.1), we obtain

$$\frac{1}{M_\alpha} \mathcal{X}_\alpha(H; \varphi_N, \bar{f}) \rightarrow c_\varphi \sum_{p=p_*}^{p_1} \frac{\beta_p}{2(p-1)} = c_\varphi \int_0^\infty y \bar{f}(y) dy \quad \text{in probability.}$$

Since $\varphi_N \geq 0$, (6.21) implies that

$$|\mathcal{X}_\alpha(H; \varphi_N, f) - \mathcal{X}_\alpha(H; \varphi_N, \bar{f})| \leq \omega \mathcal{X}_\alpha(H; \varphi_N, f_{p_*}), \quad \left| \int_0^\infty y(f - \bar{f})(y) dy \right| \leq \omega,$$

where for the second bound we also used (6.1). Hence

$$\begin{aligned} \left| \frac{1}{M_\alpha} \mathcal{X}_\alpha(H; \varphi_N, f) - c_\varphi \int_0^\infty y f(y) dy \right| \\ \leq \left| \frac{1}{M_\alpha} \mathcal{X}_\alpha(H; \varphi_N, \bar{f}) - c_\varphi \int_0^\infty y \bar{f}(y) dy \right| + \frac{\omega}{M_\alpha} \mathcal{X}_\alpha(H; \varphi_N, f_{p_*}) + \omega c_\varphi. \end{aligned}$$

By $\mathcal{P}(\alpha)$ with $p = p_*$, the second term on the right converges in probability to $\omega c_\varphi / (2(p_* - 1)) \leq \omega c_\varphi$. So the left side is bounded by $3\omega(c_\varphi + 1)$ with probability tending to one. Since ω is arbitrary, this proves (6.19).

Next, we remove the smoothing, and we may assume that $C_f > 0$. Let $\delta > 0$ be such that every E with $\text{dist}(E, \text{supp } \varphi) \leq \delta$ belongs to $(-2, 2)$. We fix $\chi \in C_c^\infty((-2, 2))$ such that $0 \leq \chi \leq 2$ and $\chi(E) = 2$ for all such E . We apply Lemma 6.2 to f/C_f , and we find that on the event $\{\|H\| \leq 3\}$, for $N \geq N_0$,

$$\frac{1}{M_\alpha} |\mathcal{X}_\alpha(H; \varphi_N, f) - \mathcal{X}_\alpha(H; \varphi, f)| \leq \frac{CC_f h_N}{M_\alpha} \mathcal{X}_\alpha(H; \chi_N, f_{p_*}) + \frac{CC_f N^2}{M_\alpha} \exp\left(-\frac{\delta^2 N^{1/5}}{4}\right).$$

By $\mathcal{P}(\alpha)$, applied with χ in place of φ and $p = p_*$, the random variable $M_\alpha^{-1} \mathcal{X}_\alpha(H; \chi_N, f_{p_*})$ converges in probability, and the right side tends to zero in probability. As the event $\{\|H\| \leq 3\}$ has probability tending to one by Proposition 3.2, the convergence (6.19) implies the same convergence with φ in place of φ_N .

Finally, we prove (6.20). Let I be a bulk interval. Let χ now denote a function in $C_c^\infty((-2, 2))$ such that $0 \leq \chi \leq 2$ and $\chi = 2$ on I . By part (5) of Lemma 5.1, we have

$$\|\chi_N - \chi\|_\infty \leq Ch_N \leq 1$$

for $N \geq N_0$, so $\chi_N \geq 1$ on I . All terms of $\mathcal{X}_\alpha(H; \chi_N, f_{p_*})$ are nonnegative, and therefore the sum in (6.20) is bounded by $\mathcal{X}_\alpha(H; \chi_N, f_{p_*})$. By $\mathcal{P}(\alpha)$, we have

$$\frac{1}{M_\alpha} \mathcal{X}_\alpha(H; \chi_N, f_{p_*}) \rightarrow \frac{1}{2(p_* - 1)} \int_{-2}^2 \chi(E) c(E) dE \quad \text{in probability.}$$

Hence (6.20) holds with

$$C_I = 1 + \int_{-2}^2 \chi(E) c(E) dE.$$

□

6.3 The induction step

Proposition 6.7. *Let α, α' satisfy*

$$\frac{1}{10} \leq \alpha < \alpha' < \frac{1}{2}, \quad \alpha' < \frac{6\alpha}{5}, \quad \alpha' < \frac{1}{6} + \frac{2\alpha}{3}. \quad (6.22)$$

If $\mathcal{P}(\alpha)$ holds, then $\mathcal{P}(\alpha')$ holds.

To prove Proposition 6.7, we use Lemma 6.3 to replace H by H_{t_α} , with t_α as in (6.23) below, and we approximate $\mathcal{X}_{\alpha'}(H_{t_\alpha}; \varphi_N, f_p)$ by a sum of conditionally independent contributions of 2×2 blocks, whose conditional means are computed by Lemma 3.10 and determined by $\mathcal{P}(\alpha)$. We first fix the parameters and define the blocks. Lemma 6.8 collects the properties of the clusters that we need, Lemma 6.9 bounds the error of the approximation, Lemma 6.10 computes the conditional mean of the sum, and Lemma 6.11 shows that the sum concentrates around this mean. We prove Proposition 6.7 after the proofs of these lemmas. The main technical point is the approximation, which relies on Corollary 3.9 and on the first three constraints in (6.25).

Parameters. Throughout the rest of Section 6.3, we fix α and α' satisfying (6.22). The parameters and the notation introduced in this paragraph and the next two are used in the statements and the proofs of all lemmas below. We set

$$\kappa = \alpha - \frac{\alpha'}{6}, \quad q = N^{-1-\kappa}, \quad t_\alpha = N\ell_\alpha^2 = N^{-1-2\alpha}. \quad (6.23)$$

Then (6.22) implies that

$$\frac{2\alpha'}{3} < \kappa < 2\alpha - \alpha' < \alpha, \quad \alpha' - \kappa = \frac{7\alpha'}{6} - \alpha > \frac{\alpha}{6} \geq \frac{1}{60}. \quad (6.24)$$

Indeed, the first two inequalities are equivalent to $\alpha' < 6\alpha/5$, the third is equivalent to $\alpha < \alpha'$, and the last two follow from $\alpha < \alpha'$ and $\alpha \geq 1/10$.

We fix $\varepsilon > 0$ such that

$$100\varepsilon < \min \left\{ \frac{\alpha'}{6}, \alpha - \frac{5\alpha'}{6}, 3\alpha - \frac{5\alpha'}{2}, \frac{1}{2} + 2\alpha - 3\alpha', \frac{1}{2} - \alpha' \right\}. \quad (6.25)$$

Every number in this minimum is positive. For the second and third numbers, this follows from $\alpha' < 6\alpha/5$, for the fourth from $\alpha' < 1/6 + 2\alpha/3$, and for the fifth from $\alpha' < 1/2$. In terms of κ , the first three numbers are $\alpha - \kappa$, $2\alpha - \kappa - \alpha'$, and $3\kappa - 2\alpha'$. Moreover,

$$\varepsilon < \frac{\alpha'}{6} < \kappa,$$

where the last inequality is equivalent to $\alpha' < 3\alpha$.

For an integer $p \geq p_*$ and a nonnegative $\varphi \in C_c^\infty((-2, 2))$, we set

$$L = \frac{1}{2(p-1)} \int_{-2}^2 \varphi(E) c(E) dE. \quad (6.26)$$

In the statements and the proofs of the lemmas below, constants may depend on $\alpha, \alpha', \varepsilon, p$, and φ , and N_0 may increase from line to line.

Gaussian blocks. Let $U = (u_1, \dots, u_N)$ be the orthogonal matrix whose columns are the chosen eigenvectors of H . We set

$$y_i = e^{-t_\alpha/2} \lambda_i \quad (i \in [N]), \quad D_y = \text{diag}(y_1, \dots, y_N), \quad W = (1 - e^{-t_\alpha})^{1/2} U^T G U. \quad (6.27)$$

Then

$$U^T H_{t_\alpha} U = D_y + W, \quad \text{so} \quad \lambda_i(H_{t_\alpha}) = \lambda_i(D_y + W) \quad \text{for all } i \in [N]. \quad (6.28)$$

The matrix G is independent of H , and the law of G is invariant under conjugation by a fixed orthogonal matrix. Since U is a measurable function of H , the conditional law of $U^T G U$ given H is the law of G . Consequently, W is independent of H , and its entries W_{ij} with $i \leq j$ are independent centered Gaussian random variables satisfying

$$\mathbb{E} W_{ij}^2 = \frac{1 - e^{-t_\alpha}}{N} \quad (i < j), \quad \mathbb{E} W_{ii}^2 = \frac{2(1 - e^{-t_\alpha})}{N}. \quad (6.29)$$

We next bound the entries of W . By (6.29), we have for all $i \leq j$

$$\mathbb{E} W_{ij}^2 \leq \frac{2(1 - e^{-t_\alpha})}{N} \leq \frac{2t_\alpha}{N} = 2\ell_\alpha^2.$$

We set

$$w_0 = N^\varepsilon \ell_\alpha. \quad (6.30)$$

The standard Gaussian tail bound and a union bound over the $N(N+1)/2$ entries give

$$\mathbb{P}\left(\max_{i \leq j} |W_{ij}| > w_0\right) \leq 2N^2 \exp\left(-\frac{N^{2\varepsilon}}{4}\right) \rightarrow 0. \quad (6.31)$$

Clusters. Let $\mathfrak{K}$ denote the family of clusters obtained from the points $y_1 \leq \dots \leq y_N$ by joining i and $i+1$ whenever $y_{i+1} - y_i < q$, as in Corollary 3.9. (The threshold q is much larger than ℓ_α , since $\kappa < \alpha$, and much smaller than N^{-1} .) Every cluster is a set of consecutive integers, and the family $\mathfrak{K}$ is a function of H .

Let $\mathcal{K}, \mathcal{K}' \in \mathfrak{K}$ satisfy $\max \mathcal{K} < \min \mathcal{K}'$. Then the labels $\max \mathcal{K}$ and $\max \mathcal{K} + 1$ lie in different clusters, and since y is nondecreasing,

$$y_{\min \mathcal{K}'} - y_{\max \mathcal{K}} \geq y_{\max \mathcal{K}+1} - y_{\max \mathcal{K}} \geq q. \quad (6.32)$$

Let $\mathfrak{P}_2$ denote the set of clusters with exactly two elements. For $\mathcal{K} \in \mathfrak{P}_2$, we write $\mathcal{K} = \{i, i+1\}$, and we set $g_{\mathcal{K}} = g_i$, $E_{\mathcal{K}} = E_i$, and

$$B_{\mathcal{K}} = \begin{pmatrix} y_i + W_{ii} & W_{i,i+1} \\ W_{i,i+1} & y_{i+1} + W_{i+1,i+1} \end{pmatrix}, \quad \hat{g}_{\mathcal{K}} = \lambda_2(B_{\mathcal{K}}) - \lambda_1(B_{\mathcal{K}}).$$

Thus $B_{\mathcal{K}} = (D_y + W)_{\mathcal{K}\mathcal{K}}$. For an integer $p \geq p_*$ and a nonnegative $\varphi \in C_c^\infty((-2, 2))$, we define the block statistic

$$\hat{\mathcal{X}}_{\alpha'}(\varphi, p) = \sum_{\mathcal{K} \in \mathfrak{P}_2} \varphi_N(E_{\mathcal{K}}) f_p\left(\frac{\hat{g}_{\mathcal{K}}}{\ell_{\alpha'}}\right). \quad (6.33)$$

Lemma 6.8. *There exists $N_0 \geq 1$ such that the following holds for all $N \geq N_0$ on the event that $H \in \mathcal{G}_{\varepsilon/2}$. We have $2N^\varepsilon q < N^{-1}$, the points $y_1 \leq \dots \leq y_N$ lie in $[-3, 3]$, and every interval of length N^{-1} contains at most N^ε of them. Every cluster in $\mathfrak{K}$ has at most N^ε elements, and the number of pairs $i < j$ that lie in a common cluster $\mathcal{K} \in \mathfrak{K}$ with $|\mathcal{K}| \geq 3$ is at most $2N^\varepsilon \mathcal{N}_3(4q)$, where $\mathcal{N}_3(r)$ is as in Proposition 3.4. Moreover, without restriction to this event, we have*

$$\lim_{N \rightarrow \infty} \frac{1}{M_{\alpha'}} \mathbb{E}[N^\varepsilon \mathcal{N}_3(4q)] = 0.$$

Proof. Suppose that $H \in \mathcal{G}_{\varepsilon/2}$. We first verify the hypotheses of Corollary 3.9 on the points $y_1 \leq \dots \leq y_N$ and the threshold q . We have $\|H\| \leq 3$, so $y_i \in [-3, 3]$ for all i . Each λ_i with $y_i \in [E, E + N^{-1}]$ belongs to an interval of length $e^{t_\alpha/2} N^{-1} \leq 2N^{-1}$, which is covered by two intervals of length N^{-1} . Since $H \in \mathcal{G}_{\varepsilon/2}$, it follows that every interval of length N^{-1} contains at most $2N^{\varepsilon/2} \leq N^\varepsilon$ of the points y_i , for $N \geq N_0$. Further, since $\varepsilon < \kappa$, we have

$$2N^\varepsilon q = 2N^{\varepsilon-\kappa} N^{-1} < N^{-1}$$

for $N \geq N_0$. Hence Corollary 3.9 shows that every cluster has at most N^ε elements.

We next count the pairs $i < j$ in a common cluster $\mathcal{K}$ with $|\mathcal{K}| \geq 3$. Recall from Proposition 3.4 that $\mathcal{N}_3(r)$ denotes the number of $i \in [N-2]$ such that $\lambda_{i+2} - \lambda_i \leq r$. Let $\mathcal{K} \in \mathfrak{K}$ satisfy $|\mathcal{K}| \geq 3$. For each of the $|\mathcal{K}| - 2$ labels i such that $i, i+2 \in \mathcal{K}$, the definition of the clusters gives

$$\lambda_{i+2} - \lambda_i = e^{t_\alpha/2} (y_{i+2} - y_i) < 2e^{t_\alpha/2} q \leq 4q.$$

These $|\mathcal{K}| - 2$ labels are distinct for distinct clusters, which implies that

$$\sum_{\mathcal{K} \in \mathfrak{K}: |\mathcal{K}| \geq 3} (|\mathcal{K}| - 2) \leq \mathcal{N}_3(4q).$$

Recall that $|\mathcal{K}| \leq N^\varepsilon$, and note that $|\mathcal{K}| \leq 3(|\mathcal{K}| - 2)$ for $|\mathcal{K}| \geq 3$. Then the number of these pairs is bounded by

$$\sum_{\mathcal{K} \in \mathfrak{K}: |\mathcal{K}| \geq 3} \frac{|\mathcal{K}|^2}{2} \leq \frac{N^\varepsilon}{2} \sum_{\mathcal{K} \in \mathfrak{K}: |\mathcal{K}| \geq 3} 3(|\mathcal{K}| - 2) \leq 2N^\varepsilon \mathcal{N}_3(4q). \quad (6.34)$$

Finally, using Proposition 3.4 with $r = 4q \geq N^{-2}$, we have

$$\frac{1}{M_{\alpha'}} \mathbb{E}[N^\varepsilon \mathcal{N}_3(4q)] \leq \frac{CN^\varepsilon}{M_{\alpha'}} N(4Nq)^3 \leq CN^{\varepsilon+2\alpha'-3\kappa} = CN^{\varepsilon-(3\alpha-5\alpha'/2)} \rightarrow 0, \quad (6.35)$$

where the last step follows from (6.25). $\square$

Lemma 6.9. *Let $p \geq p_*$ be an integer, let $\varphi \in C_c^\infty((-2, 2))$ be nonnegative, and let $\hat{\mathcal{X}}_{\alpha'}(\varphi, p)$ be the block statistic defined in (6.33). Then there exist $N_0 \geq 1$ and a deterministic sequence $o(1)$ such that for all $N \geq N_0$, on the event that $H \in \mathcal{G}_{\varepsilon/2}$ and $\max_{i \leq j} |W_{ij}| \leq w_0$, with W and w_0 as in (6.27) and (6.30), we have*

$$\begin{aligned} |\mathcal{X}_{\alpha'}(H_{t_\alpha}; \varphi_N, f_p) - \hat{\mathcal{X}}_{\alpha'}(\varphi, p)| \\ \leq o(1) \hat{\mathcal{X}}_{\alpha'}(\varphi, p) + 2\|\varphi\|_\infty N^\varepsilon \mathcal{N}_3(4q) + 4^p \|\varphi\|_\infty N^2 \left(\frac{\ell_{\alpha'}}{q}\right)^{2p}, \end{aligned} \quad (6.36)$$

where $\mathcal{N}_3(r)$ is as in Proposition 3.4.

Proof. Suppose that the event holds and that $N \geq N_0$. By Lemma 6.8, the points $y_1 \leq \dots \leq y_N$ and the threshold q satisfy the hypotheses of Corollary 3.9 that concern them. It remains to verify the condition (3.30) on W . On our event, the entries of W are bounded by $w_0 = N^\varepsilon \ell_\alpha$. Since $N \log N \leq q^{-1}$ for $N \geq N_0$ and $\alpha - \kappa = \alpha'/6$, we have

$$N^\varepsilon w_0 (q^{-1} + N \log N) \leq 2N^{2\varepsilon} \ell_\alpha q^{-1} = 2N^{2\varepsilon - \alpha'/6} \leq \frac{1}{48}$$

for $N \geq N_0$, where we used that $2\varepsilon < \alpha'/6$ by (6.25). Hence Corollary 3.9 applies to the points $y_1 \leq \dots \leq y_N$, the threshold q , and the matrix W . By (6.28), it shows that

$$\lambda_{i+1}(H_{t_\alpha}) - \lambda_i(H_{t_\alpha}) \geq \frac{q}{2} \quad \text{whenever } i \text{ and } i+1 \text{ belong to different clusters,} \quad (6.37)$$

and that for every cluster $\mathcal{K} = \{i_0 + 1, \dots, i_0 + |\mathcal{K}|\}$ and every $i \in [|\mathcal{K}|]$,

$$\begin{aligned} |\lambda_{i_0+i}(H_{t_\alpha}) - \lambda_i((D_y + W)_{\mathcal{K}\mathcal{K}})| &\leq 24N^{2\varepsilon} w_0^2 (q^{-1} + N \log N) \leq 48N^{4\varepsilon} \ell_\alpha^2 q^{-1} \\ &= 48N^{4\varepsilon - \alpha + 5\alpha'/6} \ell_{\alpha'} \leq N^{-\varepsilon} \ell_{\alpha'} \end{aligned} \quad (6.38)$$

for $N \geq N_0$. In the equality, we used that

$$\ell_\alpha^2 q^{-1} = N^{-1-2\alpha+\kappa}, \quad -1 - 2\alpha + \kappa = -1 - \alpha' + (5\alpha'/6 - \alpha),$$

and in the last inequality, we used $5\varepsilon < \alpha - 5\alpha'/6$, which follows from (6.25).

We now sort the pairs $i < j$ in $\mathcal{X}_{\alpha'}(H_{t_\alpha}; \varphi_N, f_p)$ into three types. The first type consists of the pairs with $i \in \mathcal{K}$ and $j \in \mathcal{K}'$ for distinct clusters $\mathcal{K}$ and $\mathcal{K}'$, so that $\max \mathcal{K} < \min \mathcal{K}'$.

For such a pair, the labels $\max \mathcal{K}$ and $\max \mathcal{K} + 1$ belong to different clusters, so (6.37) gives

$$\lambda_j(H_{t_\alpha}) - \lambda_i(H_{t_\alpha}) \geq \lambda_{\max \mathcal{K}+1}(H_{t_\alpha}) - \lambda_{\max \mathcal{K}}(H_{t_\alpha}) \geq \frac{q}{2}.$$

Hence

$$f_p\left(\frac{\lambda_j(H_{t_\alpha}) - \lambda_i(H_{t_\alpha})}{\ell_{\alpha'}}\right) \leq \left(1 + \frac{q^2}{4\ell_{\alpha'}^2}\right)^{-p} \leq 4^p \left(\frac{\ell_{\alpha'}}{q}\right)^{2p}.$$

Using $\varphi_N \leq \|\varphi\|_\infty$, we see that the total contribution of the pairs of the first type is bounded by $4^p \|\varphi\|_\infty N^2 (\ell_{\alpha'}/q)^{2p}$.

The second type consists of the pairs $i < j$ in a common cluster $\mathcal{K}$ with $|\mathcal{K}| \geq 3$. By Lemma 6.8, there are at most $2N^\varepsilon \mathcal{N}_3(4q)$ such pairs. Each of their terms is bounded by $\|\varphi\|_\infty$, since $f_p \leq 1$.

The third type consists of the pairs $(i, i+1)$ such that $\{i, i+1\} \in \mathfrak{P}_2$. Fix such a cluster $\mathcal{K} = \{i, i+1\}$. By (6.38), we have

$$|(\lambda_{i+1}(H_{t_\alpha}) - \lambda_i(H_{t_\alpha})) - \hat{g}_{\mathcal{K}}| \leq 2N^{-\varepsilon} \ell_{\alpha'}.$$

The right side of (6.11) is at most $4p$, so integrating (6.11), we find

$$\left| \log f_p\left(\frac{\lambda_{i+1}(H_{t_\alpha}) - \lambda_i(H_{t_\alpha})}{\ell_{\alpha'}}\right) - \log f_p\left(\frac{\hat{g}_{\mathcal{K}}}{\ell_{\alpha'}}\right) \right| \leq 8pN^{-\varepsilon}.$$

We now turn to the midpoints. Observe that

$$\lambda_1(B_{\mathcal{K}}) + \lambda_2(B_{\mathcal{K}}) = \text{Tr } B_{\mathcal{K}} = y_i + y_{i+1} + W_{ii} + W_{i+1, i+1}.$$

Using this identity, (6.38), and the bound $|E_{\mathcal{K}}| \leq 3$, we obtain

$$\left| \frac{\lambda_i(H_{t_\alpha}) + \lambda_{i+1}(H_{t_\alpha})}{2} - E_{\mathcal{K}} \right| \leq N^{-\varepsilon} \ell_{\alpha'} + w_0 + 3(1 - e^{-t_\alpha/2}) \leq 5N^\varepsilon \ell_{\alpha'},$$

where in the last inequality we used $\ell_{\alpha'} \leq \ell_\alpha$ and

$$1 - e^{-t_\alpha/2} \leq \frac{t_\alpha}{2} \leq \ell_\alpha.$$

Part (2) of Lemma 5.1, applied at the point $E_{\mathcal{K}} \in [-3, 3]$, shows that φ_N at this midpoint differs from $\varphi_N(E_{\mathcal{K}})$ by at most

$$(\exp(40N^\varepsilon \ell_\alpha h_N^{-2}) - 1) \varphi_N(E_{\mathcal{K}}) = (\exp(40N^{\varepsilon-4/5-\alpha}) - 1) \varphi_N(E_{\mathcal{K}}).$$

(This bound holds at every point of $[-3, 3]$, so midpoints near the edges of the spectrum or outside the support of φ require no separate treatment.)

Combining the last display with the bound on $\log f_p$ above, we conclude that

$$\begin{aligned} \left| \varphi_N \left(\frac{\lambda_i(H_{t_\alpha}) + \lambda_{i+1}(H_{t_\alpha})}{2} \right) f_p \left(\frac{\lambda_{i+1}(H_{t_\alpha}) - \lambda_i(H_{t_\alpha})}{\ell_{\alpha'}} \right) - \varphi_N(E_{\mathcal{K}}) f_p \left(\frac{\hat{g}_{\mathcal{K}}}{\ell_{\alpha'}} \right) \right| \\ \leq o(1) \varphi_N(E_{\mathcal{K}}) f_p \left(\frac{\hat{g}_{\mathcal{K}}}{\ell_{\alpha'}} \right), \end{aligned}$$

where the $o(1)$ term is deterministic and uniform in $\mathcal{K}$. Summing the bounds for the three types of pairs, we obtain (6.36). $\square$

Lemma 6.10. *Suppose that $\mathcal{P}(\alpha)$ holds. Let $p \geq p_*$ be an integer, let $\varphi \in C_c^\infty((-2, 2))$ be nonnegative, and let L be as in (6.26). Let $\hat{\mathcal{X}}_{\alpha'}(\varphi, p)$ be the block statistic defined in (6.33). Then*

$$\frac{1}{M_{\alpha'}} \mathbb{E}[\hat{\mathcal{X}}_{\alpha'}(\varphi, p) \mid H] \rightarrow L \quad \text{in probability.} \quad (6.39)$$

Proof. Let $\mathcal{K} \in \mathfrak{P}_2$, and write $\mathcal{K} = \{i, i+1\}$. We first compute the conditional law of $\hat{g}_{\mathcal{K}}$. Since $y_i - y_{i+1} = -e^{-t_\alpha/2} g_{\mathcal{K}}$, the formula (4.10) for the gap of a real symmetric 2×2 matrix gives

$$\frac{\hat{g}_{\mathcal{K}}}{\ell_\alpha} = \left| \left(-\frac{e^{-t_\alpha/2} g_{\mathcal{K}}}{\ell_\alpha} + \frac{2}{\ell_\alpha} \cdot \frac{W_{ii} - W_{i+1,i+1}}{2}, \frac{2}{\ell_\alpha} W_{i,i+1} \right) \right|. \quad (6.40)$$

By (6.29), conditionally on H , the variables $(W_{ii} - W_{i+1,i+1})/2$ and $W_{i,i+1}$ are independent centered Gaussian random variables, each with variance $(1 - e^{-t_\alpha})/N$. Hence the two random coordinates in (6.40) are independent centered Gaussian random variables with variance v_N , where

$$v_N = \frac{4(1 - e^{-t_\alpha})}{N\ell_\alpha^2} = \frac{4(1 - e^{-t_\alpha})}{t_\alpha} \in [4 - 2t_\alpha, 4]. \quad (6.41)$$

Here we used that

$$1 - \frac{t}{2} \leq \frac{1 - e^{-t}}{t} \leq 1 \quad \text{for } t \in (0, 1].$$

The law of a centered Gaussian vector in $\mathbb{R}^2$ with covariance $v_N \text{Id}_2$ is invariant under the reflection that changes the sign of the first coordinate, so we conclude that, conditionally on H ,

$$\frac{\hat{g}_{\mathcal{K}}}{\ell_\alpha} \stackrel{d}{=} |G_x|, \quad (6.42)$$

where G_x is as in Lemma 3.10, with $n = 2$ and

$$x = \frac{g_{\mathcal{K}}}{\ell_\alpha}, \quad \varrho = e^{-t_\alpha/2}, \quad v = v_N.$$

We now compute the conditional mean. Recall that

$$\ell_{\alpha'} = N^{\alpha-\alpha'} \ell_\alpha, \quad M_{\alpha'} = N^{2(\alpha-\alpha')} M_\alpha. \quad (6.43)$$

Using the first relation and (6.42), we find

$$\mathbb{E}\left[f_p\left(\frac{\hat{g}\mathcal{K}}{\ell_{\alpha'}}\right) \middle| H\right] = \mathbb{E}f_p(|G_x|/N^{\alpha-\alpha'}).$$

Since $\varrho \in [1/2, 2]$ and $v_N \in [2, 4]$, we may apply Lemma 3.10 with $n = 2$, the exponent $p > 3/2$, $v_0 = 4$, and $N^{\alpha-\alpha'} \leq 1/2$ in place of δ . For the density $\mathfrak{g}_4$ of a centered Gaussian vector in $\mathbb{R}^2$ with covariance 4 times the identity, we have

$$\mathfrak{g}_4(x, 0) = (8\pi)^{-1}e^{-x^2/8}, \quad \int_{\mathbb{R}^2} f_p(|y|) dy = 2\pi \int_0^\infty r f_p(r) dr = \frac{\pi}{p-1},$$

where the last equality follows from (6.1). Therefore Lemma 3.10 gives

$$\begin{aligned} \left| N^{2(\alpha'-\alpha)} \mathbb{E}\left[f_p\left(\frac{\hat{g}\mathcal{K}}{\ell_{\alpha'}}\right) \middle| H\right] - \frac{e^{-g_{\mathcal{K}}^2/(8\ell_{\alpha}^2)}}{8(p-1)} \right| \\ \leq C(N^{\alpha-\alpha'} + |e^{-t_{\alpha}/2} - 1| + |v_N - 4|) f_p\left(\frac{g_{\mathcal{K}}}{\ell_{\alpha}}\right), \end{aligned}$$

and by (6.41), the factor in parentheses on the right is $o(1)$, uniformly in $\mathcal{K}$.

We multiply by $\varphi_N(E_{\mathcal{K}})$ and sum over $\mathcal{K} \in \mathfrak{P}_2$. The terms with $\mathcal{K} \in \mathfrak{P}_2$ form part of the sum defining $\mathcal{X}_{\alpha}(H; \varphi_N, f_p)$, so the second relation in (6.43) gives

$$\left| \frac{1}{M_{\alpha'}} \mathbb{E}[\hat{\mathcal{X}}_{\alpha'}(\varphi, p) \middle| H] - \frac{1}{8(p-1)M_{\alpha}} \sum_{\mathcal{K} \in \mathfrak{P}_2} \varphi_N(E_{\mathcal{K}}) e^{-g_{\mathcal{K}}^2/(8\ell_{\alpha}^2)} \right| \leq \frac{o(1)}{M_{\alpha}} \mathcal{X}_{\alpha}(H; \varphi_N, f_p). \quad (6.44)$$

By $\mathcal{P}(\alpha)$, the right side tends to zero in probability.

We now evaluate the sum in (6.44). The function $y \mapsto e^{-y^2/8}$ satisfies the hypothesis of Lemma 6.6, and

$$\int_0^\infty y e^{-y^2/8} dy = 4.$$

So $\mathcal{P}(\alpha)$ and Lemma 6.6 give

$$\frac{1}{M_{\alpha}} \sum_{1 \leq i < j \leq N} \varphi_N\left(\frac{\lambda_i + \lambda_j}{2}\right) \exp\left(-\frac{(\lambda_j - \lambda_i)^2}{8\ell_{\alpha}^2}\right) \rightarrow 4 \int_{-2}^2 \varphi(E) c(E) dE \quad \text{in probability.}$$

The sum on the left differs from the sum in (6.44) by the terms of the pairs $i < j$ that do not form a cluster in $\mathfrak{P}_2$, which are nonnegative.

Suppose that $H \in \mathcal{G}_{\varepsilon/2}$ and $N \geq N_0$. A pair that does not form a cluster in $\mathfrak{P}_2$ either is contained in a cluster with at least three elements, or meets two distinct clusters $\mathcal{K}$ and $\mathcal{K}'$, with $i \in \mathcal{K}$ and $j \in \mathcal{K}'$. By Lemma 6.8, there are at most $2N^{\varepsilon} \mathcal{N}_3(4q)$ pairs of the former kind. For a pair of the latter kind, (6.32) gives

$$\lambda_j - \lambda_i \geq e^{t_{\alpha}/2} (y_{\min \mathcal{K}'} - y_{\max \mathcal{K}}) \geq q = N^{\alpha'/6} \ell_{\alpha},$$

where we used that $\alpha - \kappa = \alpha'/6$. Hence the difference of the two sums is bounded by

$$2\|\varphi\|_\infty N^\varepsilon \mathcal{N}_3(4q) + \|\varphi\|_\infty N^2 \exp\left(-\frac{q^2}{8\ell_\alpha^2}\right) = 2\|\varphi\|_\infty N^\varepsilon \mathcal{N}_3(4q) + \|\varphi\|_\infty N^2 \exp\left(-\frac{N^{\alpha'/3}}{8}\right).$$

Since $M_\alpha \geq M_{\alpha'}$, Lemma 6.8 and Proposition 3.2 show that this difference, divided by M_α , tends to zero in probability. Inserting this into (6.44), and recalling the definition (6.26) of L , we obtain (6.39). $\square$

Lemma 6.11. *Assume the hypotheses of Lemma 6.10. Then*

$$\frac{1}{M_{\alpha'}} \widehat{\mathcal{X}}_{\alpha'}(\varphi, p) \rightarrow L \quad \text{in probability.} \quad (6.45)$$

Proof. Conditionally on H , the matrices B_K , $K \in \mathfrak{P}_2$, are independent. Indeed, $\mathfrak{P}_2$ and y are functions of H , and each B_K depends on W only through the entries W_{ii} , $W_{i+1,i+1}$, and $W_{i,i+1}$. These sets of entries are disjoint for distinct clusters, and W is independent of H with independent entries. Hence $\widehat{\mathcal{X}}_{\alpha'}(\varphi, p)$ is a sum of conditionally independent random variables. Using $0 \leq \varphi_N \leq \|\varphi\|_\infty$ and $0 \leq f_p^2 \leq f_p \leq 1$, we find

$$\text{Var}(\widehat{\mathcal{X}}_{\alpha'}(\varphi, p) \mid H) \leq \sum_{K \in \mathfrak{P}_2} \varphi_N(E_K)^2 \mathbb{E}\left[f_p\left(\frac{\widehat{g}_K}{\ell_{\alpha'}}\right)^2 \mid H\right] \leq \|\varphi\|_\infty \mathbb{E}[\widehat{\mathcal{X}}_{\alpha'}(\varphi, p) \mid H].$$

Let $\omega > 0$. The conditional Chebyshev inequality gives

$$\mathbb{P}\left(\left|\widehat{\mathcal{X}}_{\alpha'}(\varphi, p) - \mathbb{E}[\widehat{\mathcal{X}}_{\alpha'}(\varphi, p) \mid H]\right| > \omega M_{\alpha'} \mid H\right) \leq \min\left\{1, \frac{\|\varphi\|_\infty}{\omega^2 M_{\alpha'}} \cdot \frac{1}{M_{\alpha'}} \mathbb{E}[\widehat{\mathcal{X}}_{\alpha'}(\varphi, p) \mid H]\right\}.$$

By Lemma 6.10 and $M_{\alpha'} \rightarrow \infty$, the right side tends to zero in probability. We take expectations and use the bounded convergence theorem, and we find that the unconditional probability tends to zero as well. Together with (6.39), this implies (6.45). $\square$

Proof of Proposition 6.7. Fix an integer $p \geq p_*$ and a nonnegative $\varphi \in C_c^\infty((-2, 2))$, which we may assume is not identically zero, and let L be as in (6.26). By (6.25), we have $\alpha' + 5\varepsilon < 1/2$ and

$$t_\alpha N^{1/2+3\alpha'+12\varepsilon} = N^{-1/2-2\alpha+3\alpha'+12\varepsilon} \rightarrow 0.$$

Hence Lemma 6.3, applied with α' in place of α and $t_N = t_\alpha$, shows that it suffices to prove

$$\frac{1}{M_{\alpha'}} \mathcal{X}_{\alpha'}(H_{t_\alpha}; \varphi_N, f_p) \rightarrow L \quad \text{in probability.} \quad (6.46)$$

The event that $H \in \mathcal{G}_{\varepsilon/2}$ and $\max_{i \leq j} |W_{ij}| \leq w_0$ has probability tending to one, by Proposition 3.2 and (6.31). On this event, Lemma 6.9 gives

$$\begin{aligned} \left| \frac{1}{M_{\alpha'}} \mathcal{X}_{\alpha'}(H_{t_\alpha}; \varphi_N, f_p) - L \right| &\leq \left| \frac{1}{M_{\alpha'}} \widehat{\mathcal{X}}_{\alpha'}(\varphi, p) - L \right| + \frac{o(1)}{M_{\alpha'}} \widehat{\mathcal{X}}_{\alpha'}(\varphi, p) \\ &\quad + \frac{2\|\varphi\|_\infty N^\varepsilon \mathcal{N}_3(4q)}{M_{\alpha'}} + \frac{4^p \|\varphi\|_\infty N^2}{M_{\alpha'}} \left(\frac{\ell_{\alpha'}}{q}\right)^{2p}. \end{aligned}$$

We now bound the last term on the right side. Since $M_{\alpha'} = N^{1-2\alpha'}$ and $\ell_{\alpha'}/q = N^{\kappa-\alpha'}$, this term is a constant multiple of $N^{1+2\alpha'-2p(\alpha'-\kappa)}$, so it tends to zero if and only if $2p(\alpha'-\kappa) > 1+2\alpha'$. This is the only step in the proof of Theorem 1.3 that uses the value $p_* = 61$ and the condition $\alpha \geq 1/10$. By (6.24), we have $\alpha' - \kappa > \alpha/6$, and hence

$$2p(\alpha' - \kappa) > \frac{p\alpha}{3} \geq \frac{p}{30} \geq \frac{61}{30} > 1 + 2\alpha',$$

where we used $\alpha \geq 1/10$, $p \geq p_* = 61$, and $\alpha' < 1/2$. We conclude that

$$\frac{1}{M_{\alpha'}} N^2 \left(\frac{\ell_{\alpha'}}{q} \right)^{2p} = N^{1+2\alpha'-2p(\alpha'-\kappa)} \leq N^{1+2\alpha'-p/30} \leq N^{2-61/30} \rightarrow 0. \quad (6.47)$$

By Lemma 6.8 and Markov's inequality, the third term on the right side tends to zero in probability. Since $\mathcal{P}(\alpha)$ holds, Lemma 6.11 shows that so do the first two terms, which proves (6.46). $\square$

6.4 Iteration and conclusion

We now iterate Proposition 6.7, starting from Proposition 6.5.

Corollary 6.12. *For every $\alpha \in (0, 1/2)$, the statement $\mathcal{P}(\alpha)$ holds.*

Proof. We define $\alpha^{(0)} = 1/10$ and

$$\alpha^{(n+1)} = \alpha^{(n)} + \frac{1}{10} \min \left\{ \alpha^{(n)}, \frac{1}{2} - \alpha^{(n)} \right\} \quad \text{for } n \geq 0. \quad (6.48)$$

By induction on n , we have $\alpha^{(n)} \in [1/10, 1/2)$ for all n . Indeed, if $\alpha^{(n)} \in [1/10, 1/2)$, then $\alpha^{(n+1)} > \alpha^{(n)}$ and

$$\alpha^{(n+1)} \leq \alpha^{(n)} + \frac{1}{10} \left(\frac{1}{2} - \alpha^{(n)} \right) < \frac{1}{2}.$$

The sequence $(\alpha^{(n)})$ is therefore increasing and bounded by $1/2$. Its limit $\alpha^{(\infty)} \in (1/10, 1/2]$ satisfies $\min\{\alpha^{(\infty)}, 1/2 - \alpha^{(\infty)}\} = 0$ by (6.48), and hence $\alpha^{(n)} \rightarrow 1/2$.

We claim that for all $n \geq 0$ and all $\alpha' \in (\alpha^{(n)}, \alpha^{(n+1)}]$, the pair $(\alpha^{(n)}, \alpha')$ satisfies (6.22). We have $1/10 \leq \alpha^{(n)} < \alpha' \leq \alpha^{(n+1)} < 1/2$. Further, (6.48) gives

$$\alpha' - \alpha^{(n)} \leq \frac{\alpha^{(n)}}{10} < \frac{\alpha^{(n)}}{5}, \quad \alpha' - \alpha^{(n)} \leq \frac{1}{10} \left(\frac{1}{2} - \alpha^{(n)} \right) < \frac{1}{3} \left(\frac{1}{2} - \alpha^{(n)} \right),$$

so that

$$\alpha' < \frac{6\alpha^{(n)}}{5}, \quad \alpha' < \alpha^{(n)} + \frac{1}{6} - \frac{\alpha^{(n)}}{3} = \frac{1}{6} + \frac{2\alpha^{(n)}}{3},$$

as claimed.

We have $\alpha^{(0)} < 1/8$, so Proposition 6.5 shows that $\mathcal{P}(\alpha^{(0)})$ holds. Assuming that $\mathcal{P}(\alpha^{(n)})$ holds, we may apply Proposition 6.7 with $(\alpha, \alpha') = (\alpha^{(n)}, \alpha^{(n+1)})$, and we find that $\mathcal{P}(\alpha^{(n+1)})$ holds. Hence $\mathcal{P}(\alpha^{(n)})$ holds for all $n \geq 0$. Now let $\alpha' \in (0, 1/2)$. When $\alpha' \leq \alpha^{(0)}$, the statement $\mathcal{P}(\alpha')$ holds by Proposition 6.5. Otherwise, as $\alpha^{(n)} \rightarrow 1/2$, there exists $n \geq 0$ such that $\alpha^{(n)} < \alpha' \leq \alpha^{(n+1)}$, and Proposition 6.7 with $(\alpha, \alpha') = (\alpha^{(n)}, \alpha')$ shows that $\mathcal{P}(\alpha')$ holds. $\square$

Proof of Theorem 1.3. Fix $\alpha \in (0, 1/2)$, a continuous function φ with compact support in $(-2, 2)$, and a continuous function $f: [0, \infty) \rightarrow \mathbb{R}$ with compact support. Recall that $N^{2\alpha-1} = M_\alpha^{-1}$. By Corollary 6.12, $\mathcal{P}(\alpha)$ holds, so Lemma 6.6 applies. Since f has compact support, it satisfies the hypothesis of Lemma 6.6, and we fix $R \geq 1$ such that $\text{supp } f \subset [0, R]$.

Suppose first that $\varphi \in C_c^\infty((-2, 2))$ is nonnegative. We split the sum defining $\mathcal{X}_\alpha(H; \varphi, f)$ into the terms with $j = i + 1$ and the terms with $j \geq i + 2$, which gives

$$\mathcal{X}_\alpha(H; \varphi, f) = \sum_{i=1}^{N-1} \varphi(E_i) f\left(\frac{g_i}{\ell_\alpha}\right) + \sum_{\substack{1 \leq i, j \leq N \\ j \geq i+2}} \varphi\left(\frac{\lambda_i + \lambda_j}{2}\right) f\left(\frac{\lambda_j - \lambda_i}{\ell_\alpha}\right). \quad (6.49)$$

To show that the second sum is negligible, we count its nonzero terms. A term in this sum vanishes unless $\lambda_j - \lambda_i \leq R\ell_\alpha$. Let $N \geq N_0$, so that $R\ell_\alpha \leq N^{-1}$, and suppose that $H \in \mathcal{G}_{\alpha/2}$. For each i , there are at most $N^{\alpha/2}$ labels j with $\lambda_j \in [\lambda_i, \lambda_i + N^{-1}]$. A label $j \geq i + 2$ with $\lambda_j - \lambda_i \leq R\ell_\alpha$ exists only if $\lambda_{i+2} - \lambda_i \leq R\ell_\alpha$. It follows that

$$\#\{(i, j) : 1 \leq i < j \leq N, j \geq i + 2, \lambda_j - \lambda_i \leq R\ell_\alpha\} \leq N^{\alpha/2} \mathcal{N}_3(R\ell_\alpha). \quad (6.50)$$

Using Proposition 3.4 with $r = R\ell_\alpha \geq N^{-2}$, we have

$$\frac{1}{M_\alpha} \mathbb{E}[N^{\alpha/2} \mathcal{N}_3(R\ell_\alpha)] \leq \frac{CN^{\alpha/2}}{M_\alpha} N(NR\ell_\alpha)^3 = CR^3 N^{-\alpha/2} \rightarrow 0.$$

The second sum in (6.49) is bounded by $\|\varphi\|_\infty \|f\|_\infty$ times the left side of (6.50). By Markov's inequality and Proposition 3.2, it is therefore $o_{\mathbb{P}}(M_\alpha)$. Combining this with Lemma 6.6 for the unsmoothed weight φ , we obtain

$$\frac{1}{M_\alpha} \sum_{i=1}^{N-1} \varphi(E_i) f\left(\frac{g_i}{\ell_\alpha}\right) \rightarrow \int_{-2}^2 \varphi(E) c(E) dE \int_0^\infty y f(y) dy \quad \text{in probability.} \quad (6.51)$$

Next, we treat a general continuous φ with compact support in $(-2, 2)$. The positive and negative parts of φ are continuous, nonnegative, and compactly supported in $(-2, 2)$, and both sides of (6.51) are linear in φ . We may therefore assume that $\varphi \geq 0$.

Let $\delta > 0$ satisfy $\text{supp } \varphi \subset [-2 + 2\delta, 2 - 2\delta]$, and set $I = [-2 + \delta, 2 - \delta]$. Let $\omega > 0$. By convolving φ with a smooth mollifier supported in a sufficiently small interval $[-\delta', \delta'] \subset$

$[-\delta, \delta]$, we obtain a nonnegative $\tilde{\varphi} \in C_c^\infty((-2, 2))$ with $\text{supp } \tilde{\varphi} \subset I$ and $\|\varphi - \tilde{\varphi}\|_\infty \leq \omega$. We set

$$C_f = \sup_{y \geq 0} (1 + y^2)^{p_*} |f(y)| \leq (1 + R^2)^{p_*} \|f\|_\infty.$$

The function $\varphi - \tilde{\varphi}$ vanishes outside I , so

$$\left| \sum_{i=1}^{N-1} (\varphi - \tilde{\varphi})(E_i) f\left(\frac{g_i}{\ell_\alpha}\right) \right| \leq \omega C_f \sum_{i: E_i \in I} f_{p_*}\left(\frac{g_i}{\ell_\alpha}\right) \leq \omega C_f \sum_{\substack{1 \leq i < j \leq N \\ (\lambda_i + \lambda_j)/2 \in I}} f_{p_*}\left(\frac{\lambda_j - \lambda_i}{\ell_\alpha}\right).$$

By (6.20), the right side is bounded by $\omega C_f C_I M_\alpha$ with probability tending to one.

Further, the limits in (6.51) for φ and $\tilde{\varphi}$ differ by at most $\omega \int_{-2}^2 c(E) dE \int_0^\infty y |f(y)| dy$. Applying (6.51) to $\tilde{\varphi}$, we find that with probability tending to one,

$$\begin{aligned} \left| \frac{1}{M_\alpha} \sum_{i=1}^{N-1} \varphi(E_i) f\left(\frac{g_i}{\ell_\alpha}\right) - \int_{-2}^2 \varphi(E) c(E) dE \int_0^\infty y f(y) dy \right| \\ \leq \omega \left(1 + C_f C_I + \int_{-2}^2 c(E) dE \int_0^\infty y |f(y)| dy \right). \end{aligned}$$

The constant C_I does not depend on ω , and $\omega > 0$ is arbitrary. This completes the proof. $\square$

Proof of Proposition 4.1. Recall that $\alpha_* = 99/200$, $\ell = \ell_{\alpha_*}$, and $M_{\alpha_*} = \tau$. By Corollary 6.12, $\mathcal{P}(\alpha_*)$ holds. The second assertion of Proposition 4.1 is (6.20) with $\alpha = \alpha_*$.

For the first assertion, let $f \in C_c([0, \infty) \times (-2, 2))$. Let $R \geq 1$ and $\delta \in (0, 1)$ satisfy $\text{supp } f \subset [0, R] \times [-2 + 2\delta, 2 - 2\delta]$, and set $I = [-2 + \delta, 2 - \delta]$. Let $\omega > 0$. The function f is uniformly continuous, so there exists $\delta' \in (0, \delta)$ such that $|f(r, E) - f(r, E')| \leq \omega$ for all $r \geq 0$ and all E, E' with $|E - E'| \leq \delta'$.

We approximate f by a sum of products. Using a standard smooth partition of unity, we choose nonnegative functions $\chi_1, \dots, \chi_n \in C_c^\infty((-2 + \delta, 2 - \delta))$ such that each $\text{supp } \chi_k$ has diameter at most δ' and

$$\sum_{k=1}^n \chi_k \leq 1 \quad \text{on } \mathbb{R}, \quad \sum_{k=1}^n \chi_k = 1 \quad \text{on } [-2 + 2\delta, 2 - 2\delta]. \quad (6.52)$$

For each k , we choose a point $E'_k \in \text{supp } \chi_k$, and we set

$$\bar{f}(r, E) = \sum_{k=1}^n \chi_k(E) f(r, E'_k).$$

We claim that

$$|f(r, E) - \bar{f}(r, E)| \leq \omega (1 + R^2)^{p_*} f_{p_*}(r) \mathbf{1}_I(E) \quad \text{for all } (r, E) \in [0, \infty) \times (-2, 2). \quad (6.53)$$

When $E \in [-2 + 2\delta, 2 - 2\delta]$, the second relation in (6.52) gives

$$f(r, E) - \bar{f}(r, E) = \sum_{k=1}^n \chi_k(E) (f(r, E) - f(r, E'_k)).$$

When $E \notin [-2 + 2\delta, 2 - 2\delta]$, we have $f(r, E) = 0$, and the same identity holds. In either case, $\chi_k(E) \neq 0$ implies $|E - E'_k| \leq \delta'$, so

$$|f(r, E) - \bar{f}(r, E)| \leq \omega \sum_{k=1}^n \chi_k(E) \leq \omega.$$

The function $f - \bar{f}$ vanishes unless $r \leq R$ and $E \in I$, which gives (6.53).

By (6.53) and (6.20) with $\alpha = \alpha_*$, we have with probability tending to one

$$\frac{1}{\tau} \left| \sum_{1 \leq i < j \leq N} (f - \bar{f}) \left(\frac{\lambda_j - \lambda_i}{\ell}, \frac{\lambda_i + \lambda_j}{2} \right) \right| \leq \omega (1 + R^2)^{p_*} C_I.$$

On the other hand, for each k the function $r \mapsto f(r, E'_k)$ is continuous with compact support in $[0, \infty)$, and

$$\sum_{1 \leq i < j \leq N} \bar{f} \left(\frac{\lambda_j - \lambda_i}{\ell}, \frac{\lambda_i + \lambda_j}{2} \right) = \sum_{k=1}^n \mathcal{X}_{\alpha_*}(H; \chi_k, f(\cdot, E'_k)).$$

We apply Lemma 6.6 with $\alpha = \alpha_*$ and the unsmoothed weights χ_k , and we obtain

$$\begin{aligned} \frac{1}{\tau} \sum_{1 \leq i < j \leq N} \bar{f} \left(\frac{\lambda_j - \lambda_i}{\ell}, \frac{\lambda_i + \lambda_j}{2} \right) &\rightarrow \sum_{k=1}^n \int_{-2}^2 \chi_k(E) c(E) dE \int_0^\infty r f(r, E'_k) dr \\ &= \int_{-2}^2 \int_0^\infty c(E) r \bar{f}(r, E) dr dE \end{aligned}$$

in probability. Finally, (6.53) gives

$$\left| \int_{-2}^2 \int_0^\infty c(E) r (f - \bar{f})(r, E) dr dE \right| \leq \omega R^2 \int_{-2}^2 c(E) dE.$$

The constant C_I does not depend on ω , and $\omega > 0$ is arbitrary, which proves the first assertion. $\square$

Actual and linearized gaps

The goal of this section is to prove Proposition 4.5, which compares the gaps of the resampled matrix $\tilde{H}$ with the linearized gaps $2d|z_i + S_i|$. Recall from Lemma 4.3 that, when the selected positions are distinct, the linearized gap of the pair $(i, i+1)$ is the gap of the compression of $\tilde{H}$ to the span of u_i and u_{i+1} . Heuristically, when g_i is much smaller than the distance from λ_i and λ_{i+1} to the other eigenvalues of H , the gap $\tilde{g}_i$ should be close to the linearized gap, since the remaining eigenvectors of H interact with the pair only at second order in Δ .

The proof has two steps. In Section 7.1, we prove a deterministic statement, Corollary 7.2, by applying Lemma 3.8 in the eigenbasis of H ; the main task is to verify its hypotheses with explicit exponents. In Section 7.2, we show that these hypotheses hold with probability tending to one.

Throughout this section, we set $\varepsilon = 1/100$. Recall that $d = N^{-3/2}$, $\tau = N^{1/100}$, $s = N^{-7/5}$, and $m = 2\lfloor \tau/8 \rfloor \leq \tau/4$.

7.1 A deterministic comparison

In Section 7.1, H denotes a deterministic real symmetric $N \times N$ matrix with eigenvalues $\lambda_1 \leq \dots \leq \lambda_N$, and $u_1, \dots, u_N$ denotes an orthonormal basis of $\mathbb{R}^N$ such that $Hu_j = \lambda_j u_j$ for all $j \in [N]$. We define g_i and E_i for $i \in [N-1]$, and the maps L_i , by the formulas of Section 4, using these eigenvalues and eigenvectors.

Definition 7.1. Let $\mathfrak{K}(H)$ denote the set of equivalence classes of the smallest equivalence relation on $[N]$ under which j and $j+1$ are equivalent for every $j \in [N-1]$ such that $g_j \leq s$. We call the elements of $\mathfrak{K}(H)$ the *clusters* of H .

Observe that every cluster is a set of consecutive integers, and that for every $j \in [N-1]$, the labels j and $j+1$ lie in the same cluster if and only if $g_j \leq s$. In particular, $\mathfrak{K}(H)$ is a family of clusters in the sense of Section 3.2.

We let $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$, and we let $U = (u_1, \dots, u_N)$ denote the orthogonal $N \times N$ matrix with columns $u_1, \dots, u_N$, so that $H = U\Lambda U^T$. For $\mathcal{K} \in \mathfrak{K}(H)$, we let $U_{\mathcal{K}}$ denote the $N \times |\mathcal{K}|$ matrix whose columns are the vectors u_j with $j \in \mathcal{K}$, in increasing order of j , and we set $\Lambda_{\mathcal{K}} = \text{diag}(\lambda_j : j \in \mathcal{K})$. For an $N \times N$ matrix A , we use the notation $A_{\mathcal{K}\mathcal{K}} = (A_{jj'})_{j,j' \in \mathcal{K}}$ of Section 3.2.

Recall the quantities w_N , $\mathcal{L}_N$, and δ_N from (4.17). Since $\varepsilon = 1/100$, they are given by

$$w_N = \tau N^{-3/2+3\varepsilon}, \quad \mathcal{L}_N = 2s^{-1} \log N, \quad \delta_N = 8w_N^2 \mathcal{L}_N. \quad (7.1)$$

Inserting the values of τ , s , d , and ε , and using $\delta_N/d = 8(w_N/d)(w_N \mathcal{L}_N)$ and $w_N/d = \tau N^{3\varepsilon}$, we find

$$\begin{aligned} \frac{w_N}{s} &= N^{-9/100+3\varepsilon} = N^{-3/50}, & w_N \mathcal{L}_N &= 2N^{-3/50} \log N, \\ \frac{\delta_N}{d} &= 16N^{-2/25+6\varepsilon} \log N = 16N^{-1/50} \log N. \end{aligned} \quad (7.2)$$

All three quantities in (7.2) tend to zero.

Corollary 7.2. *There exists an absolute constant $N_0 \geq 1$ such that the following holds for all $N \geq N_0$. Let H and $u_1, \dots, u_N$ be as above, and suppose that*

- (i) $\|H\| \leq 3$;
- (ii) $|u_j(a)| \leq N^{-1/2+\varepsilon}$ for all $a, j \in [N]$;
- (iii) $\lambda_{j+2} - \lambda_j > 4s$ for all $j \in [N-2]$.

Let Δ be a real symmetric $N \times N$ matrix such that $\Delta_{aa} = 0$ for all $a \in [N]$, $|\Delta_{ab}| \leq 2N^{-1/2+\varepsilon}$ for all $a, b \in [N]$, and $\Delta_{ab} \neq 0$ for at most $\tau/4$ pairs (a, b) with $a < b$. Then every cluster of H has at most two elements, and the two-element clusters of H are the sets $\{i, i+1\}$ with $i \in [N-1]$ and $g_i \leq s$. Further, the following statements hold.

- (a) Let $\mathcal{K} \in \mathfrak{K}(H)$. For all $j \in \mathcal{K}$, we have $\lambda_{\min \mathcal{K}} - s/4 \leq \lambda_j(H + \Delta) \leq \lambda_{\max \mathcal{K}} + s/4$. Moreover, the numbers $\lambda_j(H + \Delta)$ with $j \in \mathcal{K}$, in increasing order, differ by at most $\delta_N/2$ from the ordered eigenvalues of $\Lambda_{\mathcal{K}} + U_{\mathcal{K}}^T \Delta U_{\mathcal{K}}$.
- (b) For all $\mathcal{K} \in \mathfrak{K}(H)$, we have

$$\left| \sum_{j \in \mathcal{K}} (\lambda_j(H + \Delta) - \lambda_j) \right| \leq |\mathcal{K}|(w_N + \delta_N/2).$$

- (c) For all $j \in [N-1]$ such that $g_j > s$, we have $\lambda_{j+1}(H + \Delta) - \lambda_j(H + \Delta) > s/2$.
- (d) For all $i \in [N-1]$ such that $g_i \leq s$, we have

$$|\lambda_{i+1}(H + \Delta) - \lambda_i(H + \Delta) - 2|L_i(H + \Delta)|| \leq \delta_N.$$

Proof. We first describe the clusters. Suppose that some $\mathcal{K} \in \mathfrak{K}(H)$ has at least three elements. As $\mathcal{K}$ is a set of consecutive integers, there exists $j \in [N-2]$ such that $j, j+1, j+2 \in \mathcal{K}$. Then $g_j \leq s$ and $g_{j+1} \leq s$, which gives $\lambda_{j+2} - \lambda_j \leq 2s$ and contradicts (iii). Hence every cluster has at most two elements.

If $i \in [N-1]$ satisfies $g_i \leq s$, then i and $i+1$ lie in the same cluster, which is therefore $\{i, i+1\}$. Conversely, every two-element cluster has the form $\{i, i+1\}$ with $g_i \leq s$, since its elements are consecutive and lie in the same cluster.

Next, we rotate Δ into the eigenbasis of H . Set

$$W = U^T \Delta U, \quad \text{so that} \quad W_{jj'} = u_j^T \Delta u_{j'} \quad \text{for } j, j' \in [N].$$

Since U is orthogonal, we have $H + \Delta = U(\Lambda + W)U^T$, and therefore

$$\lambda_j(H + \Delta) = \lambda_j(\Lambda + W) \quad \text{for all } j \in [N].$$

As Δ is symmetric with zero diagonal, the triangle inequality and (ii) give

$$|W_{jj'}| = \left| \sum_{1 \leq a < b \leq N} \Delta_{ab} (u_j(a)u_{j'}(b) + u_j(b)u_{j'}(a)) \right| \leq \frac{\tau}{4} \cdot 2N^{-1/2+\varepsilon} \cdot 2N^{-1+2\varepsilon} = w_N \quad (7.3)$$

for all $j, j' \in [N]$. In the inequality, we used that at most $\tau/4$ of the coefficients Δ_{ab} with $a < b$ are nonzero, that each of them is bounded by $2N^{-1/2+\varepsilon}$, and that the term in absolute values is bounded by $2N^{-1+2\varepsilon}$ by (ii).

We will apply Lemma 3.8 with $D_y = \Lambda$ (so that $y_j = \lambda_j$), the family of clusters $\mathfrak{K}(H)$, and the parameters

$$k_0 = 2, \quad \vartheta = s/4, \quad w_0 = w_N, \quad \mathcal{L} = \mathcal{L}_N.$$

For $\mathcal{K} \in \mathfrak{K}(H)$, the rectangle (3.24) is then

$$\mathcal{R}_{\mathcal{K}} = \{w \in \mathbb{C} : \lambda_{\min \mathcal{K}} - s/4 \leq \operatorname{Re} w \leq \lambda_{\max \mathcal{K}} + s/4, |\operatorname{Im} w| \leq s/4\}.$$

We next check that these rectangles are pairwise disjoint. Let $\mathcal{K}, \mathcal{K}' \in \mathfrak{K}(H)$ satisfy $\max \mathcal{K} + 1 = \min \mathcal{K}'$. Then $g_{\max \mathcal{K}} > s$, so the intervals $\mathcal{R}_{\mathcal{K}} \cap \mathbb{R}$ and $\mathcal{R}_{\mathcal{K}'} \cap \mathbb{R}$ are at distance $g_{\max \mathcal{K}} - s/2 > s/2$.

We now bound the harmonic sum in (3.25). Fix $\mathcal{K} \in \mathfrak{K}(H)$ and $w \in \mathcal{R}_{\mathcal{K}}$, and write $i_- = \min \mathcal{K}$ and $i_+ = \max \mathcal{K}$. We first consider the labels $j > i_+$, assuming that $i_+ < N$. The labels i_+ and $i_+ + 1$ lie in different clusters, so $g_{i_+} > s$.

Let $r \geq 0$ be an integer, and let $j \in \{i_+ + 2r + 1, i_+ + 2r + 2\}$ with $j \leq N$. Then

$$\lambda_j - \operatorname{Re} w \geq \lambda_{i_+ + 2r + 1} - \lambda_{i_+} - s/4 \geq g_{i_+} + 4sr - s/4 > 3s/4 + 4sr.$$

The first inequality follows from $\lambda_j \geq \lambda_{i_+ + 2r + 1}$ and $\operatorname{Re} w \leq \lambda_{i_+} + s/4$, and the second follows from (iii), applied r times to the labels $i_+ + 1, i_+ + 3, \dots, i_+ + 2r - 1$. Further, (i) implies that $\lambda_j - \operatorname{Re} w < 7$, so only integers r with $4sr < 7$ occur. Every such r corresponds to at most two labels j , and $|\lambda_j - w| \geq |\lambda_j - \operatorname{Re} w|$, so

$$\sum_{j=i_++1}^N \frac{1}{|\lambda_j - w|} \leq \sum_{0 \leq r < 7/(4s)} \frac{2}{3s/4 + 4sr} \leq \frac{8}{3s} + \frac{1}{2s} \sum_{1 \leq r < 7/(4s)} \frac{1}{r} \leq \frac{1}{s} \left(\frac{19}{6} + \frac{1}{2} \log \frac{7}{4s} \right),$$

where we compared the harmonic sum with an integral in the last step.

The same argument, with the roles of the two sides exchanged, bounds the sum over $j < i_-$ by the same quantity. Recalling that $s = N^{-7/5}$, we conclude that

$$\sum_{j \notin \mathcal{K}} \frac{1}{|\lambda_j - w|} \leq \frac{1}{s} \left(\frac{19}{3} + \log \frac{7}{4} + \frac{7}{5} \log N \right) \leq \frac{2 \log N}{s} = \mathcal{L}_N \quad (7.4)$$

for $N \geq N_0$, uniformly over all clusters $\mathcal{K}$ and all $w \in \mathcal{R}_{\mathcal{K}}$.

We now apply Lemma 3.8. By (7.3), we have $\max_{j,j'} |W_{jj'}| \leq w_N$. Using (7.2), we see that $w_N \mathcal{L}_N \leq 1/2$ and

$$\frac{1}{s} (k_0 w_N + 2k_0 w_N^2 \mathcal{L}_N) = \frac{2w_N}{s} + \frac{4w_N}{s} \cdot w_N \mathcal{L}_N = 2N^{-3/50} + 8N^{-3/25} \log N < \frac{1}{4}$$

for $N \geq N_0$. Thus the hypotheses of Lemma 3.8 hold.

Let $\mathcal{K} \in \mathfrak{K}(H)$. Recalling that $\lambda_j(\Lambda + W) = \lambda_j(H + \Delta)$, we deduce from Lemma 3.8 that the numbers $\lambda_j(H + \Delta)$ with $j \in \mathcal{K}$ lie within distance $k_0 w_N + 2k_0 w_N^2 \mathcal{L}_N < s/4$ of $[\lambda_{\min \mathcal{K}}, \lambda_{\max \mathcal{K}}]$, and that, in increasing order, they differ by at most $2k_0 w_N^2 \mathcal{L}_N = \delta_N/2$ from the ordered eigenvalues of

$$(\Lambda + W)_{\mathcal{K}\mathcal{K}} = \Lambda_{\mathcal{K}} + U_{\mathcal{K}}^T \Delta U_{\mathcal{K}}.$$

This proves (a).

To prove (b), we use that the sum of the eigenvalues of a matrix is its trace. Then (a) implies that

$$\left| \sum_{j \in \mathcal{K}} \lambda_j(H + \Delta) - \text{Tr}(\Lambda_{\mathcal{K}} + U_{\mathcal{K}}^T \Delta U_{\mathcal{K}}) \right| \leq \frac{|\mathcal{K}| \delta_N}{2}, \quad \text{Tr}(\Lambda_{\mathcal{K}} + U_{\mathcal{K}}^T \Delta U_{\mathcal{K}}) = \sum_{j \in \mathcal{K}} (\lambda_j + W_{jj}).$$

Combining these relations with the bound $|W_{jj}| \leq w_N$ from (7.3), we obtain (b).

Next, we prove (c). Let $j \in [N-1]$ satisfy $g_j > s$. Then j and $j+1$ lie in different clusters $\mathcal{K}$ and $\mathcal{K}'$, with $j = \max \mathcal{K}$ and $j+1 = \min \mathcal{K}'$. Using (a) for $\mathcal{K}$ and $\mathcal{K}'$, we find

$$\lambda_{j+1}(H + \Delta) - \lambda_j(H + \Delta) \geq (\lambda_{j+1} - s/4) - (\lambda_j + s/4) = g_j - s/2 > s/2.$$

It remains to prove (d). Let $i \in [N-1]$ satisfy $g_i \leq s$, so that $\mathcal{K} = \{i, i+1\}$ is a cluster. Recalling that $u_i^T H u_i = \lambda_i$, $u_{i+1}^T H u_{i+1} = \lambda_{i+1}$, and $u_i^T H u_{i+1} = 0$, we have

$$\Lambda_{\mathcal{K}} + U_{\mathcal{K}}^T \Delta U_{\mathcal{K}} = (u_j^T (H + \Delta) u_{j'})_{j,j' \in \{i, i+1\}}.$$

By (4.10) and the definition of L_i , the two eigenvalues of this matrix differ by $2|L_i(H + \Delta)|$. By (a), the eigenvalues $\lambda_i(H + \Delta)$ and $\lambda_{i+1}(H + \Delta)$ differ by at most $\delta_N/2$ from the smaller and the larger of them, respectively, which gives (d). $\square$

7.2 Proof of Proposition 4.5

We now return to the random matrix H and its chosen eigenvectors $u_1, \dots, u_N$, and we apply Corollary 7.2 to their realizations. The clusters $\mathfrak{K}(H)$ are then random sets, defined by Definition 7.1 for each realization of H .

Proof of Proposition 4.5. Let N_0 be the constant of Corollary 7.2. For $N \geq N_0$, let $\mathcal{A}_N$ denote the event

$$\mathcal{A}_N = \{H \in \mathcal{G}_\varepsilon\} \cap \{\lambda_{i+2} - \lambda_i > 4s \text{ for all } i \in [N-2]\} \cap \{|X_{ab}| \leq N^\varepsilon \text{ for all } a, b \in [N] \text{ with } a < b\}, \quad (7.5)$$

and for $N < N_0$, set $\mathcal{A}_N = \emptyset$. The event $\mathcal{A}_N$ is measurable with respect to H (the condition $H \in \mathcal{G}_\varepsilon$ concerns every unit eigenvector of H , not only $u_1, \dots, u_N$). Further, let $\mathcal{B}_N$ denote the event

$$\mathcal{B}_N = \{(a_l, b_l) \neq (a_{l'}, b_{l'}) \text{ for all } l, l' \in [m] \text{ with } l \neq l'\} \cap \{|Y_l| \leq N^\varepsilon \text{ for all } l \in [m]\}, \quad (7.6)$$

which is measurable with respect to the resampling variables. We will show that $\mathbb{P}(\mathcal{A}_N) \rightarrow 1$, that $\mathbb{P}(\mathcal{B}_N \mid H) = 1 - o(1)$ uniformly in H , and that the conclusions of Proposition 4.5 hold on $\mathcal{A}_N \cap \mathcal{B}_N$. We may assume that $N \geq N_0$.

We begin with $\mathcal{A}_N$. Proposition 3.2 with $t = 0$ shows that $H \in \mathcal{G}_\varepsilon$ with overwhelming probability. By Markov's inequality and the second statement of Proposition 3.4 with $C_1 = 1$, which applies since $4s = 4N^{-7/5} \geq N^{-2}$, we have

$$\mathbb{P}\left(\min_{i \in [N-2]} (\lambda_{i+2} - \lambda_i) \leq 4s\right) \leq \mathbb{E}\mathcal{N}_3(4s) \leq CN(4Ns)^3 = 64CN^{-1/5}. \quad (7.7)$$

Note that (7.7) covers all $i \in [N-2]$, including labels at the spectral edges. This is needed, although only pairs in the bulk are used in Section 4, because Lemma 3.8 is applied to the whole spectrum.

Finally, the standardized entries X_{ab} with $a < b$ are copies of X , so the sub-Gaussian tail bound and a union bound over fewer than N^2 pairs give

$$\mathbb{P}(|X_{ab}| > N^\varepsilon \text{ for some } a, b \in [N] \text{ with } a < b) \leq C_0 N^2 e^{-c_0 N^{2\varepsilon}}.$$

Combining these three bounds, we obtain

$$\mathbb{P}(\mathcal{A}_N^c) \leq CN^{-1/5}.$$

Next, we bound the conditional probability of $\mathcal{B}_N^c$. The selected positions and the variables Y_l are independent of H , so $\mathbb{P}(\mathcal{B}_N \mid H) = \mathbb{P}(\mathcal{B}_N)$. For distinct $l, l' \in [m]$, the positions (a_l, b_l) and $(a_{l'}, b_{l'})$ are independent and uniformly distributed on a set of

$N(N-1)/2$ pairs, so they coincide with probability $2/(N(N-1))$. A union bound over the fewer than $m^2/2$ pairs $\{l, l'\}$ and over $l \in [m]$, together with $m \leq \tau/4$, gives

$$\mathbb{P}(\mathcal{B}_N^c \mid H) \leq \frac{m^2}{N(N-1)} + mC_0e^{-c_0N^{2\varepsilon}} \leq \frac{\tau^2}{8N^2} + C_0\tau e^{-c_0N^{2\varepsilon}}. \quad (7.8)$$

The right side of (7.8) is deterministic and tends to zero.

We now fix a realization in $\mathcal{A}_N \cap \mathcal{B}_N$ and verify the hypotheses of Corollary 7.2. As $H \in \mathcal{G}_\varepsilon$, we have $\|H\| \leq 3$, and every unit eigenvector of H , in particular each u_j , satisfies $\|u_j\|_\infty \leq N^{-1/2+\varepsilon}$. Hypothesis (iii) holds by the definition of $\mathcal{A}_N$. The selected positions are distinct on $\mathcal{B}_N$, so Lemma 4.3 gives

$$\Delta = \sum_{l=1}^m N^{-1/2} (Y_l - X_{a_l b_l}) V^{a_l b_l}.$$

Hence Δ is real symmetric with zero diagonal, $\Delta_{ab} \neq 0$ for at most $m \leq \tau/4$ pairs (a, b) with $a < b$, and

$$|\Delta_{a_l b_l}| \leq N^{-1/2} (|Y_l| + |X_{a_l b_l}|) \leq 2N^{-1/2+\varepsilon} \quad \text{for all } l \in [m],$$

where we used the bounds on Y_l and $X_{a_l b_l}$ contained in $\mathcal{B}_N$ and $\mathcal{A}_N$. Thus Corollary 7.2 applies to H and Δ , with $H + \Delta = \tilde{H}$.

By Corollary 7.2, every cluster of H has at most two elements, and the two-element clusters are the sets $\{i, i+1\}$ with $g_i \leq s$. Statement (c) of Corollary 7.2 gives $\tilde{g}_j > s/2$ for all $j \in [N-1]$ with $g_j > s$.

Now let $i \in [N-1]$ satisfy $g_i \leq s$, so that $\{i, i+1\}$ is a cluster. By Lemma 4.3, we have $L_i(\tilde{H}) = d(z_i + S_i)$. Statement (d) of Corollary 7.2, statement (b) with $\mathcal{K} = \{i, i+1\}$, and (7.2) therefore give

$$\left| \frac{\tilde{g}_i}{d} - 2|z_i + S_i| \right| \leq \frac{\delta_N}{d} = 16N^{-1/50} \log N, \quad |\tilde{E}_i - E_i| \leq w_N + \frac{\delta_N}{2}. \quad (7.9)$$

Both bounds in (7.9) are deterministic and tend to zero by (7.2). Together with $\mathbb{P}(\mathcal{A}_N) \rightarrow 1$ and (7.8), this establishes all assertions of Proposition 4.5. $\square$

A local limit theorem for the linearized increments

The goal of this section is to prove Proposition 4.4, whose normalization is explained heuristically in Section 2.2. Recall from Section 4 that, for a k -tuple $\mathbf{i}$, the vector $S_{\mathbf{i}}$ is a sum of $m = 2\lfloor \tau/8 \rfloor$ increments that are independent and identically distributed conditionally on H . The central limit theorem describes $S_{\mathbf{i}}$ on the scale $\tau^{1/2}$, but Proposition 4.4 concerns the probability that $S_{\mathbf{i}} + x$ belongs to a bounded set, which is a statement on the unit scale. For sums of independent copies of a fixed law, local limit theorems of this kind hold when the law is nonlattice [67]. Here the law of an increment depends on N and on H , and the main task of this section is to show that it stays away from lattice laws, uniformly over H outside an event of small probability.

This is a genuine issue when X is discrete. Suppose, for example, that X takes the values 1 and -1 , so that $Y - X_{ab} \in \{-2, 0, 2\}$, and that every coordinate of u_i and u_{i+1} has modulus $N^{-1/2}$. Then every $D_{ab}^{(i)}$ belongs to $\{(\pm 2, 0), (0, \pm 2)\}$, and S_i takes values in the lattice $4\mathbb{Z}^2$. The conclusion of Proposition 4.4 would then fail for test functions supported near $(2, 2)$, although such eigenvectors satisfy the delocalization bound in the definition of $\mathcal{G}_\varepsilon$.

After computing the moments of one increment in Section 8.1, we show in Section 8.2 that with overwhelming probability every unit eigenvector of H is far, in a normalized ℓ^2 norm, from all vectors whose coordinates take at most a small power of τ distinct values. In Section 8.3, we show by a deterministic argument that this property prevents the phases in the characteristic function of one increment from concentrating near $2\pi\mathbb{Z}$. A symmetrization in Section 8.4 then bounds the squared modulus of this characteristic function away from 1 at all frequencies between a small negative power and a small positive power of N . With the parameters chosen in Section 8.5, this bound reads $1 - c\tau^{-1/4}$. This bound is the main input to the local limit theorem for sums of increments, Lemma 8.13, which we prove in Section 8.5 by Fourier inversion. Finally, in Section 8.6 we add a large deviation estimate for pairs with large gaps and complete the proof of Proposition 4.4. Throughout this section, we fix an integer $k \geq 1$.

8.1 Notation and moments of one increment

Throughout Section 8, $\mathbf{i} = (i_1, \dots, i_k)$ denotes a k -tuple of integers in $[N - 1]$ such that $i_{j+1} \geq i_j + 2$ for all $j \in [k - 1]$. For $w \in \mathbb{R}^{2k}$, we let $w_j = (w_{j,1}, w_{j,2}) \in \mathbb{R}^2$ denote its

j -th block, so that $w = (w_1, \dots, w_k)$, and we let $w(\eta)$ denote its η -th coordinate, where $\eta \in [2k]$. We let U_i denote the $N \times 2k$ matrix with columns $u_{i_1}, u_{i_1+1}, \dots, u_{i_k}, u_{i_k+1}$, in this order. Since these $2k$ indices are distinct, we have $U_i^T U_i = \text{Id}_{2k}$. For $b \in [N]$, let $r_b \in \mathbb{R}^{2k}$ denote the b -th row of $N^{1/2} U_i$, so that $r_{b,j} = N^{1/2}(u_{i_j}(b), u_{i_j+1}(b))$. The relation $U_i^T U_i = \text{Id}_{2k}$ is equivalent to

$$\frac{1}{N} \sum_{b=1}^N r_b r_b^T = \text{Id}_{2k}. \quad (8.1)$$

For $\xi \in \mathbb{R}^{2k}$, let $\Gamma(\xi)$ denote the block-diagonal $2k \times 2k$ matrix whose j -th diagonal block is

$$\begin{pmatrix} \xi_{j,1} & \xi_{j,2} \\ \xi_{j,2} & -\xi_{j,1} \end{pmatrix}. \quad (8.2)$$

For $k = 1$, this defines $\Gamma(\xi_j)$ for $\xi_j \in \mathbb{R}^2$, and we have $\Gamma(\xi) = \text{diag}(\Gamma(\xi_1), \dots, \Gamma(\xi_k))$. The matrix $\Gamma(\xi_j)$ is symmetric with eigenvalues $|\xi_j|$ and $-|\xi_j|$. Hence $\|\Gamma(\xi)\| = \max_j |\xi_j| \leq |\xi|$, and $\Gamma(\xi_j)$ is invertible with $\|\Gamma(\xi_j)^{-1}\| = |\xi_j|^{-1}$ when $\xi_j \neq 0$. The map $\xi \mapsto \Gamma(\xi)$ is linear, and we let Γ_η denote its value at the η -th standard basis vector of $\mathbb{R}^{2k}$, so that $\Gamma(\xi) = \sum_{\eta=1}^{2k} \xi(\eta) \Gamma_\eta$.

Recall the vectors $D_{ab}^{(i)} \in \mathbb{R}^2$ from (4.11). In Section 8, we use the formula on the right side of (4.11) to define $D_{ab}^{(i)}$ for all $a, b \in [N]$, so that $D_{ab}^{(i)} = D_{ba}^{(i)}$, and we let $D_{ab}^{(i)} \in \mathbb{R}^{2k}$ denote the vector with blocks $D_{ab}^{(i_1)}, \dots, D_{ab}^{(i_k)}$. Expanding the quadratic forms on the right side and comparing with the definition of $D_{ab}^{(i_j)}$, we find

$$\xi \cdot D_{ab}^{(i)} = r_a^T \Gamma(\xi) r_b, \quad a, b \in [N], \quad \xi \in \mathbb{R}^{2k}. \quad (8.3)$$

Further, after identifying $\mathbb{R}^2$ with $\mathbb{C}$, we have

$$D_{ab}^{(i_j)} = N(u_{i_j}(a) + i u_{i_j+1}(a))(u_{i_j}(b) + i u_{i_j+1}(b)), \quad (8.4)$$

and therefore

$$|D_{ab}^{(i_j)}| = |r_{a,j}| |r_{b,j}|, \quad a, b \in [N], \quad j \in [k]. \quad (8.5)$$

Definition 8.1. Let (a, b) be a random pair with the uniform distribution on $\{(a', b') \in [N]^2 : a' < b'\}$, and let Y be a copy of X , such that H , (a, b) , and Y are independent. We define the *increment*

$$J = (Y - X_{ab}) D_{ab}^{(i)} \in \mathbb{R}^{2k}. \quad (8.6)$$

We define the conditional mean $\bar{J} = \mathbb{E}[J \mid H]$, the conditional covariance $\Sigma_H = \mathbb{E}[(J - \bar{J})(J - \bar{J})^T \mid H]$, and the conditional characteristic function

$$\Phi_H(\xi) = \mathbb{E}[e^{i\xi \cdot J} \mid H], \quad \xi \in \mathbb{R}^{2k}. \quad (8.7)$$

For an integer $n \geq 1$, let $S^{(n)}$ denote a random vector in $\mathbb{R}^{2k}$ whose conditional law given H is the law of a sum of n independent random vectors, each with the conditional law of J given H . We let Φ_Y denote the characteristic function of Y , so that $\Phi_Y(t') = \mathbb{E}e^{it'Y}$ for $t' \in \mathbb{R}$.

These objects depend on $\mathbf{i}$, which we suppress from the notation. By the definition of $S^{(n)}$, we have

$$\mathbb{E}[e^{i\zeta \cdot S^{(n)}} \mid H] = \Phi_H(\zeta)^n, \quad \zeta \in \mathbb{R}^{2k}. \quad (8.8)$$

Recall from Section 4 that $S_{\mathbf{i}} = \sum_{l=1}^m (Y_l - X_{a_l b_l}) D_{a_l b_l}^{(\mathbf{i})}$, where $X_{a_l b_l}$ is the entry of H for every l , also when a position is selected more than once. Conditionally on H , the m summands are therefore independent, and each has the conditional law of J . Hence the conditional law of $S_{\mathbf{i}}$ given H is that of $S^{(m)}$.

We next introduce the event on which the estimates below take place. It involves the matrix $\hat{X}$ below, which records the fluctuations of the squared entries; these fluctuations appear in the covariance of J .

Definition 8.2. Let $\hat{X}$ denote the $N \times N$ real symmetric matrix with zero diagonal and entries $\hat{X}_{ab} = X_{ab}^2 - 1$ for $a \neq b$. For $\varepsilon > 0$, we let $\mathcal{A}_N^{\text{reg}}(\varepsilon)$ denote the event that

$$H \in \mathcal{G}_\varepsilon, \quad \max_{a \leq b} |X_{ab}| \leq N^\varepsilon, \quad \|\hat{X}\| \leq N^{1/2+\varepsilon}. \quad (8.9)$$

Recall from Definition 3.1 that $H \in \mathcal{G}_\varepsilon$ implies $\|H\| \leq 3$ and $\|v\|_\infty \leq N^{-1/2+\varepsilon}$ for every unit eigenvector v of H . Hence, on $\mathcal{A}_N^{\text{reg}}(\varepsilon)$, we have for every k -tuple $\mathbf{i}$ and all $a, b \in [N]$ that

$$|r_{b,j}| \leq 2^{1/2} N^\varepsilon, \quad |r_b| \leq (2k)^{1/2} N^\varepsilon, \quad |D_{ab}^{(\mathbf{i})}| \leq 2k^{1/2} N^{2\varepsilon}, \quad (8.10)$$

where the last bound follows from (8.5).

Lemma 8.3. *For every $\varepsilon > 0$, the event $\mathcal{A}_N^{\text{reg}}(\varepsilon)$ holds with overwhelming probability.*

Proof. The three conditions in (8.9) become weaker as ε increases, so we may assume that $\varepsilon \leq 1/4$. By Proposition 3.2, we have $H \in \mathcal{G}_\varepsilon$ with overwhelming probability. The sub-Gaussian tail bound for X and Z and a union bound over the at most N^2 entries give

$$\mathbb{P}\left(\max_{a \leq b} |X_{ab}| > N^{\varepsilon/4}\right) \leq C_0 N^2 \exp(-c_0 N^{\varepsilon/2}), \quad (8.11)$$

which is smaller than every negative power of N for large N . This gives the second condition in (8.9).

To bound $\|\hat{X}\|$, we truncate the entries at $N^{\varepsilon/4}$ and center them. Set

$$c_N = \mathbb{E}[(X^2 - 1)\mathbf{1}\{|X| \leq N^{\varepsilon/4}\}] = -\mathbb{E}[(X^2 - 1)\mathbf{1}\{|X| > N^{\varepsilon/4}\}], \quad (8.12)$$

where we used $\mathbb{E}X^2 = 1$. The sub-Gaussian tail of X gives

$$|c_N| \leq C \exp(-cN^{\varepsilon/2}). \quad (8.13)$$

Recall that $V^{ab} = e_a e_b^T + e_b e_a^T$ for $a < b$, and define

$$A_{ab} = \left((X_{ab}^2 - 1) \mathbf{1}\{|X_{ab}| \leq N^{\varepsilon/4}\} - c_N \right) V^{ab}, \quad \dot{X}' = \sum_{a < b} A_{ab}. \quad (8.14)$$

The matrices A_{ab} are independent, centered, and real symmetric. Since $\|V^{ab}\| = 1$ and $|c_N| \leq 1$ for large N , we have $\|A_{ab}\| \leq 4N^{\varepsilon/2}$. Further, $(V^{ab})^2 = e_a e_a^T + e_b e_b^T$, so $\sum_{a < b} \mathbb{E}A_{ab}^2$ is a diagonal matrix, and

$$\left\| \sum_{a < b} \mathbb{E}A_{ab}^2 \right\| \leq (N-1) \mathbb{E}[(X^2 - 1)^2] \leq CN. \quad (8.15)$$

We now use the matrix Bernstein inequality [73, Theorem 1.4], which we state in the form needed here. Let $(A_{ab})_{a < b}$ be independent, centered, real symmetric random $N \times N$ matrices such that $\lambda_N(A_{ab}) \leq R'$ almost surely for all $a < b$. Then for all $t' \geq 0$,

$$\mathbb{P}\left(\lambda_N\left(\sum_{a < b} A_{ab}\right) \geq t'\right) \leq N \exp\left(-\frac{t'^2/2}{\left\|\sum_{a < b} \mathbb{E}A_{ab}^2\right\| + R't'/3}\right). \quad (8.16)$$

We apply it to the matrices A_{ab} and to their negatives, with $R' = 4N^{\varepsilon/2}$ and $t' = N^{1/2+\varepsilon}/2$. Our assumption $\varepsilon \leq 1/4$ gives $R't' \leq 2N$, and we obtain

$$\mathbb{P}\left(\|\dot{X}'\| \geq \frac{N^{1/2+\varepsilon}}{2}\right) \leq 2N \exp(-cN^{2\varepsilon}). \quad (8.17)$$

When $\max_{a < b} |X_{ab}| \leq N^{\varepsilon/4}$, we have

$$\dot{X} = \dot{X}' + c_N \sum_{a < b} V^{ab}, \quad \text{and hence} \quad \|\dot{X}\| \leq \|\dot{X}'\| + N^2 |c_N| \leq \|\dot{X}'\| + CN^2 e^{-cN^{\varepsilon/2}}. \quad (8.18)$$

Combining the last two displays with (8.11), we conclude that $\|\dot{X}\| \leq N^{1/2+\varepsilon}$ with overwhelming probability. $\square$

For $v, v' \in \mathbb{R}^N$, we let $v \odot v' \in \mathbb{R}^N$ denote the vector with coordinates $(v \odot v')(a) = v(a)v'(a)$. The following lemma identifies the first three conditional moments of one increment.

Lemma 8.4. *Let $0 < \varepsilon \leq 1/8$. There exists $C > 0$ such that on $\mathcal{A}_N^{\text{reg}}(\varepsilon)$, for every k -tuple $\mathbf{i}$,*

$$\begin{aligned} |\bar{J}| &\leq CN^{-1/2}, & \mathbb{E}[|J|^3 \mid H] &\leq CN^{9\varepsilon}, \\ \|\mathbb{E}[JJ^T \mid H] - 4\text{Id}_{2k}\| + \|\Sigma_H - 4\text{Id}_{2k}\| &\leq CN^{3\varepsilon-1/2}. \end{aligned} \quad (8.19)$$

Moreover, on $\mathcal{A}_N^{\text{reg}}(\varepsilon)$ we have $|J| \leq 2k^{1/2}N^{2\varepsilon}(|Y| + N^\varepsilon)$.

Proof. The last claim follows from (8.6), (8.10), and $|X_{ab}| \leq N^\varepsilon$. We begin with the mean. Since $\mathbb{E}Y = 0$ and Y is independent of H and (a, b) , we have

$$\bar{J} = -\binom{N}{2}^{-1} \sum_{a < b} X_{ab} D_{ab}^{(i)}. \quad (8.20)$$

The main point is that the sum in (8.20) can be computed exactly with the eigenvalue equation; a bound by absolute values would give only $|\bar{J}| \leq CN^{3\varepsilon}$, while the drift $m\bar{J}$ must be small on the unit scale. Fix $j \in [k]$, and abbreviate $u = u_{i_j}$ and $v = u_{i_{j+1}}$. Recall that $X_{ba} = X_{ab}$ and $X_{aa} = N^{1/2}H_{aa}$, so that

$$\sum_{a,b=1}^N X_{ab} u(a)u(b) = N^{1/2} u^T H u = N^{1/2} \lambda_{i_j}. \quad (8.21)$$

Separating the diagonal terms, we obtain

$$\sum_{a < b} X_{ab} u(a)u(b) = \frac{1}{2} \left(N^{1/2} \lambda_{i_j} - \sum_{a=1}^N X_{aa} u(a)^2 \right), \quad (8.22)$$

and the same identity holds with v and $\lambda_{i_{j+1}}$ in place of u and λ_{i_j} . Similarly, since $Hv = \lambda_{i_{j+1}}v$ and $u^T v = 0$, we have

$$\sum_{a < b} X_{ab} (u(a)v(b) + v(a)u(b)) = N^{1/2} u^T H v - \sum_{a=1}^N X_{aa} u(a)v(a) = - \sum_{a=1}^N X_{aa} u(a)v(a). \quad (8.23)$$

Inserting these identities into the definition of $D_{ab}^{(i_j)}$, we find

$$\sum_{a < b} X_{ab} D_{ab}^{(i_j)} = N \left(\frac{N^{1/2}(\lambda_{i_j} - \lambda_{i_{j+1}}) - \sum_a X_{aa}(u(a)^2 - v(a)^2)}{2}, - \sum_a X_{aa} u(a)v(a) \right). \quad (8.24)$$

We have $|X_{aa}| \leq N^\varepsilon$ on $\mathcal{A}_N^{\text{reg}}(\varepsilon)$, and

$$|\lambda_{i_j} - \lambda_{i_{j+1}}| \leq 2\|H\| \leq 6. \quad (8.25)$$

Hence the norm of the right side of (8.24) is bounded by

$$N(3N^{1/2} + 2N^\varepsilon) \leq 5N^{3/2}. \quad (8.26)$$

Inserting this into (8.20) gives $|\bar{J}| \leq 20k^{1/2}N^{-1/2}$.

We now compute the second moment. Since $\mathbb{E}(Y - y)^2 = 1 + y^2$ for $y \in \mathbb{R}$, we have

$$\mathbb{E}[JJ^T \mid H] = \binom{N}{2}^{-1} \sum_{a < b} (1 + X_{ab}^2) D_{ab}^{(i)} (D_{ab}^{(i)})^T = \binom{N}{2}^{-1} \sum_{a < b} (2 + \dot{X}_{ab}) D_{ab}^{(i)} (D_{ab}^{(i)})^T. \quad (8.27)$$

To evaluate the first term, observe that $D_{ab}^{(i)}(\eta) = r_a^T \Gamma_\eta r_b$ by (8.3). Using (8.1) twice and the symmetry of $\Gamma_{\eta'}$, we obtain for $\eta, \eta' \in [2k]$ that

$$\begin{aligned} \sum_{a,b=1}^N D_{ab}^{(i)}(\eta) D_{ab}^{(i)}(\eta') &= \sum_{a=1}^N r_a^T \Gamma_\eta \left(\sum_{b=1}^N r_b r_b^T \right) \Gamma_{\eta'} r_a = N \operatorname{Tr} \left(\Gamma_\eta \Gamma_{\eta'} \sum_{a=1}^N r_a r_a^T \right) \\ &= N^2 \operatorname{Tr}(\Gamma_\eta \Gamma_{\eta'}) = 2N^2 \delta_{\eta\eta'}. \end{aligned} \quad (8.28)$$

In the last equality, we used that $\Gamma_\eta \Gamma_{\eta'} = 0$ when η and η' belong to different blocks, that Γ_η^2 is the identity on the block of η , and that $\Gamma_{2j-1} \Gamma_{2j}$ has trace zero, which can be read off from (8.2).

Since $D_{ab}^{(i)} = D_{ba}^{(i)}$ and $\binom{N}{2} = N(N-1)/2$, the identity (8.28) gives

$$2 \binom{N}{2}^{-1} \sum_{a < b} D_{ab}^{(i)} (D_{ab}^{(i)})^T = \frac{2}{N(N-1)} \left(2N^2 \operatorname{Id}_{2k} - \sum_{a=1}^N D_{aa}^{(i)} (D_{aa}^{(i)})^T \right). \quad (8.29)$$

Using $|D_{aa}^{(i)}|^2 \leq 4kN^{4\epsilon}$ from (8.10), we conclude that

$$\left\| 2 \binom{N}{2}^{-1} \sum_{a < b} D_{ab}^{(i)} (D_{ab}^{(i)})^T - 4 \operatorname{Id}_{2k} \right\| \leq \frac{4}{N-1} + \frac{8kN^{4\epsilon}}{N-1} \leq CN^{4\epsilon-1}. \quad (8.30)$$

Next, we bound the contribution of $\mathring{X}$ to (8.27). Since $\mathring{X}$ is symmetric with zero diagonal, the entry (η, η') of this contribution is

$$\binom{N}{2}^{-1} \sum_{a < b} \mathring{X}_{ab} D_{ab}^{(i)}(\eta) D_{ab}^{(i)}(\eta') = \frac{1}{N(N-1)} \sum_{a,b=1}^N \mathring{X}_{ab} D_{ab}^{(i)}(\eta) D_{ab}^{(i)}(\eta'). \quad (8.31)$$

By the definition of $D_{ab}^{(i)}$, each coordinate $D_{ab}^{(i)}(\eta)$ is N times a sum of two terms of the form $\pm v(a)v'(b)$, where v and v' are columns of U_i . Hence the sum on the right side of (8.31) is N^2 times a sum of four terms of the form $\pm (v_1 \odot v_2)^T \mathring{X} (v_3 \odot v_4)$, where $v_1, \dots, v_4$ are columns of U_i . Every two columns v, v' of U_i satisfy, on $\mathcal{A}_N^{\text{reg}}(\epsilon)$,

$$\|v \odot v'\|_2 \leq \|v\|_\infty \|v'\|_2 \leq N^{\epsilon-1/2}, \quad (8.32)$$

and therefore

$$|(v_1 \odot v_2)^T \mathring{X} (v_3 \odot v_4)| \leq \|\mathring{X}\| N^{2\epsilon-1} \leq N^{3\epsilon-1/2}. \quad (8.33)$$

It follows that the absolute value of (8.31) is bounded by

$$\frac{4N^2 N^{3\epsilon-1/2}}{N(N-1)} \leq 8N^{3\epsilon-1/2}. \quad (8.34)$$

The operator norm of a $2k \times 2k$ matrix is bounded by $2k$ times its largest entry, so (8.27) and (8.30) give

$$\|\mathbb{E}[JJ^T \mid H] - 4\operatorname{Id}_{2k}\| \leq CN^{4\epsilon-1} + 16kN^{3\epsilon-1/2} \leq CN^{3\epsilon-1/2}, \quad (8.35)$$

where we used $\varepsilon \leq 1/8$. Since $\Sigma_H = \mathbb{E}[JJ^T \mid H] - \bar{J}\bar{J}^T$ and $|\bar{J}|^2 \leq CN^{-1}$, the same bound holds for Σ_H .

Finally, the bound on $|J|$ and the elementary inequality $(y_1 + y_2)^3 \leq 4(y_1^3 + y_2^3)$ for $y_1, y_2 \geq 0$ give

$$\mathbb{E}[|J|^3 \mid H] \leq 8k^{3/2}N^{6\varepsilon}\mathbb{E}[(|Y| + N^\varepsilon)^3] \leq 32k^{3/2}N^{6\varepsilon}(\mathbb{E}|Y|^3 + N^{3\varepsilon}) \leq CN^{9\varepsilon}. \quad (8.36)$$

This completes the proof. $\square$

8.2 Eigenvectors with few coordinate values

For an integer $n \geq 1$, let $\mathcal{V}_n$ denote the set of vectors $v \in \mathbb{R}^N$ whose coordinates $v(1), \dots, v(N)$ take at most n distinct values. For $v \in \mathbb{R}^N$, we set

$$\|v\|_{2,N} = \left(\frac{1}{N} \sum_{b=1}^N v(b)^2 \right)^{1/2} = N^{-1/2} \|v\|_2. \quad (8.37)$$

Observe that $\mathcal{V}_n$ is the union of the finitely many linear subspaces of vectors that are constant on the blocks of a partition of $[N]$ into at most n blocks. Hence $\mathcal{V}_n$ is closed, and the infimum of $\|w - v\|_{2,N}$ over $v \in \mathcal{V}_n$ is attained for every $w \in \mathbb{R}^N$.

The following lemma states that, with overwhelming probability, no delocalized unit eigenvector of H is close to $\mathcal{V}_n$ when n grows slowly. Its proof follows the net and small-ball arguments developed for the invertibility of random matrices [60, 74].

Lemma 8.5. *Let $p > 0$ and $\alpha > 4p$. Let (Ω_N) and (K_N) be deterministic sequences of positive numbers such that $\Omega_N \rightarrow \infty$, $\Omega_N^p = o(N)$, $\Omega_N^{\alpha+1} = o(N^{1/2})$, and $1 \leq K_N \leq \Omega_N$ for all sufficiently large N , and write $\Omega = \Omega_N$ and $K = K_N$. Then there exists $c > 0$ such that for all sufficiently large N , the event that $\|H\| \leq 3$ and that some unit eigenvector w of H satisfies*

$$N^{1/2} \|w\|_\infty \leq K, \quad \inf_{v \in \mathcal{V}_{\lfloor \Omega^p \rfloor}} \|N^{1/2} w - v\|_{2,N} \leq \Omega^{-\alpha} \quad (8.38)$$

has probability at most $\exp(-cN \log \Omega)$.

Proof. Throughout the proof, we write $n = \lfloor \Omega^p \rfloor$, and we take N large enough that $1 \leq K \leq \Omega$ and $4\Omega^{-\alpha} \leq 1$. We first replace w and its eigenvalue by elements of a deterministic finite set. Suppose that the event in the statement holds. Let $E \in [-3, 3]$ be the eigenvalue of w , and choose $v \in \mathcal{V}_n$ such that $\|N^{1/2}w - v\|_{2,N} \leq \Omega^{-\alpha}$. We obtain $\hat{v} \in \mathbb{R}^N$ from v in two steps. We replace each coordinate $v(b)$ by its closest point in $[-K, K]$; since $|N^{1/2}w(b)| \leq K$, this does not increase $|N^{1/2}w(b) - v(b)|$. We then replace the result by the smallest among its closest points in the grid $\Omega^{-\alpha}\mathbb{Z} \cap [-K, K]$, which changes each coordinate by at most $\Omega^{-\alpha}$. Both replacements act on each coordinate

through a function of the value of that coordinate, so equal coordinates remain equal. Hence

$$\hat{v} \in \mathcal{V}_n, \quad \|\hat{v}\|_\infty \leq K, \quad \|N^{1/2}w - \hat{v}\|_{2,N} \leq 2\Omega^{-a}. \quad (8.39)$$

Set $y = N^{-1/2}\hat{v}$. Let E' be a closest point to E in $\Omega^{-a}\mathbb{Z} \cap [-3, 3]$. The previous paragraph gives

$$\|y - w\|_2 \leq 2\Omega^{-a}, \quad \frac{1}{2} \leq \|y\|_2 \leq 2, \quad \|y\|_\infty \leq KN^{-1/2}. \quad (8.40)$$

Using $(H - E)w = 0$, $\|H - E\| \leq 6$, and $|E - E'| \leq \Omega^{-a}$, we find

$$\|(H - E')y\|_2 \leq \|(H - E)(y - w)\|_2 + |E - E'| \|y\|_2 \leq 14\Omega^{-a}. \quad (8.41)$$

Let $\mathcal{Y}$ denote the set of vectors $y = N^{-1/2}\hat{v}$ such that $\hat{v} \in \mathcal{V}_n$ has all coordinates in $\Omega^{-a}\mathbb{Z} \cap [-K, K]$ and $1/2 \leq \|y\|_2 \leq 2$. The grid $\Omega^{-a}\mathbb{Z} \cap [-K, K]$ has at most $2K\Omega^a + 1 \leq 3\Omega^{a+1}$ points. We choose a set of at most n grid values, and we assign one of these values to each coordinate. Counting these choices, we obtain

$$|\mathcal{Y}| \leq (3\Omega^{a+1} + 1)^n n^N \leq (4\Omega^{a+1})^n n^N. \quad (8.42)$$

The grid $\Omega^{-a}\mathbb{Z} \cap [-3, 3]$ has at most $7\Omega^a$ points. By (8.41), the event in the statement is a subset of

$$\bigcup_{y \in \mathcal{Y}} \bigcup_{E' \in \Omega^{-a}\mathbb{Z} \cap [-3, 3]} \{ \|(H - E')y\|_2 \leq 14\Omega^{-a} \}. \quad (8.43)$$

These vectors and energies are deterministic.

We now fix $y \in \mathcal{Y}$ and $E' \in \Omega^{-a}\mathbb{Z} \cap [-3, 3]$, and we bound the probability that $\|(H - E')y\|_2 \leq 14\Omega^{-a}$. Let $\mathcal{T} \subset [N]$ be a set of $\lfloor N/2 \rfloor$ indices b for which $|y(b)|$ is largest, with complement $\mathcal{T}^c = [N] \setminus \mathcal{T}$. Let $y_{\mathcal{T}} \in \mathbb{R}^N$ denote the vector with $y_{\mathcal{T}}(b) = y(b)$ for $b \in \mathcal{T}$ and $y_{\mathcal{T}}(b) = 0$ otherwise. The average of $y(b)^2$ over $b \in \mathcal{T}$ is at least its average over $[N]$, so

$$\|y_{\mathcal{T}}\|_2^2 \geq \frac{\lfloor N/2 \rfloor}{N} \|y\|_2^2 \geq \frac{1}{12}. \quad (8.44)$$

Suppose that $\|(H - E')y\|_2 \leq 14\Omega^{-a}$. Then fewer than $N/4$ indices $a \in [N]$ satisfy $|((H - E')y)(a)| > 28\Omega^{-a}N^{-1/2}$. Since $|\mathcal{T}^c| \geq N/2$, at least $N/4$ indices $a \in \mathcal{T}^c$ satisfy

$$|N^{1/2}((H - E')y)(a)| \leq 28\Omega^{-a}. \quad (8.45)$$

The rows indexed by $\mathcal{T}^c$ are conditionally independent once we fix the entries outside $\mathcal{T}^c \times \mathcal{T}$. Indeed, for $a \in \mathcal{T}^c$ we have

$$N^{1/2}((H - E')y)(a) = \sum_{b \in \mathcal{T}} X_{ab}y(b) + \chi_a, \quad \chi_a = \sum_{b \in \mathcal{T}^c} X_{ab}y(b) - N^{1/2}E'y(a). \quad (8.46)$$

Let $\mathcal{F}_{\mathcal{T}}$ denote the σ -algebra generated by the entries X_{ab} , $a \leq b$, for which a and b do not lie on opposite sides of the partition $\{\mathcal{T}, \mathcal{T}^c\}$. Each χ_a is $\mathcal{F}_{\mathcal{T}}$ -measurable. The entries X_{ab} with $a \in \mathcal{T}^c$ and $b \in \mathcal{T}$ are off-diagonal, and distinct pairs (a, b) of this kind give distinct entries. Hence these entries are independent copies of X , independent of $\mathcal{F}_{\mathcal{T}}$. In particular, conditionally on $\mathcal{F}_{\mathcal{T}}$, the random variables $\sum_{b \in \mathcal{T}} X_{ab} y(b) + \chi_a$, $a \in \mathcal{T}^c$, are independent, because they are functions of disjoint families of these entries.

We bound the conditional probability of (8.45) for one row with the Berry–Esseen inequality for sums of independent, not necessarily identically distributed, random variables [58, Chapter V]. Since $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$, it states that the Kolmogorov distance between the law of $\|y_{\mathcal{T}}\|_2^{-1} \sum_{b \in \mathcal{T}} X_{ab} y(b)$ and the standard Gaussian law is bounded by $C_{\text{BE}} \|y_{\mathcal{T}}\|_2^{-3} \sum_{b \in \mathcal{T}} \mathbb{E}|X|^3 |y(b)|^3$, where C_{BE} is an absolute constant. The standard Gaussian probability of a closed interval is bounded by its length times $(2\pi)^{-1/2}$, and the probabilities of an interval under two laws differ by at most twice their Kolmogorov distance. Using (8.40) and (8.44), we therefore find for every $\chi \in \mathbb{R}$ that

$$\begin{aligned} \mathbb{P}\left(\left|\sum_{b \in \mathcal{T}} X_{ab} y(b) + \chi\right| \leq 28 \Omega^{-\alpha}\right) &\leq \frac{56 \Omega^{-\alpha}}{(2\pi)^{1/2} \|y_{\mathcal{T}}\|_2} + \frac{2C_{\text{BE}} \mathbb{E}|X|^3 \|y\|_{\infty} \|y_{\mathcal{T}}\|_2^2}{\|y_{\mathcal{T}}\|_2^3} \\ &\leq C(\Omega^{-\alpha} + KN^{-1/2}). \end{aligned} \quad (8.47)$$

Recall that $K \leq \Omega$ and $\Omega^{\alpha+1} N^{-1/2} \rightarrow 0$. Hence the right side is bounded by $C_2 \Omega^{-\alpha}$ for some $C_2 > 0$ and all large N , and we take N large enough that $C_2 \Omega^{-\alpha} \leq 1$.

We apply this bound with $\chi = \chi_a$, use the conditional independence of the rows, and take a union bound over the at most 2^N sets of $\lceil N/4 \rceil$ rows in $\mathcal{T}^c$. This gives

$$\mathbb{P}(\|(H - E')y\|_2 \leq 14 \Omega^{-\alpha} \mid \mathcal{F}_{\mathcal{T}}) \leq 2^N (C_2 \Omega^{-\alpha})^{N/4}. \quad (8.48)$$

The right side is deterministic, so the same bound holds for the unconditional probability.

Combining the last display with (8.42) and the bound on the number of energies, we find that the logarithm of the probability of the event in the statement is bounded by

$$n \log(4\Omega^{\alpha+1}) + N \log n + \log(7\Omega^{\alpha}) + N \log 2 + \frac{N}{4} \log C_2 - \frac{\alpha}{4} N \log \Omega. \quad (8.49)$$

We have $N \log n \leq pN \log \Omega$. Since $n \leq \Omega^p = o(N)$ and $\log \Omega \rightarrow \infty$, the first, third, fourth, and fifth terms are $o(N \log \Omega)$. Hence (8.49) is bounded by $-(\alpha/4 - p - o(1))N \log \Omega$, and the lemma holds with $c = (\alpha - 4p)/8$. $\square$

Remark 8.6. The vectors $y \in \mathcal{Y}$ and energies E' in the union bound in the proof of Lemma 8.5 are deterministic, and the eigenvector w enters only through the existence of a nearby pair (y, E') . This is why the lemma applies to the eigenvectors of H , which are functions of the entries of H . An argument that conditions on an eigenvector and treats the entries of H as independent would not be valid.

8.3 A deterministic nonlattice estimate

In Section 8.3, U denotes a deterministic $N \times 2k$ matrix with $U^T U = \text{Id}_{2k}$. We use the notation of Section 8.1 for such general U , so that r_b denotes the b -th row of $N^{1/2}U$, and (8.1) holds. For $\xi \in \mathbb{R}^{2k}$, we define

$$\mathcal{D}_U(\xi) = \frac{1}{N^2} \sum_{a,b=1}^N [1 - \cos(r_a^T \Gamma(\xi) r_b)]. \quad (8.50)$$

For $U = U_i$, the identity (8.3) shows that $\mathcal{D}_U(\xi)$ is the average of $1 - \cos(\xi \cdot D_{ab}^{(i)})$ over all $(a, b) \in [N]^2$. It is small exactly when most of the phases $\xi \cdot D_{ab}^{(i)}$ are close to $2\pi\mathbb{Z}$. We also define the exponents

$$\mathfrak{p}_k = 4k + 4, \quad \mathfrak{a}_k = 100(k + 1), \quad \mathfrak{q}_k = 2\mathfrak{a}_k + 2k + 10, \quad (8.51)$$

and we note that $\mathfrak{a}_k > 4\mathfrak{p}_k$.

Lemma 8.7. *There exists $\Omega_0(k) \geq 3$ such that the following holds. Let $K \geq 1$ and $\Omega \geq \max\{\Omega_0(k), K^2\}$. Suppose that $\max_b |r_b| \leq K$ and that every column u of U satisfies*

$$\inf_{v \in \mathcal{V}_{[\Omega^{\mathfrak{p}_k}]}} \|N^{1/2}u - v\|_{2,N} > \Omega^{-\mathfrak{a}_k}. \quad (8.52)$$

Then $\mathcal{D}_U(\xi) \geq \Omega^{-\mathfrak{q}_k}$ for all $\xi \in \mathbb{R}^{2k}$ such that $\Omega^{-1} \leq |\xi| \leq \Omega K^{-2}$.

Proof. We argue by contradiction. Suppose that there exists ξ with $\Omega^{-1} \leq |\xi| \leq \Omega K^{-2}$ and $\mathcal{D}_U(\xi) < \Omega^{-\mathfrak{q}_k}$, and abbreviate $\mathcal{D} = \mathcal{D}_U(\xi)$. For $a, b \in [N]$, let

$$\Theta_{ab} = \text{dist}(r_a^T \Gamma(\xi) r_b, 2\pi\mathbb{Z}) \in [0, \pi]. \quad (8.53)$$

For real t' with $\text{dist}(t', 2\pi\mathbb{Z}) = \Theta$, we have

$$1 - \cos t' = 2 \sin^2(\Theta/2) \geq \frac{2\Theta^2}{\pi^2}, \quad (8.54)$$

where we used $\sin y' \geq 2y'/\pi$ for $y' \in [0, \pi/2]$. Summing over a and b , we obtain

$$\frac{1}{N^2} \sum_{a,b=1}^N \Theta_{ab}^2 \leq \frac{\pi^2}{2} \mathcal{D}. \quad (8.55)$$

We first discard the rows with large errors. For $a \in [N]$, set

$$\bar{\Theta}_a = \frac{1}{N} \sum_{b=1}^N \Theta_{ab}^2, \quad \mathcal{O} = \{a \in [N] : \bar{\Theta}_a \leq 2\pi^2 K^2 \mathcal{D}\}. \quad (8.56)$$

When $\mathcal{D} > 0$, Markov's inequality and (8.55) give

$$|[N] \setminus \mathcal{O}| \leq \frac{N}{4K^2}, \quad (8.57)$$

and when $\mathcal{D} = 0$, we have $\mathcal{O} = [N]$. Using (8.1) and $|r_a| \leq K$, we find, in the sense of quadratic forms,

$$\frac{1}{N} \sum_{a \in \mathcal{O}} r_a r_a^T = \text{Id}_{2k} - \frac{1}{N} \sum_{a \notin \mathcal{O}} r_a r_a^T \geq \left(1 - \frac{K^2 |[N] \setminus \mathcal{O}|}{N}\right) \text{Id}_{2k} \geq \frac{3}{4} \text{Id}_{2k}. \quad (8.58)$$

Next, we select $2k$ rows that form a well-conditioned matrix. For a set $\mathcal{O}_0 \subset \mathcal{O}$ with $2k$ elements, let $Q(\mathcal{O}_0)$ denote the $2k \times 2k$ matrix with rows r_a^T , $a \in \mathcal{O}_0$, in increasing order of a . The Cauchy–Binet formula, applied to the matrix with rows r_a^T , $a \in \mathcal{O}$, gives

$$\det\left(\sum_{a \in \mathcal{O}} r_a r_a^T\right) = \sum_{\mathcal{O}_0 \subset \mathcal{O}, |\mathcal{O}_0|=2k} \det(Q(\mathcal{O}_0))^2. \quad (8.59)$$

By (8.58), the left side is at least $(3N/4)^{2k}$; in particular, $|\mathcal{O}| \geq 2k$. The sum has at most $\binom{N}{2k} \leq N^{2k}/(2k)!$ terms, so there exists $\mathcal{O}_0$ such that $Q = Q(\mathcal{O}_0)$ satisfies

$$|\det Q| \geq (3/4)^k ((2k)!)^{1/2}. \quad (8.60)$$

We now bound $\|Q^{-1}\|$. The rows of Q have norm at most K , so $\|Q\| \leq (2k)^{1/2} K$. The determinant is the product of the $2k$ singular values of Q , so the smallest singular value of Q is at least $|\det Q| \|Q\|^{1-2k}$. Therefore

$$\|Q^{-1}\| \leq \frac{\|Q\|^{2k-1}}{|\det Q|} \leq C_k K^{2k-1}. \quad (8.61)$$

Next, we round the phases against the rows of Q . For $a \in \mathcal{O}_0$ and $b \in [N]$, choose $\mathbf{n}_b(a) \in \mathbb{Z}$ such that

$$|r_a^T \Gamma(\xi) r_b - 2\pi \mathbf{n}_b(a)| = \Theta_{ab}. \quad (8.62)$$

We let $\mathbf{n}_b \in \mathbb{Z}^{2k}$ denote the vector with coordinates $\mathbf{n}_b(a)$, $a \in \mathcal{O}_0$, in increasing order of a . The coordinates of $Q\Gamma(\xi)r_b$ are the phases $r_a^T \Gamma(\xi)r_b$, $a \in \mathcal{O}_0$. Since $\mathcal{O}_0 \subset \mathcal{O}$, we find

$$\frac{1}{N} \sum_{b=1}^N |Q\Gamma(\xi)r_b - 2\pi \mathbf{n}_b|^2 = \sum_{a \in \mathcal{O}_0} \bar{\Theta}_a \leq 4k\pi^2 K^2 \mathcal{D}. \quad (8.63)$$

Further, the bound $|\xi| \leq \Omega K^{-2}$ gives

$$|r_a^T \Gamma(\xi) r_b| \leq \|\Gamma(\xi)\| K^2 \leq |\xi| K^2 \leq \Omega, \quad \text{so} \quad |\mathbf{n}_b(a)| \leq \frac{\Omega}{2\pi} + \frac{1}{2} \leq \Omega. \quad (8.64)$$

Hence, as b ranges over $[N]$, the vector $\mathbf{n}_b$ takes at most $(2\Omega + 1)^{2k} \leq (3\Omega)^{2k}$ values.

We now reconstruct one block of the rows r_b from the integer vectors $\mathbf{n}_b$. Choose $j \in [k]$ such that $|\xi_j| = \max_{j'} |\xi_{j'}|$, so that

$$|\xi_j| \geq k^{-1/2} |\xi| \geq k^{-1/2} \Omega^{-1}. \quad (8.65)$$

Since $\Gamma(\xi)$ is block diagonal, the j -th block of $\Gamma(\xi)r_b$ is $\Gamma(\xi_j)r_{b,j}$. For $b \in [N]$, define

$$\hat{r}_b = \Gamma(\xi_j)^{-1} (Q^{-1} 2\pi \mathbf{n}_b)_j \in \mathbb{R}^2. \quad (8.66)$$

Then

$$r_{b,j} - \hat{r}_b = \Gamma(\xi_j)^{-1} \left(Q^{-1} (Q\Gamma(\xi)r_b - 2\pi \mathbf{n}_b) \right)_j. \quad (8.67)$$

Using $\|\Gamma(\xi_j)^{-1}\| = |\xi_j|^{-1}$, (8.61), and (8.63), we obtain

$$\begin{aligned} \left(\frac{1}{N} \sum_{b=1}^N |r_{b,j} - \hat{r}_b|^2 \right)^{1/2} &\leq \frac{\|Q^{-1}\|}{|\xi_j|} (4k\pi^2 K^2 \mathcal{D})^{1/2} \leq C_k \Omega K^{2k} \mathcal{D}^{1/2} \\ &\leq C_k \Omega^{k+1-q_k/2} = C_k \Omega^{-a_k-4}. \end{aligned} \quad (8.68)$$

In the third inequality, we used $K^2 \leq \Omega$ and $\mathcal{D} < \Omega^{-q_k}$, and the last equality follows from the definition of q_k in (8.51).

It remains to derive the contradiction. Let u be the $(2j-1)$ -th column of U , so that $N^{1/2}u(b)$ is the first coordinate of $r_{b,j}$. Let $v \in \mathbb{R}^N$ be the vector whose b -th coordinate is the first coordinate of $\hat{r}_b$. The vector $\hat{r}_b$ depends on b only through $\mathbf{n}_b$, so v has at most $(3\Omega)^{2k}$ distinct coordinates. For $\Omega \geq 3$, we have $(3\Omega)^{2k} \leq \lfloor \Omega^{p_k} \rfloor$, and hence $v \in \mathcal{V}_{\lfloor \Omega^{p_k} \rfloor}$. By (8.68), we have

$$\|N^{1/2}u - v\|_{2,N} \leq C_k \Omega^{-a_k-4} \leq \Omega^{-a_k}, \quad (8.69)$$

where the second inequality holds when $\Omega^4 \geq C_k$. This contradicts (8.52), and the lemma follows with $\Omega_0(k) = \max\{3, C_k^{1/4}\}$, where C_k is the constant in the last line of (8.68). $\square$

Remark 8.8. The resampling selects only pairs $a < b$, so the characteristic function Φ_H involves the phases $r_a^T \Gamma(\xi)r_b$ only for $a < b$. We therefore need the following version of Lemma 8.7 without diagonal terms.

For $\xi \in \mathbb{R}^{2k}$, we define

$$\mathcal{D}_U^{\text{off}}(\xi) = \binom{N}{2}^{-1} \sum_{a < b} [1 - \cos(r_a^T \Gamma(\xi)r_b)]. \quad (8.70)$$

Corollary 8.9. *Let K , Ω , and U satisfy the hypotheses of Lemma 8.7, and suppose in addition that $\Omega^{q_k} \leq N/4$. Then $\mathcal{D}_U^{\text{off}}(\xi) \geq \Omega^{-q_k}/2$ for all $\xi \in \mathbb{R}^{2k}$ such that $\Omega^{-1} \leq |\xi| \leq \Omega K^{-2}$.*

Proof. The summand in (8.50) is symmetric in (a, b) and takes values in $[0, 2]$, so

$$\mathcal{D}_U(\xi) = \left(1 - \frac{1}{N}\right) \mathcal{D}_U^{\text{off}}(\xi) + \frac{1}{N^2} \sum_{a=1}^N [1 - \cos(r_a^T \Gamma(\xi) r_a)] \leq \mathcal{D}_U^{\text{off}}(\xi) + \frac{2}{N}. \quad (8.71)$$

For $\Omega^{-1} \leq |\xi| \leq \Omega K^{-2}$, Lemma 8.7 and the condition $\Omega^{q_k} \leq N/4$ therefore give $\mathcal{D}_U^{\text{off}}(\xi) \geq \Omega^{-q_k} - 2N^{-1} \geq \Omega^{-q_k}/2$. $\square$

8.4 The characteristic function of one increment

We now convert Lemma 8.7 into a bound on $|\Phi_H|$. Let Y' be a copy of Y independent of Y . Since $\mathbb{E}(Y - Y')^2 = 2$, we have $\mathbb{P}(Y \neq Y') > 0$, and we may choose constants $0 < c_- < c_+$, depending only on the law of X , such that

$$p_Y = \mathbb{P}(c_- \leq |Y - Y'| \leq c_+) > 0. \quad (8.72)$$

Lemma 8.10. *Let $K \geq 1$ and $\Omega \geq \max\{\Omega_0(k), K^2\}$ with $\Omega^{q_k} \leq N/4$. Let $\mathbf{i}$ be a k -tuple, and consider a realization of H for which $U = U_{\mathbf{i}}$ satisfies the hypotheses of Lemma 8.7. Then*

$$1 - |\Phi_H(\xi)|^2 \geq \frac{p_Y}{2} \Omega^{-q_k} \quad \text{for all } \xi \in \mathbb{R}^{2k} \text{ such that } \frac{1}{c_- \Omega} \leq |\xi| \leq \frac{\Omega}{c_+ K^2}. \quad (8.73)$$

Proof. Conditioning on (a, b) in (8.7) and using the independence of Y , we find

$$\Phi_H(\xi) = \binom{N}{2}^{-1} \sum_{a < b} e^{-iX_{ab}\xi \cdot D_{ab}^{(\mathbf{i})}} \Phi_Y(\xi \cdot D_{ab}^{(\mathbf{i})}). \quad (8.74)$$

We symmetrize with the Cauchy–Schwarz inequality, which removes the phases $e^{-iX_{ab}\xi \cdot D_{ab}^{(\mathbf{i})}}$. Using

$$|\Phi_Y(t')|^2 = \mathbb{E} e^{it'(Y - Y')} = \mathbb{E} \cos(t'(Y - Y')), \quad t' \in \mathbb{R}, \quad (8.75)$$

we obtain

$$|\Phi_H(\xi)|^2 \leq \binom{N}{2}^{-1} \sum_{a < b} |\Phi_Y(\xi \cdot D_{ab}^{(\mathbf{i})})|^2 = \mathbb{E} \left[\binom{N}{2}^{-1} \sum_{a < b} \cos((Y - Y') \xi \cdot D_{ab}^{(\mathbf{i})}) \right]. \quad (8.76)$$

By (8.3) and the linearity of Γ , we have

$$(Y - Y') \xi \cdot D_{ab}^{(\mathbf{i})} = r_a^T \Gamma((Y - Y') \xi) r_b. \quad (8.77)$$

Hence (8.76) and (8.70) give

$$1 - |\Phi_H(\xi)|^2 \geq \mathbb{E} [\mathcal{D}_U^{\text{off}}((Y - Y') \xi)] \geq \mathbb{E} [\mathcal{D}_U^{\text{off}}((Y - Y') \xi) \mathbf{1}\{c_- \leq |Y - Y'| \leq c_+\}], \quad (8.78)$$

where the second inequality holds because $\mathcal{D}_U^{\text{off}} \geq 0$. For ξ as in (8.73) and every realization with $c_- \leq |Y - Y'| \leq c_+$, we have

$$\Omega^{-1} \leq c_- |\xi| \leq |(Y - Y')\xi| \leq c_+ |\xi| \leq \Omega K^{-2}. \quad (8.79)$$

Therefore Corollary 8.9 gives

$$\mathcal{D}_U^{\text{off}}((Y - Y')\xi) \geq \Omega^{-q_k}/2 \quad (8.80)$$

on this event. Taking expectations and using (8.72) completes the proof. $\square$

8.5 A local limit theorem for sums of increments

We now fix the parameters. Recall that $\tau = N^{1/100}$, and set

$$\varepsilon_k = \frac{1}{4000 q_k}, \quad \Omega = \tau^{1/(4q_k)} = N^{10\varepsilon_k}, \quad K = (2k)^{1/2} N^{\varepsilon_k}. \quad (8.81)$$

Let $C_{\text{mom}} \geq 1$ denote the constant C of Lemma 8.4 for $\varepsilon = \varepsilon_k$, and set

$$\varrho_N = (12C_{\text{mom}})^{-1} N^{-9\varepsilon_k}. \quad (8.82)$$

In the proof of Lemma 8.13, ϱ_N will be the radius of the ball of frequencies on which a second-order Taylor expansion of Φ_H is effective. The exponent of Ω is chosen so that $\Omega^{-q_k} = \tau^{-1/4}$, which makes $m\Omega^{-q_k}$ a positive power of τ . Since Ω , K^2 , and ϱ_N^{-1} are constant multiples of $N^{10\varepsilon_k}$, $N^{2\varepsilon_k}$, and $N^{9\varepsilon_k}$, respectively, there exists N_0 such that for $N \geq N_0$,

$$\Omega \geq \max\{\Omega_0(k), K^2\}, \quad \Omega^{q_k} \leq \frac{N}{4}, \quad c_+ K^2 \leq \Omega^{1/2}, \quad \frac{1}{c_- \Omega} \leq \varrho_N. \quad (8.83)$$

The last condition dictates the choice of ε_k . Since $\Omega = N^{1/(400q_k)}$ is determined by τ , this condition can hold for large N only if $9\varepsilon_k \leq 1/(400q_k)$, and our choice gives $9\varepsilon_k < 10\varepsilon_k = 1/(400q_k)$. Further, we have $10\varepsilon_k p_k < 1$ and $10\varepsilon_k(a_k + 1) < 1/2$, so

$$\Omega^{p_k} = o(N), \quad \Omega^{a_k+1} = o(N^{1/2}). \quad (8.84)$$

Definition 8.11. We let $\mathcal{A}_{N,k}$ denote the event that $\mathcal{A}_N^{\text{reg}}(\varepsilon_k)$ holds and that every unit eigenvector w of H satisfies

$$\inf_{v \in \mathcal{V}_{[\Omega^{p_k}]}} \|N^{1/2}w - v\|_{2,N} > \Omega^{-a_k}. \quad (8.85)$$

The event $\mathcal{A}_{N,k}$ does not refer to any tuple $\mathbf{i}$. This is what makes the estimates below uniform over tuples, including tuples selected according to properties of H such as small gaps.

Lemma 8.12. *The event $\mathcal{A}_{N,k}$ holds with overwhelming probability. Moreover, for $N \geq N_0$, on $\mathcal{A}_{N,k}$ and for every k -tuple $\mathbf{i}$, we have*

$$1 - |\Phi_H(\xi)|^2 \geq \frac{p_Y}{2} \tau^{-1/4} \quad \text{for all } \xi \in \mathbb{R}^{2k} \text{ such that } q_N \leq |\xi| \leq \Omega^{1/2}. \quad (8.86)$$

Proof. Recall that $\|H\| \leq 3$ and that every unit eigenvector w of H satisfies $N^{1/2}\|w\|_\infty \leq N^{\varepsilon_k} \leq K$ on $\mathcal{A}_N^{\text{reg}}(\varepsilon_k)$. Hence the complement of $\mathcal{A}_{N,k}$ is a subset of the union of the complement of $\mathcal{A}_N^{\text{reg}}(\varepsilon_k)$ and the event of Lemma 8.5 with $\mathfrak{p} = \mathfrak{p}_k$, $\mathfrak{a} = \mathfrak{a}_k$, and Ω and K as in (8.81). The hypotheses of Lemma 8.5 hold by (8.84), and $K \leq \Omega$ for large N . Lemma 8.3 and Lemma 8.5 therefore show that $\mathcal{A}_{N,k}$ holds with overwhelming probability.

Now let $N \geq N_0$, fix a realization of H in $\mathcal{A}_{N,k}$, and let $\mathbf{i}$ be a k -tuple. The rows of $N^{1/2}U_{\mathbf{i}}$ satisfy $|r_b| \leq K$ by (8.10). The columns of $U_{\mathbf{i}}$ are unit eigenvectors of H , so the definition of $\mathcal{A}_{N,k}$ gives (8.52). Together with the first two conditions in (8.83), this verifies the hypotheses of Lemma 8.10. The last two conditions in (8.83) show that every ξ with $q_N \leq |\xi| \leq \Omega^{1/2}$ satisfies the condition in (8.73). Since $\Omega^{-q_k} = \tau^{-1/4}$, the bound (8.73) gives (8.86). $\square$

For a positive definite $2k \times 2k$ matrix Σ , let $\mathfrak{g}_\Sigma$ denote the density of the centered Gaussian law on $\mathbb{R}^{2k}$ with covariance Σ ,

$$\mathfrak{g}_\Sigma(y) = (2\pi)^{-k} (\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2} y^T \Sigma^{-1} y\right), \quad y \in \mathbb{R}^{2k}. \quad (8.87)$$

This extends the notation (3.34) on $\mathbb{R}^{2k}$, so that $\mathfrak{g}_v = \mathfrak{g}_{v\text{Id}_{2k}}$ for $v > 0$. In particular, $\mathfrak{g}_{4n\text{Id}_{2k}}(y) = (8\pi n)^{-k} e^{-|y|^2/(8n)}$. We now state the local limit theorem. Its proof is the classical Fourier-analytic argument for local limit theorems (see [7]), made uniform in H , in the tuple $\mathbf{i}$, and in the number of summands. The input specific to our setting is the bound (8.86).

Lemma 8.13. *Let $\Psi \subset C_c^\infty(\mathbb{R}^{2k})$ be a family of functions such that there exists $R_\Psi > 0$ with $\text{supp } \psi \subset \{y \in \mathbb{R}^{2k} : |y| \leq R_\Psi\}$ for all $\psi \in \Psi$, and such that each partial derivative of each order is bounded uniformly over $\psi \in \Psi$. Then on $\mathcal{A}_{N,k}$, for every k -tuple $\mathbf{i}$, every integer n with $\tau/10 \leq n \leq \tau/3$, every $x \in \mathbb{R}^{2k}$, and every $\psi \in \Psi$,*

$$\mathbb{E}[\psi(S^{(n)} + x) \mid H] = \mathfrak{g}_{4n\text{Id}_{2k}}(x) \int_{\mathbb{R}^{2k}} \psi(y) \, dy + o(\tau^{-k}), \quad (8.88)$$

where the error term is bounded in absolute value by a deterministic quantity that is $o(\tau^{-k})$ and depends on Ψ , but not on $\mathbf{i}$, n , x , ψ , or the realization of H in $\mathcal{A}_{N,k}$. Further, for every bounded Borel set $B \subset \mathbb{R}^{2k}$, there exists $C_B > 0$ such that on $\mathcal{A}_{N,k}$, for every k -tuple $\mathbf{i}$ and every integer n with $\tau/10 \leq n \leq \tau/3$,

$$\sup_{x \in \mathbb{R}^{2k}} \mathbb{P}(S^{(n)} + x \in B \mid H) \leq C_B \tau^{-k}. \quad (8.89)$$

Proof. In this proof, constants may depend on Ψ in addition to k and the laws of X and Z . We first record a trivial bound that holds for all N . By the hypothesis on the partial derivatives of order zero, the functions in Ψ are bounded uniformly. They are also supported in the ball of radius R_Ψ centered at the origin, and therefore

$$\sup_{\psi \in \Psi} \|\psi\|_\infty < \infty, \quad \sup_{\psi \in \Psi} \int_{\mathbb{R}^{2k}} |\psi(y)| \, dy \leq (2R_\Psi)^{2k} \sup_{\psi \in \Psi} \|\psi\|_\infty < \infty. \quad (8.90)$$

Every integer n with $\tau/10 \leq n \leq \tau/3$ is positive, so $\mathfrak{g}_{4n\text{Id}_{2k}}(x) \leq (8\pi n)^{-k} \leq (8\pi)^{-k}$ for all $x \in \mathbb{R}^{2k}$. Hence, for every realization of H and all $\mathbf{i}$, n , x , and ψ as in the statement, the error term in (8.88) is bounded in absolute value by $\sup_{\psi \in \Psi} \|\psi\|_\infty + (8\pi)^{-k} \sup_{\psi \in \Psi} \int |\psi|$, which is finite by (8.90). We take this number as the deterministic quantity in the statement when $N < N_0$, $\tau < 20$, or $C_{\text{mom}} N^{3\varepsilon_k - 1/2} > 1$. Each of these conditions holds only for finitely many N , and the deterministic quantity only needs to be $o(\tau^{-k})$ as $N \rightarrow \infty$. We may therefore assume from now on that $N \geq N_0$, that $\tau \geq 20$, and that $C_{\text{mom}} N^{3\varepsilon_k - 1/2} \leq 1$. Fix a realization of H in $\mathcal{A}_{N,k}$, a k -tuple $\mathbf{i}$, an integer n with $\tau/10 \leq n \leq \tau/3$, a vector $x \in \mathbb{R}^{2k}$, and $\psi \in \Psi$. Then $n \geq \tau/10 \geq 2$.

We begin with bounds on the Fourier transform

$$\hat{\psi}(\xi) = \int_{\mathbb{R}^{2k}} e^{-i\xi \cdot y} \psi(y) \, dy, \quad \xi \in \mathbb{R}^{2k}. \quad (8.91)$$

Set

$$\mathfrak{s}_k = 4k\varepsilon_k + k + 1. \quad (8.92)$$

Integrating by parts $2\mathfrak{s}_k$ times, differentiating under the integral, and using that the functions in Ψ are supported in a common ball and have uniformly bounded derivatives, we find

$$|\hat{\psi}(\xi)| \leq \frac{C}{1 + |\xi|^{2\mathfrak{s}_k}}, \quad |\hat{\psi}(\xi) - \hat{\psi}(0)| \leq C|\xi|, \quad \xi \in \mathbb{R}^{2k}. \quad (8.93)$$

Since $2\mathfrak{s}_k > 2k$, the first bound implies

$$\int_{\mathbb{R}^{2k}} |\hat{\psi}(\xi)| \, d\xi \leq C. \quad (8.94)$$

Next, we write both sides of (8.88) as Fourier integrals. Since $\hat{\psi}$ is integrable, the inversion formula

$$\psi(y) = (2\pi)^{-2k} \int_{\mathbb{R}^{2k}} e^{i\xi \cdot y} \hat{\psi}(\xi) \, d\xi \quad (8.95)$$

holds for all y . Inserting it into the conditional expectation, and using Fubini's theorem and (8.8), we obtain

$$\mathbb{E}[\psi(S^{(n)} + x) \mid H] = (2\pi)^{-2k} \int_{\mathbb{R}^{2k}} e^{i\xi \cdot x} \hat{\psi}(\xi) \Phi_H(\xi)^n \, d\xi. \quad (8.96)$$

Further, we have

$$\hat{\psi}(0) = \int \psi, \quad \int_{\mathbb{R}^{2k}} e^{i\tilde{\zeta} \cdot x} e^{-2n|\tilde{\zeta}|^2} d\tilde{\zeta} = (2\pi)^{2k} \mathfrak{g}_{4n\text{Id}_{2k}}(x). \quad (8.97)$$

Subtracting, we find

$$\left| \mathbb{E}[\psi(S^{(n)} + x) \mid H] - \mathfrak{g}_{4n\text{Id}_{2k}}(x) \int \psi \right| \leq (2\pi)^{-2k} \int_{\mathbb{R}^{2k}} |\hat{\psi}(\tilde{\zeta}) \Phi_H(\tilde{\zeta})^n - \hat{\psi}(0) e^{-2n|\tilde{\zeta}|^2}| d\tilde{\zeta}. \quad (8.98)$$

The right side of (8.98) does not involve x , which gives the uniformity in x .

We split the integral in (8.98) into three ranges of frequencies. Since $n \geq \tau/10$, the right side of (8.98) is bounded by $C\tau^{-k}(\mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4)$, where

$$\begin{aligned} \mathcal{I}_1 &= n^k \int_{|\tilde{\zeta}| \leq \varrho_N} |\hat{\psi}(\tilde{\zeta}) \Phi_H(\tilde{\zeta})^n - \hat{\psi}(0) e^{-2n|\tilde{\zeta}|^2}| d\tilde{\zeta}, & \mathcal{I}_2 &= n^k \int_{\varrho_N < |\tilde{\zeta}| \leq \Omega^{1/2}} |\hat{\psi}(\tilde{\zeta})| |\Phi_H(\tilde{\zeta})|^n d\tilde{\zeta}, \\ \mathcal{I}_3 &= n^k \int_{|\tilde{\zeta}| > \Omega^{1/2}} |\hat{\psi}(\tilde{\zeta})| d\tilde{\zeta}, & \mathcal{I}_4 &= n^k |\hat{\psi}(0)| \int_{|\tilde{\zeta}| > \varrho_N} e^{-2n|\tilde{\zeta}|^2} d\tilde{\zeta}, \end{aligned} \quad (8.99)$$

and where we used $|\Phi_H| \leq 1$ to define $\mathcal{I}_3$. All constants above depend on ψ only through R_Ψ and the uniform bounds on finitely many derivatives. It therefore suffices to show that $\mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4$ is bounded by a deterministic quantity that tends to zero.

We first treat small frequencies. Let

$$\Phi_H^\circ(\tilde{\zeta}) = e^{-i\tilde{\zeta} \cdot \bar{J}} \Phi_H(\tilde{\zeta}) = \mathbb{E}[e^{i\tilde{\zeta} \cdot (J - \bar{J})} \mid H], \quad \tilde{\zeta} \in \mathbb{R}^{2k}. \quad (8.100)$$

By Lemma 8.4 with $\varepsilon = \varepsilon_k$ and our assumption $C_{\text{mom}} N^{3\varepsilon_k - 1/2} \leq 1$, we have

$$3\text{Id}_{2k} \leq \Sigma_H \leq 5\text{Id}_{2k} \quad (8.101)$$

and

$$\mathbb{E}[|J - \bar{J}|^3 \mid H] \leq 4(\mathbb{E}[|J|^3 \mid H] + |\bar{J}|^3) \leq 8\mathbb{E}[|J|^3 \mid H] \leq 8C_{\text{mom}} N^{9\varepsilon_k}. \quad (8.102)$$

We apply the inequality $|e^{it'} - 1 - it' + t'^2/2| \leq |t'|^3/6$, valid for real t' , with $t' = \tilde{\zeta} \cdot (J - \bar{J})$, and obtain

$$\left| \Phi_H^\circ(\tilde{\zeta}) - 1 + \frac{1}{2} \tilde{\zeta}^T \Sigma_H \tilde{\zeta} \right| \leq \frac{4}{3} C_{\text{mom}} N^{9\varepsilon_k} |\tilde{\zeta}|^3, \quad \tilde{\zeta} \in \mathbb{R}^{2k}. \quad (8.103)$$

For $|\tilde{\zeta}| \leq \varrho_N$, the definition (8.82) gives

$$\frac{3}{2} |\tilde{\zeta}|^2 \leq \frac{1}{2} \tilde{\zeta}^T \Sigma_H \tilde{\zeta} \leq \frac{5}{2} |\tilde{\zeta}|^2 \leq 1, \quad \frac{4}{3} C_{\text{mom}} N^{9\varepsilon_k} |\tilde{\zeta}|^3 \leq \frac{1}{9} |\tilde{\zeta}|^2. \quad (8.104)$$

Hence (8.103) implies

$$|\Phi_H(\tilde{\zeta})| = |\Phi_H^\circ(\tilde{\zeta})| \leq 1 - \frac{3}{2} |\tilde{\zeta}|^2 + \frac{1}{9} |\tilde{\zeta}|^2 \leq e^{-|\tilde{\zeta}|^2}, \quad |\tilde{\zeta}| \leq \varrho_N. \quad (8.105)$$

We now change variables $\xi = n^{-1/2}\xi'$ in $\mathcal{I}_1$, which gives

$$\mathcal{I}_1 = \int_{|\xi'| \leq \varrho_N n^{1/2}} |\hat{\psi}(n^{-1/2}\xi') \Phi_H(n^{-1/2}\xi')^n - \hat{\psi}(0) e^{-2|\xi'|^2}| d\xi'. \quad (8.106)$$

By (8.93) and (8.105), the integrand is bounded by

$$C n^{-1/2} |\xi'| e^{-|\xi'|^2} + C |\Phi_H(n^{-1/2}\xi')^n - e^{-2|\xi'|^2}|. \quad (8.107)$$

We have $\Phi_H(\xi)^n = e^{in\xi \cdot \bar{J}} \Phi_H^\circ(\xi)^n$, and therefore

$$\begin{aligned} |\Phi_H(n^{-1/2}\xi')^n - e^{-2|\xi'|^2}| &\leq |e^{in^{1/2}\xi' \cdot \bar{J}} - 1| e^{-|\xi'|^2} + |\Phi_H^\circ(n^{-1/2}\xi')^n - e^{-\xi'^T \Sigma_H \xi' / 2}| \\ &\quad + |e^{-\xi'^T \Sigma_H \xi' / 2} - e^{-2|\xi'|^2}|. \end{aligned} \quad (8.108)$$

We bound the three terms on the right side in turn. By Lemma 8.4, the first term satisfies

$$|e^{in^{1/2}\xi' \cdot \bar{J}} - 1| e^{-|\xi'|^2} \leq n^{1/2} |\xi'| |\bar{J}| e^{-|\xi'|^2} \leq C \tau^{1/2} N^{-1/2} |\xi'| e^{-|\xi'|^2}. \quad (8.109)$$

The two exponents in the third term are at least $3|\xi'|^2/2$, so the mean value theorem and Lemma 8.4 give

$$|e^{-\xi'^T \Sigma_H \xi' / 2} - e^{-2|\xi'|^2}| \leq \frac{1}{2} \|\Sigma_H - 4\text{Id}_{2k}\| |\xi'|^2 e^{-3|\xi'|^2/2} \leq C N^{3\varepsilon_k - 1/2} |\xi'|^2 e^{-3|\xi'|^2/2}. \quad (8.110)$$

To bound the second term in (8.108), we set

$$z' = \Phi_H^\circ(n^{-1/2}\xi'), \quad z'' = e^{-\xi'^T \Sigma_H \xi' / (2n)}.$$

The bound (8.105) and $\Sigma_H \geq 3\text{Id}_{2k}$ give

$$\max\{|z'|, |z''|\} \leq e^{-|\xi'|^2/n}. \quad (8.111)$$

Using (8.103) and $|e^{-t'} - 1 + t'| \leq t'^2/2$ for $t' \geq 0$, we find

$$|z' - z''| \leq \frac{1}{2} \left(\frac{5|\xi'|^2}{2n} \right)^2 + \frac{4}{3} C_{\text{mom}} N^{9\varepsilon_k} \frac{|\xi'|^3}{n^{3/2}} \leq \frac{4|\xi'|^4}{n^2} + \frac{2C_{\text{mom}} N^{9\varepsilon_k} |\xi'|^3}{n^{3/2}}. \quad (8.112)$$

We now use the elementary inequality

$$|z'^n - z''^n| \leq n |z' - z''| \max\{|z'|, |z''|\}^{n-1} \leq n |z' - z''| e^{-|\xi'|^2(n-1)/n} \leq n |z' - z''| e^{-|\xi'|^2/2}, \quad (8.113)$$

where we used $n \geq 2$ in the last step, and we obtain

$$|\Phi_H^\circ(n^{-1/2}\xi')^n - e^{-\xi'^T \Sigma_H \xi' / 2}| \leq \left(\frac{4|\xi'|^4}{n} + \frac{2C_{\text{mom}} N^{9\varepsilon_k} |\xi'|^3}{n^{1/2}} \right) e^{-|\xi'|^2/2}. \quad (8.114)$$

We now collect these bounds. Using $\tau/10 \leq n \leq \tau/3$, $\tau = N^{1/100}$, and $\tau^{1/2}N^{-1/2} + N^{3\epsilon_k-1/2} \leq 2\tau^{-1/2}$, we find that (8.107) is bounded by

$$CN^{9\epsilon_k}\tau^{-1/2}(1 + |\zeta'|)^4 e^{-|\zeta'|^2/2} = CN^{9\epsilon_k-1/200}(1 + |\zeta'|)^4 e^{-|\zeta'|^2/2}. \quad (8.115)$$

Integrating over $\zeta' \in \mathbb{R}^{2k}$, we conclude that

$$\mathcal{I}_1 \leq CN^{9\epsilon_k-1/200}, \quad (8.116)$$

which tends to zero by (8.81).

Next, we bound $\mathcal{I}_2$ using Lemma 8.12. For $\varrho_N < |\zeta| \leq \Omega^{1/2}$, the bound (8.86) gives

$$|\Phi_H(\zeta)|^n \leq \left(1 - \frac{p_Y}{2}\tau^{-1/4}\right)^{n/2} \leq \exp\left(-\frac{p_Y}{4}n\tau^{-1/4}\right) \leq \exp\left(-\frac{p_Y}{40}\tau^{3/4}\right). \quad (8.117)$$

Using (8.94) and $n \leq \tau$, we obtain

$$\mathcal{I}_2 \leq C\tau^k \exp(-p_Y\tau^{3/4}/40). \quad (8.118)$$

For large frequencies, we use the first bound in (8.93) and $n \leq \tau$, which give

$$\mathcal{I}_3 \leq C\tau^k \int_{|\zeta| > \Omega^{1/2}} |\zeta|^{-2s_k} d\zeta = C\tau^k \Omega^{k-s_k} = C\tau^{k+(k-s_k)/(4q_k)} = C\tau^{-1/(4q_k)}, \quad (8.119)$$

where the last equality follows from the definition of s_k . Finally, changing variables as for $\mathcal{I}_1$, we find

$$\mathcal{I}_4 = |\hat{\psi}(0)| \int_{|\zeta'| > \varrho_N n^{1/2}} e^{-2|\zeta'|^2} d\zeta' \leq Ce^{-\varrho_N^2 n} \leq C \exp(-cN^{1/100-18\epsilon_k}). \quad (8.120)$$

The four bounds tend to zero, which proves (8.88).

It remains to prove (8.89). Since B is bounded, there exists a nonnegative $\psi_B \in C_c^\infty(\mathbb{R}^{2k})$ with $\psi_B \geq 1$ on B . Applying (8.88) with $\Psi = \{\psi_B\}$, we obtain

$$\mathbb{P}(S^{(n)} + x \in B \mid H) \leq \mathbb{E}[\psi_B(S^{(n)} + x) \mid H] \leq (8\pi n)^{-k} \int \psi_B + o(\tau^{-k}). \quad (8.121)$$

The deterministic quantity that bounds the error term for the family $\{\psi_B\}$ is finite for every N and $o(\tau^{-k})$, so it is bounded by $C\tau^{-k}$ for all N . Since $n \geq \tau/10$, the right side is therefore bounded by $C_B\tau^{-k}$. $\square$

8.6 Pairs with large gaps

The second claim of Proposition 4.4 concerns a single pair. In Lemma 8.14 and its proof, we apply the preceding results with $k = 1$. In particular, $\mathcal{A}_{N,1}$ denotes the event of

Definition 8.11 for $k = 1$, and for $i \in [N - 1]$ the objects J , $\bar{J}$, and $S^{(n)}$ refer to the 1-tuple $\mathbf{i} = (i)$. Recall that $z_i = (-g_i/(2d), 0)$ and $\ell = d\tau^{1/2}$, so that

$$|z_i| = \frac{g_i}{2d} = \frac{g_i}{2\ell} \tau^{1/2}. \quad (8.122)$$

When $g_i \geq \ell$, the vector z_i lies at distance at least $\tau^{1/2}/2$ from the origin, so $z_i + S_i$ can be small only if S_i makes a large deviation. The following lemma quantifies this.

Lemma 8.14. *There exists $c > 0$ such that for every $x_0 > 0$ there exists $C_{x_0} > 0$ with the following property. On $\mathcal{A}_{N,1}$, for every $i \in [N - 1]$ such that $g_i \geq \ell$,*

$$\mathbb{P}(|z_i + S_i| \leq x_0 \mid H) \leq C_{x_0} \tau^{-1} e^{-c g_i / \ell}. \quad (8.123)$$

Proof. Let $N_1 \geq N_0$ be an integer depending on x_0 , which will be chosen below; until the last paragraph of the proof, we assume that $N \geq N_1$. We begin with a bound on the moment generating function of one increment. The increment is unbounded because the fresh variable Y is, so we restrict the parameter of the moment generating function to a range in which only $\mathbb{E}e^{|Y|/2} < \infty$ is needed. Set $\beta_0 = N^{-3\varepsilon_1}/4$, and let $\beta \in \mathbb{R}$ with $|\beta| \leq \beta_0$. By Lemma 8.4 with $k = 1$, we have on $\mathcal{A}_{N,1}$ that, for $\eta \in \{1, 2\}$,

$$|J(\eta)| \leq |J| \leq 2N^{2\varepsilon_1}(|Y| + N^{\varepsilon_1}), \quad \text{and hence} \quad |\beta J(\eta)| \leq \frac{|Y| + 1}{2}. \quad (8.124)$$

Using $e^{t'} \leq 1 + t' + t'^2 e^{|t'|}/2$ for real t' , we obtain

$$\mathbb{E}[e^{\beta J(\eta)} \mid H] \leq 1 + \beta \bar{J}(\eta) + \frac{e^{1/2}}{2} \beta^2 \mathbb{E}[J(\eta)^2 e^{|Y|/2} \mid H]. \quad (8.125)$$

To bound the last expectation, recall from (8.27) that

$$\mathbb{E}[J(\eta)^2 \mid H] = \binom{N}{2}^{-1} \sum_{a < b} (1 + X_{ab}^2) (D_{ab}^{(i)}(\eta))^2, \quad (8.126)$$

and that this quantity is bounded by 5 by Lemma 8.4, provided N_1 is large enough. Conditioning on (a, b) and using $(Y - X_{ab})^2 \leq 2Y^2 + 2X_{ab}^2$, we find

$$\begin{aligned} \mathbb{E}[J(\eta)^2 e^{|Y|/2} \mid H] &\leq 2 \binom{N}{2}^{-1} \sum_{a < b} (D_{ab}^{(i)}(\eta))^2 \left(\mathbb{E}[Y^2 e^{|Y|/2}] + X_{ab}^2 \mathbb{E}[e^{|Y|/2}] \right) \\ &\leq C \mathbb{E}[J(\eta)^2 \mid H] \leq C, \end{aligned} \quad (8.127)$$

where in the second inequality we used that Y is sub-Gaussian. Inserting this into (8.125) and using $1 + t' \leq e^{t'}$, we obtain, for some constant $C_* > 0$,

$$\mathbb{E}[e^{\beta(J(\eta) - \bar{J}(\eta))} \mid H] \leq e^{-\beta \bar{J}(\eta)} (1 + \beta \bar{J}(\eta) + C_* \beta^2) \leq e^{C_* \beta^2}, \quad |\beta| \leq \beta_0. \quad (8.128)$$

We next derive a tail bound for sums. Let $n \geq 1$ and $y > 0$. Conditionally on H , the vector $S^{(n)}$ is a sum of n independent copies of J . Hence (8.128) and Markov's inequality give for $0 < \beta \leq \beta_0$, $\eta \in \{1, 2\}$, and either choice of sign,

$$\mathbb{P}\left(\pm(S^{(n)} - n\bar{J})(\eta) \geq \frac{y}{\sqrt{2}} \mid H\right) \leq \exp\left(-\frac{\beta y}{\sqrt{2}} + C_* n \beta^2\right). \quad (8.129)$$

We choose

$$\beta = \begin{cases} y/(2\sqrt{2}C_*n) & \text{if } y/(2\sqrt{2}C_*n) \leq \beta_0, \\ \beta_0 & \text{otherwise.} \end{cases} \quad (8.130)$$

In the second case, we have $C_* n \beta_0^2 \leq \beta_0 y/(2\sqrt{2})$. Further, one of the two coordinates of $S^{(n)} - n\bar{J}$ has modulus at least $|S^{(n)} - n\bar{J}|/\sqrt{2}$. Summing over the four cases, we conclude that

$$\mathbb{P}(|S^{(n)} - n\bar{J}| \geq y \mid H) \leq 4 \exp\left(-\min\left\{\frac{y^2}{8C_*n}, \frac{\beta_0 y}{2\sqrt{2}}\right\}\right). \quad (8.131)$$

We now split the m resampling steps into two halves. Recall that m is even, and write $S_i = S' + S''$, where S' is the sum of the first $m/2$ summands in (4.12) and S'' is the sum of the remaining $m/2$ summands. Conditionally on H , the vectors S' and S'' are independent, and each has the conditional law of $S^{(m/2)}$. We also take N_1 large enough that for $N \geq N_1$ we have $\tau/10 \leq m/2 \leq \tau/3$, $m|\bar{J}| \leq 1$ on $\mathcal{A}_{N,1}$ (which is possible by Lemma 8.4), and $\tau^{1/2} \geq 4(x_0 + 1)$.

Suppose that $|z_i + S_i| \leq x_0$. Using (8.122) and $g_i \geq \ell$, we find $|z_i| \geq \tau^{1/2}/2$, and therefore

$$\left|S' - \frac{m}{2}\bar{J}\right| + \left|S'' - \frac{m}{2}\bar{J}\right| \geq |S' + S'' - m\bar{J}| \geq |z_i| - x_0 - 1 \geq \frac{|z_i|}{2}, \quad (8.132)$$

where the last inequality follows from $|z_i| \geq \tau^{1/2}/2$ and $\tau^{1/2} \geq 4(x_0 + 1)$. Hence

$$|S' - (m/2)\bar{J}| \geq |z_i|/4 \quad \text{or} \quad |S'' - (m/2)\bar{J}| \geq |z_i|/4. \quad (8.133)$$

Let $B_{x_0} \subset \mathbb{R}^2$ be the closed disk of radius x_0 centered at 0, and let $C_{B_{x_0}}$ be the constant of (8.89) for $k = 1$. To bound the probability of the first event, we condition on S' and use that $|z_i + S_i| \leq x_0$ holds only if $S'' + z_i + S' \in B_{x_0}$. The vector S'' is conditionally independent of S' , so (8.89) with $n = m/2$, followed by (8.131), gives

$$\mathbb{P}\left(|z_i + S_i| \leq x_0, \left|S' - \frac{m}{2}\bar{J}\right| \geq \frac{|z_i|}{4} \mid H\right) \leq \frac{C_{B_{x_0}}}{\tau} \mathbb{P}\left(|S^{(m/2)} - \frac{m}{2}\bar{J}| \geq \frac{|z_i|}{4} \mid H\right). \quad (8.134)$$

The second event is treated in the same way, with the roles of S' and S'' exchanged.

We now evaluate (8.131) with $n = m/2 \leq \tau/8$ and

$$y = \frac{|z_i|}{4} = \frac{g_i \tau^{1/2}}{8\ell}. \quad (8.135)$$

Our choice of ε_1 in (8.81) gives $\varepsilon_1 \leq 1/1200$, and hence $\tau^{1/2}N^{-3\varepsilon_1} \geq 1$. Therefore

$$\frac{y^2}{8C_*n} \geq \frac{1}{64C_*} \left(\frac{g_i}{\ell}\right)^2, \quad \frac{\beta_0 y}{2\sqrt{2}} = \frac{\tau^{1/2}N^{-3\varepsilon_1}}{64\sqrt{2}} \frac{g_i}{\ell} \geq \frac{1}{64\sqrt{2}} \frac{g_i}{\ell}. \quad (8.136)$$

Since $g_i/\ell \geq 1$, the minimum in (8.131) is at least cg_i/ℓ , where

$$c = \min\{(64C_*)^{-1}, (64\sqrt{2})^{-1}\}. \quad (8.137)$$

Combining the two events, we conclude that

$$\mathbb{P}(|z_i + S_i| \leq x_0 \mid H) \leq \frac{8C_{B_{x_0}}}{\tau} e^{-cg_i/\ell}, \quad N \geq N_1. \quad (8.138)$$

Finally, let $N < N_1$. On $\mathcal{A}_{N,1}$, we have $g_i \leq 2\|H\| \leq 6$ and $\ell \geq N^{-3/2} \geq N_1^{-3/2}$, so

$$\tau e^{cg_i/\ell} \leq N_1 \exp(6cN_1^{3/2}). \quad (8.139)$$

Hence (8.123) holds for all N with $C_{x_0} = \max\{8C_{B_{x_0}}, N_1 \exp(6cN_1^{3/2})\}$. $\square$

Proof of Proposition 4.4. We show that the events $\mathcal{A}_{N,k}$ of Definition 8.11 have the properties stated in Proposition 4.4, with the constant c of Lemma 8.14. By Lemma 8.12, we have $\mathbb{P}(\mathcal{A}_{N,k}) \rightarrow 1$. These events are measurable with respect to H . Indeed, $\mathcal{G}_{\varepsilon_k}$ is a Borel set, the other two conditions in (8.9) are closed conditions on H , and the set of matrices A that have a unit eigenvector w violating the inequality in Definition 8.11 is closed, as the projection along the compact unit sphere of the closed set of pairs (A, w) such that $\|w\|_2 = 1$, $Aw = (w^T Aw)w$, and w violates this inequality.

We now prove the local limit theorem. Let Ψ be a family of test functions as in the statement of Proposition 4.4. The supports of its members lie in a common ball $\{y \in \mathbb{R}^{2k} : |y| \leq R_\Psi\}$, so Ψ satisfies the hypotheses of Lemma 8.13, and (8.90) holds by the proof of that lemma. Let ϵ'_N denote the deterministic quantity of Lemma 8.13 that bounds the error term in (8.88) for this family Ψ . By Lemma 8.13, ϵ'_N is finite for every N , and $\tau^k \epsilon'_N \rightarrow 0$. We define

$$\epsilon_N = \begin{cases} \tau^k \epsilon'_N + 16(k+1)\pi^{-k}\tau^{-1} \sup_{\psi \in \Psi} \int_{\mathbb{R}^{2k}} |\psi(y)| dy & \text{if } \tau \geq 16, \\ \sup_{\psi \in \Psi} (\tau^k \|\psi\|_\infty + (2\pi)^{-k} \int_{\mathbb{R}^{2k}} |\psi(y)| dy) & \text{if } \tau < 16. \end{cases} \quad (8.140)$$

Since ϵ'_N is finite, (8.90) shows that the numbers ϵ_N are finite. They depend only on k and Ψ , and they tend to zero because $\tau^k \epsilon'_N \rightarrow 0$. Next, we show that (4.15) holds on $\mathcal{A}_{N,k}$.

If $\tau < 16$, then (4.15) holds without any restriction on H , since $|\mathbb{E}[\psi(S_i + x) \mid H]| \leq \|\psi\|_\infty$ and $e^{-|x|^2/(2\tau)} \leq 1$. Now suppose that $\tau \geq 16$. Recall from Section 8.1 that the conditional law of S_i given H is that of $S^{(m)}$. Since $m = 2\lfloor \tau/8 \rfloor$, we have $\tau - 8 < 4m \leq \tau$,

and hence $\tau/10 \leq m \leq \tau/3$ and $\tau/2 \leq 4m \leq \tau$. By Lemma 8.13 with $n = m$, on $\mathcal{A}_{N,k}$, we have

$$\left| \mathbb{E}[\psi(S_{\mathbf{i}} + x) \mid H] - \mathfrak{g}_{4m\text{Id}_{2k}}(x) \int \psi \right| \leq \epsilon'_N \quad (8.141)$$

for all k -tuples $\mathbf{i}$, all $x \in \mathbb{R}^{2k}$, and all $\psi \in \Psi$. It remains to replace $4m$ by τ . Differentiating $\mathfrak{g}_{r'\text{Id}_{2k}}(x)$ with respect to $r' > 0$ and using $y'e^{-y'} \leq 1$ for $y' \geq 0$, we find

$$|\partial_{r'} \mathfrak{g}_{r'\text{Id}_{2k}}(x)| \leq \frac{k+1}{r'} (2\pi r')^{-k}, \quad x \in \mathbb{R}^{2k}.$$

The numbers $4m$ and τ lie in $[\tau/2, \tau]$ and differ by at most 8, so the mean value theorem gives

$$\sup_{x \in \mathbb{R}^{2k}} |\mathfrak{g}_{4m\text{Id}_{2k}}(x) - \mathfrak{g}_{\tau\text{Id}_{2k}}(x)| \leq \frac{16(k+1)}{\tau} (\pi\tau)^{-k}. \quad (8.142)$$

Inserting this into (8.141), multiplying by τ^k , and using $|\int \psi| \leq \int |\psi|$ and

$$\tau^k \mathfrak{g}_{\tau\text{Id}_{2k}}(x) = (2\pi)^{-k} e^{-|x|^2/(2\tau)}, \quad (8.143)$$

we obtain (4.15) with ϵ_N as in (8.140).

The second claim of Proposition 4.4 concerns the events $\mathcal{A}_{N,1}$. It follows from Lemma 8.14, applied with $x_0/2$ in place of x_0 . This completes the proof. $\square$

Complex Hermitian matrices

In this section, we sketch the proof of Theorem 1.5. The argument uses the comparison estimate of Section 5 and the induction on the scale of Section 6, but not the resampling argument of Sections 7 and 8; we explain the reason at the beginning of Section 9.2. We indicate the changes to the arguments of the previous sections and give the details of the two estimates that are new. The first is a relative bound on the number of triples of nearby eigenvalues, Proposition 9.2, and the second is a bound on the probability that a Gaussian perturbation of a 3×3 diagonal matrix has a small gap, Lemma 9.3. We prove both in full, given the inputs collected in Section 9.3. Where we only sketch an argument, we name the result of the earlier sections whose proof we follow and the properties that this proof uses, and we explain why these properties hold in the complex Hermitian case.

9.1 Setup

Throughout this section, H is a complex Hermitian Wigner matrix as in Definition 1.4, and $\lambda_1 \leq \dots \leq \lambda_N$, g_i , and $c_2(E) = (\pi^2/3)\rho_{\text{sc}}(E)^4$ are as in Theorem 1.5; we also let $E_i = (\lambda_i + \lambda_{i+1})/2$. We use the notation of the previous sections with the following changes. In this section, G denotes a complex Hermitian matrix from the Gaussian unitary ensemble (GUE), independent of H , whose entries G_{ab} with $a < b$ are independent complex Gaussian variables with $\mathbb{E}G_{ab} = \mathbb{E}G_{ab}^2 = 0$ and $\mathbb{E}|G_{ab}|^2 = 1/N$, and whose diagonal entries are independent real Gaussian variables with variance $1/N$, independent of the entries above the diagonal. The Ornstein–Uhlenbeck interpolation H_t is defined by (3.2) with this G . We set

$$d_2 = N^{-4/3}, \quad M_\alpha^{\mathbb{C}} = N^4 \ell_\alpha^3 = N^{1-3\alpha} \quad (\alpha \in (0, 1/3)). \quad (9.1)$$

Theorem 1.5 states that the smallest gaps in the bulk are of order d_2 , and we call d_2 the *critical scale*. By Lemma 9.1 below, for $\alpha \in (0, 1/4)$, the number of gaps of the GUE of size at most $\ell_\alpha = N^{-1-\alpha}$ with midpoints in a bulk interval is of order $M_\alpha^{\mathbb{C}}$.

Let $p \geq 1$ be an integer, let $\chi \in C_c^\infty((-2, 2))$ be nonnegative, and let $q > 0$. We define the triple and quadruple statistics of a Hermitian matrix A by

$$\begin{aligned} U_q^{(3)}(A) &= \sum_{i < j < k} \chi_N \left(\frac{\lambda_i(A) + \lambda_j(A) + \lambda_k(A)}{3} \right) \prod_{\{i', i''\} \subset \{i, j, k\}} f_p \left(\frac{\lambda_{i'}(A) - \lambda_{i''}(A)}{q} \right), \\ U_q^{(4)}(A) &= \sum_{i < j < k < l} \chi_N \left(\frac{\lambda_i(A) + \lambda_j(A) + \lambda_k(A) + \lambda_l(A)}{4} \right) \\ &\quad \times \prod_{\{i', i''\} \subset \{i, j, k, l\}} f_p \left(\frac{\lambda_{i'}(A) - \lambda_{i''}(A)}{q} \right), \end{aligned} \quad (9.2)$$

where χ_N is the smoothing (5.2) of χ , and the products are over the three, respectively six, two-element subsets. We suppress the dependence on p and χ in the notation. Observe that if $\lambda_k(A) - \lambda_i(A) \leq 4q$ and $\chi_N \geq 1$ at the mean of $\lambda_i(A)$, $\lambda_j(A)$, and $\lambda_k(A)$, then the term of (i, j, k) in $U_q^{(3)}(A)$ is at least 17^{-3p} , since each of its three rational factors is at least $f_p(4) = 17^{-p}$. Similarly, if $\lambda_l(A) - \lambda_i(A) \leq 6q$ and $\chi_N \geq 1$ at the mean of the four eigenvalues, then the term of (i, j, k, l) in $U_q^{(4)}(A)$ is at least 37^{-6p} , since $f_p(6) = 37^{-p}$.

9.2 Strategy

The smallest gaps of H are of order $d_2 = N^{-4/3}$. A single entry moves the eigenvalues by about $N^{-3/2} = o(d_2)$, so the strip estimate of Lemma 5.2 holds at the scale d_2 , which corresponds to the exponent $\alpha = 1/3 < 1/2$. Consequently, the comparison estimate of Proposition 5.9 applies to pair statistics at the scale d_2 , with the error

$$CtN^{1/2}(Nd_2)^{-3}N^{C\varepsilon} = CtN^{3/2+C\varepsilon}. \quad (9.3)$$

This is small if $t = t_{\alpha_1} = N^{-1-2\alpha_1}$ with $\alpha_1 > 1/4$. For such t , the Gaussian part of H_t is of order $\ell_{\alpha_1} = N^{-1-\alpha_1}$, and the computation of Section 6 through independent 2×2 Gaussian blocks determines the gaps of H_t at the scale d_2 , provided that the gap counts of H at the scale ℓ_{α_1} are known and that eigenvalues of H forming clusters of three or more at a threshold $q = N^{-1-\kappa}$ do not produce gaps of order d_2 . The gap counts are supplied by an induction on the scale as in Section 6, which reaches the exponent $\alpha_1 = 8/25$ that we use.

We do not use the resampling argument, which would not be available in the same form. Heuristically, a resampling step moves the traceless part of the 2×2 compression of the matrix to the span of u_i and u_{i+1} (described by the map L_i of (4.8), with a third coordinate in the complex Hermitian case) by $O(N^{-3/2}) = o(d_2)$, so order $N^{1/3}$ steps would be needed to move it by d_2 , which is beyond the range of the crude entrywise bound (2.9).

The main difference with Section 6 concerns clusters of three eigenvalues. In Section 6, we used the bound of Nguyen, Tao, and Vu on three consecutive eigenvalues, Proposition 3.4. In the complex Hermitian case, the expected number of triples of span $N^{-1-\kappa}$ for the GUE is of order $N(N^{-\kappa})^8$, and triples become absent when $\kappa > 1/8$. However, transferring information on the scale $N^{-1-\kappa}$ from the GUE to H through the coupling of Proposition 3.5 and the comparison of Proposition 5.9 requires a time T with

$$N^{-1+\kappa} \ll T \ll N^{-1/2-3\kappa}.$$

The lower bound makes the coupling error $N^\varepsilon/(N^2T)$ smaller than $N^{-1-\kappa}$, and the upper bound makes the comparison error, which is of order $TN^{1/2+3\kappa}$, small. Such a time exists only for $\kappa < 1/8$, so triples cannot be excluded in this way.

We therefore use relative bounds. For $\kappa < 1/8$, the number of triples of span at most $4N^{-1-\kappa}$ in the bulk is at most $N^{1-8\kappa+\varepsilon}$ with probability tending to one, by Proposition 9.2. In the induction step, this is negligible compared to $M_{\alpha'}^{\mathbb{C}} = N^{1-3\alpha'}$ when $8\kappa > 3\alpha'$. In the final step, Lemma 9.3 shows that each triple produces a gap of order d_2 with probability bounded by $N^{C\varepsilon}(d_2/\ell_{\alpha_1})^3 = N^{-1+3\alpha_1+C\varepsilon}$, and the expected number of such gaps is at most $N^{3\alpha_1-8\kappa+C\varepsilon}$, which tends to zero when $8\kappa > 3\alpha_1$. Since $\alpha_1 < 1/3$, this is compatible with $\kappa < 1/8$.

9.3 Inputs

Regularity and coupling. In this section, $\mathcal{G}_\varepsilon$ denotes the set of complex Hermitian matrices with the three properties of Definition 3.1, where unit eigenvectors are taken in $\mathbb{C}^N$. The estimates of Proposition 3.2 and Lemma 3.3 hold for H_t and G with the same proofs, since [4, Definition 2.2] concerns complex Hermitian matrices and [29, Theorem 2.2] covers complex Hermitian generalized Wigner matrices. These proofs use the entries only through their independence, their centering, their sub-Gaussian tails, and their variances, which are now $\mathbb{E}|(H_t)_{ab}|^2 = 1/N$ for $a \neq b$ and $\mathbb{E}(H_t)_{aa}^2 = (e^{-t}\mathbb{E}Z^2 + 1 - e^{-t})/N$.

The coupling of Proposition 3.5 holds with GUE in place of GOE. In this case, B in (3.11) is a complex Hermitian matrix of Brownian motions whose diagonal entries are standard real Brownian motions and whose entries above the diagonal are standard complex Brownian motions with $\mathbb{E}|B_{ab}(t)|^2 = t$, all independent, and the Brownian motions in (3.12) satisfy $\mathbb{E}\beta_i(t)^2 = t$. This is the dynamics with symmetry parameter 2 in [10, (2.2)], whose equilibrium measure is the GUE, and [10, Theorem 3.1 and Corollary 3.3] are stated for this dynamics as well; by [10, Remark 3.2], they require only optimal rigidity estimates for the initial conditions. The verification of these estimates in the proof of Proposition 3.5 applies without change, since it uses only Lemma 3.3 and Proposition 3.2.

One step requires a modification. In the proof of Proposition 3.5, the dynamics is started from H , whose spectrum is simple with overwhelming probability by [57,

Corollary 2.2]. The extension of this result to Hermitian matrices stated in [57, Section 2] assumes that the real and imaginary parts of the entries are independent and identically distributed, which we do not assume. We therefore start the dynamics from H_{t_*} with $t_* = N^{-10}$. Conditionally on H , the matrix H_{t_*} has a density with respect to Lebesgue measure on the space of Hermitian matrices, and the Hermitian matrices with a multiple eigenvalue form a Lebesgue null set (the zero set of the discriminant of the characteristic polynomial), so the spectrum of H_{t_*} is simple almost surely.

Moreover, for $t' \geq 0$, we have

$$H_{t_*+t'} \stackrel{d}{=} e^{-t'/2} H_{t_*} + (1 - e^{-t'})^{1/2} \tilde{G},$$

where $\tilde{G}$ is a GUE matrix independent of H_{t_*} , and the rigidity estimates for these matrices follow from Lemma 3.3, whose proof applies for all times in $[0, 2]$. Applying [10, Corollary 3.3] with the initial condition $\lambda(H_{t_*})$ at time $t - t_*$, and using $t - t_* \geq t/2$, we obtain (3.13) with $2N^\varepsilon/(N^2t)$ in place of $N^\varepsilon/(N^2t)$, for all t with $N^{-1+\varepsilon} \leq t < 1$, and the spectra of H_t and of a GUE matrix can be coupled accordingly.

Comparison. The definition of φ_N , the strip estimate of Lemma 5.2, the holomorphic extension of Definition 5.4 and Lemma 5.5, and the derivative bound of Lemma 5.6 carry over verbatim to complex Hermitian matrices, since their proofs use only the spectral decomposition of a Hermitian matrix, Sylvester's determinant identity, and the properties of $\mathcal{G}_\varepsilon$. For $a < b$, we let $\tilde{V}^{ab} = i(e_a e_b^T - e_b e_a^T)$, so that the entry $(H_t)_{ab}$ contributes $\operatorname{Re}(H_t)_{ab} V^{ab} + \operatorname{Im}(H_t)_{ab} \tilde{V}^{ab}$ to H_t . For complex ζ_1 and ζ_2 , we have

$$\zeta_1 V^{ab} + \zeta_2 \tilde{V}^{ab} = (\zeta_1 + i\zeta_2) e_a e_b^T + (\zeta_1 - i\zeta_2) e_b e_a^T,$$

and the proof of Lemma 5.2, which uses only the factorization of a perturbation supported on the rows and columns a and b through Sylvester's identity, applies verbatim to perturbations $\zeta e_a e_b^T + \zeta' e_b e_a^T$ with $|\zeta|, |\zeta'| \leq 8c_p N^{-\alpha-4\varepsilon}$.

In Lemma 5.7, the sum runs over the diagonal entries, with $v_{aa}^G = 1/N$, and over the real and imaginary parts of the entries above the diagonal, with the directions V^{ab} and $\tilde{V}^{ab}$ and with $v^G = 1/(2N)$ for each of them; its proof uses only Gaussian integration by parts in the independent real coordinates of G .

The real and imaginary parts of X_C need not be independent, so in Steps 2 and 3 of the proof of Proposition 5.9, we treat the two terms of each entry above the diagonal together. For $t' \in (0, 1]$ and $a < b$, let $H_{t'}^{[ab]}$ denote the matrix obtained from $H_{t'}$ by setting its (a, b) and (b, a) entries to zero. This matrix is independent of $(\operatorname{Re}(H_{t'})_{ab}, \operatorname{Im}(H_{t'})_{ab})$. Let $\mathcal{E}_{ab}(t')$ be the event that $\operatorname{Spec}(H_{t'}^{[ab]} + \zeta_1 V^{ab} + \zeta_2 \tilde{V}^{ab}) \subset \mathfrak{R}_{\alpha,p}$ for all complex ζ_1, ζ_2 with $|\zeta_1|, |\zeta_2| \leq 2r_N$. As in Step 2 of the proof of Proposition 5.9, this event is determined by $H_{t'}^{[ab]}$. By the preceding paragraphs, it contains $\{H_{t'} \in \mathcal{G}_\varepsilon\} \cap \{ |(H_{t'})_{ab}| \leq r_N \}$, since the corresponding perturbation of $H_{t'}$ has coefficients of modulus at most $6r_N \leq 8c_p N^{-\alpha-4\varepsilon}$.

(Indeed, $H_{t'}^{[ab]} + \zeta_1 V^{ab} + \zeta_2 \tilde{V}^{ab}$ is obtained from $H_{t'}$ by adding $\zeta_1 - \operatorname{Re}(H_{t'})_{ab}$ times V^{ab} and $\zeta_2 - \operatorname{Im}(H_{t'})_{ab}$ times $\tilde{V}^{ab}$, and both coefficients have modulus at most $3r_N$.) Hence the complement of $\mathcal{E}_{ab}(t')$ has probability at most $C_D N^{-D}$, as in (5.52).

On $\mathcal{E}_{ab}(t')$, the function $(\zeta_1, \zeta_2) \mapsto F(H_{t'}^{[ab]} + \zeta_1 V^{ab} + \zeta_2 \tilde{V}^{ab})$ is holomorphic and bounded by one in modulus on the closed polydisk of radius $2r_N$, by Lemma 5.5. Cauchy's estimates on the polydisks of radius r_N centered at the points of $[-r_N, r_N]^2$ bound its partial derivatives of order n on $[-r_N, r_N]^2$ by $C_n r_N^{-n}$. We now Taylor expand jointly in the two real variables. We write $\mathbf{x} = (\operatorname{Re}(H_{t'})_{ab}, \operatorname{Im}(H_{t'})_{ab})$ and

$$Y(x_1, x_2) = F(H_{t'}^{[ab]} + x_1 V^{ab} + x_2 \tilde{V}^{ab}), \quad (x_1, x_2) \in \mathbb{R}^2.$$

The contribution of the entry (a, b) to the integrand in (5.36) is then $\mathbb{E}[(2N)^{-1}(\partial_1^2 + \partial_2^2)Y(\mathbf{x}) - \mathbf{x} \cdot \nabla Y(\mathbf{x})]$, where ∇ denotes the gradient in (x_1, x_2) , and ∂_1 and ∂_2 denote the partial derivatives in x_1 and x_2 . The vector $(\operatorname{Re} X_C, \operatorname{Im} X_C)$ is centered with covariance $2^{-1} \operatorname{Id}_2$, since $\mathbb{E} X_C^2 = 0$ and $\mathbb{E} |X_C|^2 = 1$, and this agrees with the covariance of $N^{1/2}(\operatorname{Re} G_{ab}, \operatorname{Im} G_{ab})$. Hence $\mathbb{E} \mathbf{x} = 0$ and $\mathbb{E} \mathbf{x} \mathbf{x}^T = (2N)^{-1} \operatorname{Id}_2$.

We expand ∇Y to first order and $(\partial_1^2 + \partial_2^2)Y$ to zeroth order around the origin, on the event $\{|\mathbf{x}| \leq r_N\}$, as in the proof of Lemma 5.8. By the last two identities, the terms of order zero and one cancel, and we obtain, on $\mathcal{E}_{ab}(t')$,

$$\left| \mathbb{E} \left[\frac{(\partial_1^2 + \partial_2^2)Y(\mathbf{x})}{2N} - \mathbf{x} \cdot \nabla Y(\mathbf{x}) \mid H_{t'}^{[ab]} \right] \right| \leq C r_N^{-3} \mathbb{E} |\mathbf{x}|^3 + C_D N^{-D} \leq C N^{-3/2} r_N^{-3} + C_D N^{-D}.$$

Here the term $C_D N^{-D}$ bounds the contribution of the event $\{|\mathbf{x}| > r_N\}$, which we control with the crude bound of Lemma 5.6 as in Step 3 of the proof of Proposition 5.9; the complement of $\mathcal{E}_{ab}(t')$ is treated in the same way. The diagonal entries are treated as in the proof of Proposition 5.9, and the variance mismatch $|\mathbb{E}(H_{t'})_{aa}^2 - 1/N| \leq C N^{-1}$ gives the term r_N^{-2} . Summing over the coordinates as in Step 4 of that proof, we obtain the analogue of (5.47) for complex Hermitian matrices,

$$|\mathbb{E} F(H_t) - \mathbb{E} F(H)| \leq C t (N^{1/2} r_N^{-3} + r_N^{-2}) + C_D N^{-D} \leq C t N^{1/2+3\alpha+12\epsilon} + C_D N^{-D} \quad (9.4)$$

for F as in Definition 5.4, $\alpha + 5\epsilon < 1/2$, and $r_N = c_p N^{-\alpha-4\epsilon}$.

Triple and quadruple statistics. The comparison estimate also applies to the statistics (9.2). We first write them as traces. Let $s_3(x_1, x_2, x_3)$ and $s_4(x_1, x_2, x_3, x_4)$ denote the symmetric functions such that the summands of $U_q^{(3)}$ and $U_q^{(4)}$ equal $s_3(\lambda_i, \lambda_j, \lambda_k)$ and $s_4(\lambda_i, \lambda_j, \lambda_k, \lambda_l)$, where $\lambda_i = \lambda_i(A)$. Summing over all tuples of indices and removing those with coincident indices by inclusion–exclusion over the partitions of the set of positions, we find

$$U_q^{(3)}(A) = \frac{1}{6} \left(\sum_{i,j,k} s_3(\lambda_i, \lambda_j, \lambda_k) - 3 \sum_{i,j} s_3(\lambda_i, \lambda_i, \lambda_j) + 2 \sum_i s_3(\lambda_i, \lambda_i, \lambda_i) \right) \quad (9.5)$$

and

$$\begin{aligned}
24 U_q^{(4)}(A) = & \sum_{i,j,k,l} s_4(\lambda_i, \lambda_j, \lambda_k, \lambda_l) - 6 \sum_{i,j,k} s_4(\lambda_i, \lambda_i, \lambda_j, \lambda_k) + 3 \sum_{i,j} s_4(\lambda_i, \lambda_i, \lambda_j, \lambda_j) \\
& + 8 \sum_{i,j} s_4(\lambda_i, \lambda_i, \lambda_i, \lambda_j) - 6 \sum_i s_4(\lambda_i, \lambda_i, \lambda_i, \lambda_i).
\end{aligned} \tag{9.6}$$

The coefficients are the values $\prod_B (-1)^{|B|-1} (|B| - 1)!$ of the Möbius function of the lattice of set partitions, where B runs over the blocks, multiplied by the number of partitions of each type.

Each sum on the right sides of (9.5) and (9.6) is the trace of a function of commuting matrices obtained from A by tensoring with identity matrices, exactly as in Definition 5.4. For example, the first sum in (9.5) is the trace over $(\mathbb{C}^N)^{\otimes 3}$ of

$$\chi_N \left(\frac{A_1 + A_2 + A_3}{3} \right) \prod_{1 \leq i < j \leq 3} (\text{Id} + q^{-2}(A_i - A_j)^2)^{-p},$$

where $A_1 = A \otimes \text{Id} \otimes \text{Id}$, $A_2 = \text{Id} \otimes A \otimes \text{Id}$, and $A_3 = \text{Id} \otimes \text{Id} \otimes A$. As in the proof of Lemma 5.5, these traces are holomorphic functions of the entries of A on the set of matrices whose eigenvalues w satisfy $|\text{Im } w| < \omega_p q$, and the identities (9.5) and (9.6) hold there with $\lambda_1, \dots, \lambda_N$ replaced by the eigenvalues of A , listed with algebraic multiplicity.

Next, we bound the arguments of the summands. Suppose now that $\text{Spec } A$ lies in the rectangle $\mathfrak{R}_{\kappa,p}$ of (5.10), where $q = \ell_\kappa = N^{-1-\kappa}$. The proof of Lemma 5.5 shows that every rational factor in (9.2) has argument of absolute value at most $1/375$, and that the weight χ_N at the mean of the eigenvalues is either zero or has argument of absolute value at most $1/200$; the mean of three or four eigenvalues satisfies the same bounds on its real and imaginary parts as the mean of two. Hence every summand of $U_q^{(3)}(A)$ and $U_q^{(4)}(A)$ is zero or has argument of absolute value at most $1/200 + 6/375 < 1/40$, so $\text{Re } U_q^{(r)}(A) \geq 0$ and $|\exp(-\theta U_q^{(r)}(A)/M)| \leq 1$ for all $\theta, M > 0$ and $r \in \{3, 4\}$. The proof of Lemma 5.6 applies to these observables with ℓ_α replaced by q . Thus the observables $\exp(-\theta U_q^{(r)}/M)$ have the two properties listed in Remark 5.10, with the polydisk above in place of the disk. Hence the proof of Proposition 5.9, with the modifications described in the paragraph on comparison above, gives, for $q = N^{-1-\kappa}$ with $\kappa \in (0, 1/2)$, all $M > 0$, $\theta > 0$, $D > 0$, $t \in [0, 1]$, and $\varepsilon > 0$ with $\kappa + 5\varepsilon < 1/2$,

$$\left| \mathbb{E} e^{-\theta U_q^{(r)}(H_t)/M} - \mathbb{E} e^{-\theta U_q^{(r)}(H)/M} \right| \leq C t N^{1/2+3\kappa+12\varepsilon} + C_D N^{-D}, \quad r \in \{3, 4\}. \tag{9.7}$$

The Gaussian unitary ensemble. The following estimates replace Proposition 3.6.

Lemma 9.1. *The following statements hold.*

- (a) Let $\alpha \in (0, 1/4)$ and $p > 3/2$. Let $f: [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$ be a continuous function for which there exist a bulk interval I and a constant $C_f > 0$ such that $f(r, E) = 0$ for $E \notin I$ and $|f(r, E)| \leq C_f(1 + r^2)^{-p}$ for all (r, E) . Then, in L^1 ,

$$\frac{1}{M_\alpha^C} \sum_{1 \leq i < j \leq N} f\left(\frac{\lambda_j(G) - \lambda_i(G)}{\ell_\alpha}, \frac{\lambda_i(G) + \lambda_j(G)}{2}\right) \longrightarrow \int_{-2}^2 \int_0^\infty c_2(E) r^2 f(r, E) dr dE. \quad (9.8)$$

- (b) Let $\kappa > 0$, set $q = N^{-1-\kappa}$, and let $p \geq 4$ be an integer. Then

$$\mathbb{E}U_q^{(3)}(G) \leq CN(Nq)^8 + o(1), \quad \mathbb{E}U_q^{(4)}(G) \leq CN(Nq)^{15} + o(1). \quad (9.9)$$

Sketch of proof. We use the kernel K_N of Definition A.2 and the estimates for it proved in Appendix A. Recall that the eigenvalues of G form a determinantal point process with kernel K_N . Thus their n -point correlation functions, defined as in (A.11) with the joint density of the eigenvalues of the GUE, are given by

$$R_{n,N}(x_1, \dots, x_n) = \det[K_N(x_i, x_j)]_{i,j=1}^n, \quad n \in [N], \quad (9.10)$$

(see [2, Section 3.2]), and (A.12) holds.

The proof follows Appendix A.5. The determinantal structure makes the correlation estimates simpler than for the GOE, and we give them below; we omit the routine verification of the dominated convergence arguments.

Throughout the proof, $\tilde{I}$ denotes a bulk interval, constants may depend on $\tilde{I}$, and we use Lemma A.5 and Lemma A.7 for $x, y \in \tilde{I}$. We first record two estimates for the correlation functions. We set

$$\mathbf{h}(x) = (N^{1/4}h_j(\sqrt{N}x))_{j=0}^{N-1} \in \mathbb{R}^N, \quad \text{so that} \quad K_N(x, y) = \mathbf{h}(x) \cdot \mathbf{h}(y).$$

The two-point function. We begin with the expansion of $R_{2,N}$ near the diagonal. Since K_N is symmetric, (9.10) with $n = 2$ gives

$$R_{2,N}(x, y) = K_N(x, x)K_N(y, y) - K_N(x, y)^2.$$

Using the symmetry of K_N again, we find

$$\begin{aligned} R_{2,N}(x, x) &= \partial_y R_{2,N}(x, y)|_{y=x} = 0, \\ \partial_y^2 R_{2,N}(x, y)|_{y=x} &= K_N(x, x) \frac{d^2}{dx^2} K_N(x, x) - 2(\partial_2 K_N(x, x))^2 - 2K_N(x, x) \partial_y^2 K_N(x, y)|_{y=x}. \end{aligned}$$

We now bound the first two terms. By the proof of Lemma A.5(c), we have

$$\frac{d}{dx} K_N(x, x) = 2\partial_2 K_N(x, x) = -N^{3/2}h_{N-1}(\sqrt{N}x)h_N(\sqrt{N}x).$$

Differentiating once more and using (A.20) and $K_N(x, x) \leq CN$, we obtain

$$K_N(x, x) \frac{d^2}{dx^2} K_N(x, x) = O(N^3), \quad (\partial_2 K_N(x, x))^2 = O(N^2),$$

so the first two terms on the right side of the expression for $\partial_y^2 R_{2,N}(x, y)|_{y=x}$ are $O(N^3)$.

For the last term, (A.19), (A.20), Lemma A.7, and the identity $1 - x^2/4 = \pi^2 \rho_{sc}(x)^2$ give

$$\begin{aligned} -2K_N(x, x) \partial_y^2 K_N(x, y)|_{y=x} &= \frac{2\pi^2}{3} N^2 \rho_{sc}(x)^2 K_N(x, x)^2 + O(N^3) \\ &= \frac{2\pi^2}{3} N^4 \rho_{sc}(x)^4 + O(N^{31/8}). \end{aligned}$$

Further, Lemma A.5(a) gives

$$|\partial_y^3 R_{2,N}(x, y)| \leq CN^5, \quad x, y \in \tilde{I}.$$

Taylor's theorem in y at $y = x$ therefore gives

$$\left| R_{2,N}(x, x+h) - \frac{\pi^2}{3} N^4 \rho_{sc}(x)^4 h^2 \right| \leq C(N^{31/8} h^2 + N^5 h^3), \quad x, x+h \in \tilde{I}, \quad 0 < h \leq N^{-1}. \quad (9.11)$$

This is the source of the function $c_2(E)$. For larger separations, we use the bounds

$$R_{2,N}(x, y) \leq K_N(x, x) K_N(y, y) \leq CN^2, \quad |K_N(x, y)| \leq C \min\{N, |x - y|^{-1}\}, \quad (9.12)$$

which follow from (9.10) and Lemma A.5(a) and (b).

Correlation functions of nearby points. We claim that for $n \in \{2, 3, 4\}$ and $x_1, \dots, x_n \in \tilde{I}$,

$$R_{n,N}(x_1, \dots, x_n) \leq CN^n \min \left\{ 1, \prod_{1 \leq i < j \leq n} (N|x_i - x_j|)^2 \right\}. \quad (9.13)$$

The matrix $[K_N(x_i, x_j)]_{i,j=1}^n$ is the Gram matrix of the vectors $\mathbf{h}(x_1), \dots, \mathbf{h}(x_n)$. Hadamard's inequality bounds its determinant by its diagonal entries, which gives

$$R_{n,N}(x_1, \dots, x_n) \leq \prod_{i=1}^n K_N(x_i, x_i) \leq CN^n.$$

To prove the second bound in (9.13), we may assume that the points are distinct. We replace $\mathbf{h}(x_1), \dots, \mathbf{h}(x_n)$ by the divided differences $\mathbf{h}[x_1], \mathbf{h}[x_1, x_2], \dots, \mathbf{h}[x_1, \dots, x_n]$. This is a triangular change of basis with determinant $\prod_{i < j} (x_j - x_i)^{-1}$, so

$$R_{n,N}(x_1, \dots, x_n) = \prod_{1 \leq i < j \leq n} (x_j - x_i)^2 \det [\mathbf{h}[x_1, \dots, x_i] \cdot \mathbf{h}[x_1, \dots, x_j]]_{i,j=1}^n.$$

We now bound the divided differences. By the Hermite–Genocchi formula, $\mathbf{h}[x_1, \dots, x_k]$ is the integral of $\mathbf{h}^{(k-1)}$ over a simplex of volume $1/(k-1)!$ mapped into the

convex hull of $x_1, \dots, x_k$. Hence the Cauchy–Schwarz inequality and the bound (A.21) in the proof of Lemma A.5, with $k - 1$ in place of k , give

$$\begin{aligned} |\mathbf{h}[x_1, \dots, x_k]|^2 &\leq \sup_{y \in \tilde{I}} |\mathbf{h}^{(k-1)}(y)|^2 \\ &= \sup_{y \in \tilde{I}} N^{k-1/2} \sum_{j=0}^{N-1} h_j^{(k-1)} (\sqrt{N}y)^2 \leq CN^{2k-1}, \quad k \leq 4. \end{aligned}$$

Applying Hadamard's inequality to the Gram determinant of the divided differences, we obtain

$$\begin{aligned} R_{n,N}(x_1, \dots, x_n) &\leq \prod_{i < j} (x_j - x_i)^2 \prod_{k=1}^n |\mathbf{h}[x_1, \dots, x_k]|^2 \\ &\leq C \prod_{i < j} (x_j - x_i)^2 \prod_{k=1}^n N^{2k-1} = CN^{n^2} \prod_{i < j} |x_i - x_j|^2, \end{aligned}$$

which is the second bound in (9.13), since $n + n(n-1) = n^2$.

Proof of (a). We first compute the mean. By (A.12), we have

$$\mathbb{E} \sum_{1 \leq i < j \leq N} f\left(\frac{\lambda_j(G) - \lambda_i(G)}{\ell_\alpha}, \frac{\lambda_i(G) + \lambda_j(G)}{2}\right) = \int_{\mathbb{R}} \int_0^\infty f\left(\frac{h}{\ell_\alpha}, x + \frac{h}{2}\right) R_{2,N}(x, x+h) dh dx.$$

Let $\delta > 0$ be small. Since $|f(r, E)| \leq C_f r^{-2p}$ and the integral of $R_{2,N}$ over $\mathbb{R}^2$ is $N(N-1)$, the pairs at distance larger than δ contribute at most

$$C_f \left(\frac{\ell_\alpha}{\delta}\right)^{2p} \int_{\mathbb{R}} \int_0^\infty R_{2,N}(x, x+h) dh dx \leq CN^2 \left(\frac{\ell_\alpha}{\delta}\right)^{2p} = o(1).$$

For the remaining pairs, both points lie in a bulk interval $\tilde{I}$, since $f(r, E) = 0$ for $E \notin I$.

We insert (9.11) for $h \leq N^{-1}$ and change variables $h = \ell_\alpha r$. We will use the identities

$$N^4 \ell_\alpha^3 = M_\alpha^{\mathbb{C}}, \quad N^{31/8} \ell_\alpha^3 = N^{-1/8} M_\alpha^{\mathbb{C}}, \quad N^5 \ell_\alpha^4 = N^{-\alpha} M_\alpha^{\mathbb{C}}.$$

By the first identity, the main term is

$$M_\alpha^{\mathbb{C}} \int_{\mathbb{R}} \int_0^{N^\alpha} f\left(r, x + \frac{\ell_\alpha r}{2}\right) \frac{\pi^2}{3} \rho_{\text{sc}}(x)^4 r^2 dr dx,$$

and since $p > 3/2$, it converges to $M_\alpha^{\mathbb{C}}$ times the right side of (9.8) by dominated convergence. Using the other two identities, we find that the error terms in (9.11) contribute at most

$$CM_\alpha^{\mathbb{C}} \left(N^{-1/8} + N^{-\alpha} \int_0^{N^\alpha} \frac{r^3 dr}{(1+r^2)^p} \right) \leq CM_\alpha^{\mathbb{C}} (N^{-1/8} + N^{-\alpha} + N^{-\alpha(2p-3)} \log N).$$

Finally, by (9.12), the pairs with $N^{-1} < h \leq \delta$ contribute at most

$$CN^2 \int_{N^{-1}}^{\delta} \left(\frac{\ell_{\alpha}}{h} \right)^{2p} dh \leq CN^2 \ell_{\alpha} \int_{N^{\alpha}}^{\infty} r^{-2p} dr = CM_{\alpha}^{\mathbb{C}} N^{-\alpha(2p-3)}.$$

Hence the mean of the sum truncated to $|\lambda_j(G) - \lambda_i(G)| \leq \delta$, divided by $M_{\alpha}^{\mathbb{C}}$, converges to the right side of (9.8).

We now bound the variance of the truncated sum. We set

$$f^{\sharp}(x, y) = f\left(\frac{|y-x|}{\ell_{\alpha}}, \frac{x+y}{2}\right) \mathbf{1}_{|y-x| \leq \delta}, \quad x, y \in \mathbb{R}.$$

As in (A.35), with $f^{\sharp}$ in place of $g^{\sharp}$, the variance is the sum of three integrals, against $R_{4,N} - R_{2,N} \otimes R_{2,N}$, against $R_{3,N}$, and against $R_{2,N}$. In the expansion of the determinant $R_{4,N}(x_1, x_2, x_3, x_4)$, the terms that do not appear in $R_{2,N}(x_1, x_2)R_{2,N}(x_3, x_4)$ come from permutations that do not preserve $\{1, 2\}$, so they contain at least two factors $K_N(x_i, x_j)$ connecting $\{1, 2\}$ and $\{3, 4\}$, in either direction. Hence, by the symmetry of K_N and (9.12), they are bounded by $C \min\{N^4, N^2 \text{dist}(\{x_1, x_2\}, \{x_3, x_4\})^{-2}\}$.

To integrate this bound, we write $x_2 = x_1 + h_1$ and $x_4 = x_3 + h_2$ and integrate first over x_3 . Where $|x_3 - x_1| \geq 2(|h_1| + |h_2|)$, the distance between $\{x_1, x_2\}$ and $\{x_3, x_4\}$ is at least $|x_3 - x_1|/2$, and the remaining values of x_3 form a set of measure at most $4(|h_1| + |h_2|)$. Hence the integral of the last bound over $x_3 \in \mathbb{R}$ is at most

$$C \int_{\mathbb{R}} \min\left\{N^4, \frac{4N^2}{|x_3 - x_1|^2}\right\} dx_3 + 4CN^4(|h_1| + |h_2|) \leq CN^3 + CN^4(|h_1| + |h_2|).$$

We now integrate over h_1 and h_2 . Since $p > 3/2$, we have

$$\int_{\mathbb{R}} |f^{\sharp}(x_1, x_1 + h)| (1 + N|h|) dh \leq C\ell_{\alpha}(1 + N\ell_{\alpha}) \leq C\ell_{\alpha},$$

and we find that the first integral is bounded by

$$C\ell_{\alpha}^2(N^3 + N^4\ell_{\alpha}) \leq CN^3\ell_{\alpha}^2 = CN^{1-2\alpha}.$$

By Fischer's inequality, we have

$$R_{3,N}(x_1, x_2, x_3) \leq K_N(x_3, x_3)R_{2,N}(x_1, x_2) \leq CNR_{2,N}(x_1, x_2),$$

so the second integral is bounded by $CN\ell_{\alpha}M_{\alpha}^{\mathbb{C}}$, and the third by $CM_{\alpha}^{\mathbb{C}}$, by the computation of the mean applied to $|f|$. Collecting the three bounds, we find that the variance divided by $(M_{\alpha}^{\mathbb{C}})^2$ is at most

$$C \frac{N^{1-2\alpha} + N\ell_{\alpha}M_{\alpha}^{\mathbb{C}} + M_{\alpha}^{\mathbb{C}}}{(M_{\alpha}^{\mathbb{C}})^2} = C(N^{4\alpha-1} + N^{2\alpha-1} + N^{3\alpha-1}),$$

which tends to zero since $\alpha < 1/4$. Hence the normalized sum converges in L^2 after truncation, and hence in L^1 .

Proof of (b). Let $\delta > 0$ be small. We first discard two kinds of tuples. The tuples containing two points at distance larger than δ , and the tuples whose mean is at distance at least δ from $\text{supp } \chi$, contribute at most

$$\|\chi\|_\infty N^4 \left(\frac{q}{\delta}\right)^{2p} = o(1) \quad \text{and} \quad CN^4 \exp\left(-\frac{\delta^2 N^{1/5}}{2}\right) = o(1),$$

respectively. For the first bound, we used that the rational factor of the two distant points is at most $(q/\delta)^{2p}$, and for the second, part (4) of Lemma 5.1. For the remaining tuples, all points lie in a bulk interval $\tilde{I}$.

We now bound the remaining tuples. Using (A.12), the second bound in (9.13), and the substitution $x_i = x_1 + qa_i$ for $i \geq 2$, we find

$$\begin{aligned} \mathbb{E}U_q^{(3)}(G) &\leq CN^9 q^8 \int_{\mathbb{R}^2} (a_2 a_3 (a_3 - a_2))^2 f_p(a_2) f_p(a_3) f_p(a_3 - a_2) da_2 da_3 + o(1) \\ &\leq CN(Nq)^8 + o(1), \end{aligned}$$

where the integral is finite since $p \geq 3$ and

$$(a_2 a_3 (a_3 - a_2))^2 \leq 2(a_2^4 a_3^2 + a_2^2 a_3^4).$$

Similarly, with $a_1 = 0$, we find

$$\mathbb{E}U_q^{(4)}(G) \leq CN^{16} q^{15} \int_{\mathbb{R}^3} \prod_{1 \leq i < j \leq 4} (a_i - a_j)^2 f_p(a_i - a_j) da_2 da_3 da_4 + o(1).$$

To bound the integral, we bound the three factors with $2 \leq i < j \leq 4$ by one, using $y^2 f_p(y) \leq 1$ for $y \in \mathbb{R}$. The other three factors are $a_j^2 f_p(a_j)$ for $j \in \{2, 3, 4\}$, so the integral is at most $(\int_{\mathbb{R}} a^2 f_p(a) da)^3$, which is finite since $2p - 2 > 1$. As $N^{16} q^{15} = N(Nq)^{15}$, this gives the bound for $U_q^{(4)}$. $\square$

9.4 Triples of nearby eigenvalues

Proposition 9.2. *Let I be a bulk interval and $\kappa \in (0, 1/8)$, and set $q = N^{-1-\kappa}$. Then for every $\varepsilon > 0$,*

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(\#\{i \in [N-2] : \lambda_i \in I, \lambda_{i+2} - \lambda_i \leq 4q\} > N^{1-8\kappa+\varepsilon}\right) = 0. \quad (9.14)$$

If moreover $\kappa > 1/15$, then with probability tending to one there exists no $i \in [N-3]$ such that $\lambda_i \in I$ and $\lambda_{i+3} - \lambda_i \leq 6q$.

Proof. We first reduce both claims to bounds on the statistics $U_q^{(3)}$ and $U_q^{(4)}$ of (9.2). Let I_+ be a bulk interval whose interior contains I . Fix a nonnegative $\chi \in C_c^\infty((-2, 2))$ with $\chi \geq 2$ on I_+ , so that $\chi_N \geq 1$ on I_+ for $N \geq N_0$ by part (5) of Lemma 5.1, and fix an integer $p \geq 4$.

Let i be a label with $\lambda_i \in I$ and $\lambda_{i+2} - \lambda_i \leq 4q$. Then the mean of λ_i , λ_{i+1} , and λ_{i+2} belongs to I_+ for $N \geq N_0$, so the term of $(i, i+1, i+2)$ in $U_q^{(3)}(H)$ is at least 17^{-3p} by the observation after (9.2). Distinct labels i give distinct triples, and all terms of $U_q^{(3)}(H)$ are nonnegative, so

$$\#\{i \in [N-2] : \lambda_i \in I, \lambda_{i+2} - \lambda_i \leq 4q\} \leq 17^{3p} U_q^{(3)}(H).$$

Similarly, a label i with $\lambda_i \in I$ and $\lambda_{i+3} - \lambda_i \leq 6q$ gives $U_q^{(4)}(H) \geq 37^{-6p}$.

It therefore suffices to show that

$$\frac{U_q^{(3)}(H)}{N^{1-8\kappa+\varepsilon}} \rightarrow 0, \quad U_q^{(4)}(H) \rightarrow 0 \quad (\kappa > 1/15), \quad (9.15)$$

in probability. We prove (9.15) for H_T , at a time T chosen below, by comparison with the GUE, and we then remove the Gaussian component with (9.7).

We now choose the time. Let $\varepsilon' \in (0, \varepsilon/10)$ be small, and set

$$T = N^{-1+\kappa+4\varepsilon'}, \quad \frac{2N^{\varepsilon'}}{N^2 T} = 2N^{-3\varepsilon'} q, \quad TN^{1/2+3\kappa+12\varepsilon'} = N^{-1/2+4\kappa+16\varepsilon'} \rightarrow 0. \quad (9.16)$$

Then $T \geq N^{-1+\varepsilon'}$, and the limit holds if ε' is small enough, since $\kappa < 1/8$. The second quantity in (9.16) is the coupling error at time T , and the third is the comparison error in (9.7).

Next, we compare H_T with the GUE. Let $\delta \in (0, 1)$ satisfy $\text{supp } \chi \subset [-2 + 3\delta, 2 - 3\delta]$, and let c_* and $\mathcal{J}$ be as in Lemma 6.4, defined with this δ . By the coupling of Proposition 3.5 at time T , in the form of Section 9.3 and applied with ε' in place of ε , there is a GUE matrix G' such that, with overwhelming probability,

$$\max_{c_* N \leq i \leq (1-c_*)N} |(\lambda_{i+1}(H_T) - \lambda_i(H_T)) - (\lambda_{i+1}(G') - \lambda_i(G'))| \leq 2N^{-3\varepsilon'} q.$$

We intersect this event with the events of Proposition 3.2, for H_T and G' and with ε' in place of ε . On the resulting event, whose probability tends to one, we claim that

$$U_q^{(r)}(H_T) \leq (1 + o(1))U_q^{(r)}(G') + o(1), \quad r \in \{3, 4\}. \quad (9.17)$$

To prove (9.17), we follow the proof of Lemma 6.4, applied to each of the three, respectively six, rational factors of the summands and to the weight χ_N at the mean. We write $x_i = \lambda_i(H_T)$ and $x'_i = \lambda_i(G')$, and we first consider the tuples whose labels all

belong to $\mathcal{J}$. For labels $i < j$ in $\mathcal{J}$, there are at most $N^{\varepsilon'}(1 + N(x_j - x_i))$ gaps between x_i and x_j , by condition (iii) of Definition 3.1. As in (6.10), it follows that

$$|(x'_j - x'_i) - (x_j - x_i)| \leq N^{\varepsilon'}(1 + N(x_j - x_i)) \cdot 2N^{-3\varepsilon'}q \leq 2N^{-2\varepsilon'}(q + x_j - x_i),$$

where we used $Nq \leq 1$.

We now compare the terms. Let $y = (x_j - x_i)/q$ and $y' = (x'_j - x'_i)/q$. The last display gives $|y - y'| \leq 2N^{-2\varepsilon'}(1 + y)$, so, as in (6.12),

$$|\log f_p(y) - \log f_p(y')| \leq \frac{8p|y - y'|}{1 + y} \leq 16pN^{-2\varepsilon'}$$

for $N \geq N_0$.

We next compare the weights. By Proposition 3.2, $|x_i - \gamma_i|$ and $|x'_i - \gamma_i|$ are at most $N^{-1+\varepsilon'}$ for $i \in \mathcal{J}$, so the means $\bar{x}$ and $\bar{x}'$ of a tuple for x and for x' satisfy $|\bar{x} - \bar{x}'| \leq 2N^{-1+\varepsilon'}$. Recall that the smoothing (5.2) has width $h_N = N^{-1/10}$. By (5.4) and $h_N^{-2} = N^{1/5}$, we have

$$e^{-16N^{-4/5+\varepsilon'}}\chi_N(\bar{x}') \leq \chi_N(\bar{x}) \leq e^{16N^{-4/5+\varepsilon'}}\chi_N(\bar{x}').$$

A summand has at most six rational factors, so the term of every such tuple for x is at most $e^{96pN^{-2\varepsilon'}+16N^{-4/5+\varepsilon'}}$ times the corresponding term for x' . Hence the sum of the terms of these tuples for H_T is at most $(1 + o(1))U_q^{(r)}(G')$.

We now consider the tuples with a label outside $\mathcal{J}$, for x and x' alike. Suppose first that such a tuple has its mean within distance δ of $\text{supp } \chi$, so that the mean lies in $[-2 + 2\delta, 2 - 2\delta]$. If its smallest label is smaller than c_*N , then (6.8) shows that its smallest point is at most $-2 + \delta$, and hence at least δ below the mean and below its largest point; the case where its largest label exceeds $(1 - c_*)N$ is symmetric. In both cases, two of its points are at distance at least δ , so its term is bounded by $\|\chi\|_\infty(q/\delta)^{2p}$.

If instead the mean of such a tuple is at distance more than δ from $\text{supp } \chi$, then its term is at most $C \exp(-\delta^2 N^{1/5}/2)$ by part (4) of Lemma 5.1. There are at most N^4 tuples, so these tuples contribute at most

$$N^4 \left(\|\chi\|_\infty \left(\frac{q}{\delta} \right)^{2p} + C \exp \left(-\frac{\delta^2 N^{1/5}}{2} \right) \right) = o(1),$$

since $p \geq 4$. This proves (9.17).

We now use the estimates for the GUE. By (9.9) and Markov's inequality, we have

$$\frac{U_q^{(3)}(G')}{N^{1-8\kappa+\varepsilon}} \rightarrow 0, \quad U_q^{(4)}(G') \rightarrow 0 \quad (\kappa > 1/15),$$

in probability, where for the second limit we used $N(Nq)^{15} = N^{1-15\kappa}$. Combining this with (9.17), we obtain (9.15) with H_T in place of H . Since these random variables are nonnegative, the bounded convergence theorem gives, for every $\theta > 0$,

$$\lim_{N \rightarrow \infty} \mathbb{E} e^{-\theta U_q^{(3)}(H_T)/N^{1-8\kappa+\varepsilon}} = 1, \quad \lim_{N \rightarrow \infty} \mathbb{E} e^{-\theta U_q^{(4)}(H_T)} = 1 \quad (\kappa > 1/15). \quad (9.18)$$

Finally, we remove the Gaussian component. By (9.7) with $t = T$ and ε' in place of ε , and by the last relation in (9.16), the same limits hold with H in place of H_T . Since $U_q^{(r)}(H) \geq 0$, these limits imply (9.15). This completes the proof. $\square$

9.5 Gaussian perturbations of 3×3 matrices

The following lemma controls the gaps produced by a cluster of three eigenvalues under a Gaussian perturbation. We state it for a general 3×3 diagonal matrix. Let $\sigma > 0$, and let W be a 3×3 complex Hermitian random matrix whose density with respect to Lebesgue measure on the space of 3×3 Hermitian matrices B is proportional to $\exp(-\text{Tr } B^2 / (2\sigma^2))$. Equivalently, the entries W_{ab} with $a \leq b$ are independent, the diagonal entries are real centered Gaussian variables with variance σ^2 , and the entries above the diagonal are complex Gaussian variables with $\mathbb{E}W_{ab} = \mathbb{E}W_{ab}^2 = 0$ and $\mathbb{E}|W_{ab}|^2 = \sigma^2$. For a 3×3 Hermitian matrix B , we let

$$\text{gap}(B) = \min\{\lambda_2(B) - \lambda_1(B), \lambda_3(B) - \lambda_2(B)\}$$

denote its smallest eigenvalue gap.

Lemma 9.3. *There exists a constant $C > 0$ such that the following holds. Let $y_1 \leq y_2 \leq y_3$ be real numbers, let $D_y = \text{diag}(y_1, y_2, y_3)$, and let W be as above. Then for all $\delta \in (0, \sigma]$ and $R \geq 1$,*

$$\mathbb{P}(\text{gap}(D_y + W) \leq \delta) \leq CR^4 \left(\frac{\delta}{\sigma}\right)^3 + Ce^{-R^2/C}. \quad (9.19)$$

Proof. Replacing D_y , W , and δ by D_y/σ , W/σ , and δ/σ , we may assume that $\sigma = 1$, so that $\delta \leq 1 \leq R$. The Gaussian tail bound for the entries of W gives

$$\mathbb{P}(\|W\| > R) \leq Ce^{-R^2/C}. \quad (9.20)$$

We distinguish three cases according to the configuration of y_1, y_2, y_3 . Every configuration belongs to at least one of them.

Case 1: $y_2 - y_1 \geq 10R$ and $y_3 - y_2 \geq 10R$. On the event $\{\|W\| \leq R\}$, Weyl's inequality shows that each eigenvalue of $D_y + W$ differs from the corresponding y_j by at most R , so

$$\lambda_{j+1}(D_y + W) - \lambda_j(D_y + W) \geq y_{j+1} - y_j - 2R \geq 8R > \delta, \quad j \in \{1, 2\}.$$

Hence the probability in (9.19) is bounded by $Ce^{-R^2/C}$, by (9.20).

Case 2: $y_3 - y_1 \leq 30R$. We use the Weyl integration formula (see [2, Section 4.1]). The density of $D_y + W$ with respect to Lebesgue measure on Hermitian matrices B is proportional to

$$\exp\left(-\frac{\|B - D_y\|_F^2}{2}\right),$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Writing $B = O \operatorname{diag}(\mu) O^*$ with $\mu_1 \leq \mu_2 \leq \mu_3$ and O unitary, the Weyl integration formula shows that the ordered eigenvalues of $D_y + W$ have the density

$$\mu \mapsto c \prod_{1 \leq i < j \leq 3} (\mu_j - \mu_i)^2 \int_{U(3)} \exp\left(-\frac{\|O \operatorname{diag}(\mu) O^* - D_y\|_F^2}{2}\right) dO \quad (9.21)$$

on $\{\mu_1 \leq \mu_2 \leq \mu_3\}$, where dO is the Haar probability measure on the unitary group $U(3)$ and $c > 0$ is an absolute constant.

By the Hoffman–Wielandt inequality for Hermitian matrices [40], we have

$$\|O \operatorname{diag}(\mu) O^* - D_y\|_F \geq |\mu - y|, \quad y = (y_1, y_2, y_3).$$

Hence the density (9.21) at μ is bounded by

$$c \prod_{1 \leq i < j \leq 3} (\mu_j - \mu_i)^2 e^{-|\mu - y|^2/2}. \quad (9.22)$$

We now bound the probability that $\mu_2 - \mu_1 \leq \delta$. On this event, we have

$$\begin{aligned} \prod_{1 \leq i < j \leq 3} (\mu_j - \mu_i)^2 &\leq \delta^2 (\mu_3 - \mu_1)^4, \quad \mu_3 - \mu_1 \leq |\mu_1 - y_1| + |\mu_3 - y_3| + 30R, \\ e^{-|\mu - y|^2/2} &\leq e^{-((\mu_1 - y_1)^2 + (\mu_3 - y_3)^2)/2}, \end{aligned}$$

where the second bound uses $y_3 - y_1 \leq 30R$. Integrating over $\mu_2 \in [\mu_1, \mu_1 + \delta]$, and then over μ_1 and μ_3 , we find

$$\begin{aligned} \mathbb{P}(\mu_2 - \mu_1 \leq \delta) &\leq c \delta^3 \int_{\mathbb{R}^2} (|\mu_1 - y_1| + |\mu_3 - y_3| + 30R)^4 e^{-((\mu_1 - y_1)^2 + (\mu_3 - y_3)^2)/2} d\mu_1 d\mu_3 \\ &\leq CR^4 \delta^3. \end{aligned} \quad (9.23)$$

The same bound holds for $\mathbb{P}(\mu_3 - \mu_2 \leq \delta)$, so the probability in (9.19) is bounded by $CR^4 \delta^3$ in this case.

Case 3: neither Case 1 nor Case 2 holds. Then $y_3 - y_1 > 30R$, and one of the gaps $y_2 - y_1$ and $y_3 - y_2$ is smaller than $10R$; the other is then larger than $20R$. Let P denote the permutation matrix that reverses the order of the coordinates. The matrix

$$-P(D_y + W)P^T = \operatorname{diag}(-y_3, -y_2, -y_1) - PWP^T$$

has the gaps of $D_y + W$ in reverse order, and $-PWP^T$ has the law of W . Replacing (y_1, y_2, y_3) by $(-y_3, -y_2, -y_1)$ if necessary, we may therefore assume that $y_2 - y_1 < 10R$ and $y_3 - y_2 > 20R$.

We work on the event $\|W\| \leq R$, whose complement is controlled by (9.20), and we set $\mu_j = \lambda_j(D_y + W)$. By Weyl's inequality, we have

$$\begin{aligned} \mu_3 &\geq y_3 - R \geq y_1 + 19R, & \mu_1, \mu_2 &\in [y_1 - R, y_2 + R] \subset [y_1 - R, y_1 + 11R], \\ \mu_3 - \mu_2 &\geq 19R - 11R = 8R > \delta. \end{aligned}$$

Thus the only possible small gap is $\mu_2 - \mu_1$, and μ_1 and μ_2 are the eigenvalues of $D_y + W$ in the interval

$$I_R = [y_1 - 12R, y_1 + 13R].$$

We reduce to a 2×2 problem by a Schur complement. Let $W_{[2]}$ denote the upper left 2×2 block of W , let $b = (W_{13}, W_{23})^T$, and for real w with $y_3 + W_{33} - w \neq 0$, let

$$A(w) = \text{diag}(y_1, y_2) + W_{[2]} - \frac{bb^*}{y_3 + W_{33} - w}. \quad (9.24)$$

Let $w \in I_R$, and suppose that $|W_{33}| \leq R$ and $|b| \leq R$, which hold if $\|W\| \leq R$. Since $y_3 - y_1 > 20R$, we have

$$\begin{aligned} y_3 + W_{33} - w &\geq 20R - R - 13R = 6R, \\ \left\| \frac{bb^*}{y_3 + W_{33} - w} \right\| &\leq \frac{R}{6}, \quad \left\| \frac{d}{dw} \frac{bb^*}{y_3 + W_{33} - w} \right\| \leq \frac{R^2}{(6R)^2} = \frac{1}{36}, \end{aligned}$$

and the last term in (9.24) is negative semidefinite.

We now locate the eigenvalues of $D_y + W$ in I_R . By Weyl's inequality, the eigenvalues $\lambda_1(A(w)) \leq \lambda_2(A(w))$ of the Hermitian matrix $A(w)$ are Lipschitz functions of $w \in I_R$ with constant $1/36$, so the functions $w \mapsto \lambda_k(A(w)) - w$ are strictly decreasing on I_R . Since $\|W_{[2]}\| \leq R$ and $y_2 < y_1 + 10R$, the bounds above give

$$\lambda_k(A(w)) \in \left[y_1 - \frac{7R}{6}, y_2 + R \right] \subset \left[y_1 - \frac{7R}{6}, y_1 + 11R \right], \quad w \in I_R,$$

so the functions $w \mapsto \lambda_k(A(w)) - w$ are positive at $w = y_1 - 12R$ and negative at $w = y_1 + 12R$. Hence each of them has exactly one zero w_k in I_R , and $w_1 \leq w_2$ since $\lambda_1(A(w)) \leq \lambda_2(A(w))$.

We now identify μ_1 and μ_2 . By the Schur complement formula, a point $w \in I_R$ is an eigenvalue of $D_y + W$ if and only if $\det(A(w) - w) = 0$, which holds exactly when $w \in \{w_1, w_2\}$. Moreover, the multiplicity of such an eigenvalue is the dimension of the kernel of $A(w) - w$. When $w_1 < w_2$, the two eigenvalues of $D_y + W$ in I_R are therefore w_1 and w_2 . When $w_1 = w_2$, we have $A(w_1) = w_1 \text{Id}_2$, so w_1 is an eigenvalue of multiplicity two. In both cases, $\mu_k = w_k$ for $k \in \{1, 2\}$.

We now transfer a small gap of $D_y + W$ to $A(w)$. Since $\lambda_k(A(\mu_k)) = \mu_k$ and $w \mapsto \lambda_2(A(w))$ is Lipschitz with constant $1/36$, if $\mu_2 - \mu_1 \leq \delta$, then

$$\lambda_2(A(\mu_1)) - \lambda_1(A(\mu_1)) = \lambda_2(A(\mu_1)) - \mu_1 \leq \lambda_2(A(\mu_2)) + \frac{\mu_2 - \mu_1}{36} - \mu_1 \leq 2\delta. \quad (9.25)$$

We now discretize. We set

$$w'_j = y_1 - 12R + j\delta, \quad j \in \mathbb{Z}, \quad 0 \leq j \leq \frac{24R}{\delta} + 1,$$

and we note that these points lie in I_R , since $\delta \leq R$. If $\mu_2 - \mu_1 \leq \delta$, we choose j with $|w'_j - \mu_1| \leq \delta$. Then the Lipschitz bound on A and (9.25), together with Weyl's inequality, give

$$\begin{aligned} \|A(w'_j) - A(\mu_1)\| &\leq \frac{\delta}{36}, \\ \lambda_2(A(w'_j)) - \lambda_1(A(w'_j)) &\leq 2\delta + \frac{\delta}{18} \leq 3\delta, \\ |\lambda_1(A(w'_j)) - w'_j| &\leq |\lambda_1(A(w'_j)) - \lambda_1(A(\mu_1))| + |\mu_1 - w'_j| \leq \frac{\delta}{36} + \delta \leq 2\delta. \end{aligned}$$

Thus the matrix $A(w'_j)$ has eigenvalue gap at most 3δ and an eigenvalue within distance 2δ of w'_j .

We bound the probability of this event for fixed j . Condition on W_{33} and b , and restrict to the event $\{|W_{33}| \leq R, |b| \leq R\}$, on which $A(w'_j) = M_j + W_{[2]}$ with a deterministic Hermitian matrix M_j . The block $W_{[2]}$ is independent of (W_{33}, b) . We set

$$\begin{aligned} v &= \left(\frac{(M_j)_{11} - (M_j)_{22}}{2}, \operatorname{Re}(M_j)_{12}, \operatorname{Im}(M_j)_{12} \right), \\ \xi &= \left(\frac{W_{11} - W_{22}}{2}, \operatorname{Re} W_{12}, \operatorname{Im} W_{12} \right). \end{aligned}$$

Then the eigenvalues of the Hermitian 2×2 matrix $M_j + W_{[2]}$ are $(\operatorname{Tr} M_j + W_{11} + W_{22})/2 \pm |v + \xi|$.

The four real variables $(W_{11} + W_{22})/2$, $(W_{11} - W_{22})/2$, $\operatorname{Re} W_{12}$, and $\operatorname{Im} W_{12}$ are independent centered Gaussian variables with variance $1/2$. The event above requires

$$|v + \xi| \leq \frac{3\delta}{2}, \quad \left| \frac{\operatorname{Tr} M_j + W_{11} + W_{22}}{2} - w'_j \right| \leq 4\delta.$$

The densities of ξ and of $(W_{11} + W_{22})/2$ are bounded, so this event has conditional probability at most $C\delta^3 \cdot C\delta$. Summing over the at most $24R/\delta + 2$ values of j , and using $\delta \leq 1 \leq R$, we conclude that in Case 3,

$$\mathbb{P}(\operatorname{gap}(D_y + W) \leq \delta) \leq C \left(\frac{R}{\delta} + 1 \right) \delta^4 + C e^{-R^2/C} \leq CR\delta^3 + C e^{-R^2/C}. \quad (9.26)$$

Combining the three cases proves (9.19). $\square$

9.6 Gap counts at the scales $N^{-1-\alpha}$

For $\alpha \in (0, 1/3)$, we say that the statement $\mathcal{P}^C(\alpha)$ holds if for every nonnegative $\varphi \in C_c^\infty((-2, 2))$ and every integer $p \geq p_*$, we have

$$\frac{1}{M_\alpha^C} \mathcal{X}_\alpha(H; \varphi_N, f_p) \rightarrow \int_{-2}^2 \varphi(E) c_2(E) dE \int_0^\infty y^2 f_p(y) dy \quad (9.27)$$

in probability, where $\mathcal{X}_\alpha$ is defined by (5.1) for Hermitian matrices. This is the analogue of the statement $\mathcal{P}(\alpha)$ of Definition 6.1.

We first record the results of Section 6 that carry over. The proofs of Lemma 6.2, Lemma 6.3, and Lemma 6.6 apply with M_α^C , $c_2(E)$, and $\int_0^\infty y^2 f(y) dy$ in place of M_α , $c(E)$, and $\int_0^\infty y f(y) dy$, and with (9.4) in place of Proposition 5.9; the integrals $\int_0^\infty y^2 f_p(y) dy$ are finite for $p \geq p_*$. Indeed, the proof of Lemma 6.2 is deterministic and uses only the bound $\|A\| \leq 3$ and Lemma 5.1. The proof of Lemma 6.3 uses the comparison estimate only through the bound (5.47) and the nonnegativity of the pair statistic, which holds for Hermitian matrices as well. The proof of Lemma 6.6 uses only $\mathcal{P}(\alpha)$, the Weierstrass approximation theorem, and Lemma 6.2, with (6.1) replaced by the finiteness of the integrals above.

Base case. The proof of Proposition 6.5 applies with the following changes, and gives $\mathcal{P}^C(\alpha)$ for every $\alpha \in (0, 1/8)$. The GUE input is Lemma 9.1(a) in place of Proposition 3.6. The proof of Lemma 6.4 is deterministic, and in it the pairs with a label outside $\mathcal{J}$ contribute at most

$$CN^{-2\alpha} + CN^2 \exp(-\delta^2 N^{1/5}/4) = o(M_\alpha^C)$$

by (6.9), so its proof gives (6.6) with M_α^C in place of M_α . With $T = N^{-3/4-\alpha}$ as in Proposition 6.5, the coupling error of Section 9.3 and the comparison error of (9.4) satisfy

$$\frac{2N^\varepsilon}{N^2 T} = 2N^{-5/4+\alpha+\varepsilon} \leq N^{-2\varepsilon} \ell_\alpha, \quad TN^{1/2+3\alpha+12\varepsilon} = N^{-1/4+2\alpha+12\varepsilon} \rightarrow 0$$

for $\alpha < 1/8$ and ε small. These exponents are the same as in the real symmetric case, and they do not use any condition on the third moments.

Induction step. Suppose that $\mathcal{P}^C(\alpha)$ holds, and let $\alpha' \in (\alpha, 1/3)$ and $\kappa > 0$ satisfy

$$\frac{3\alpha'}{8} < \kappa < \min \left\{ \alpha, 2\alpha - \alpha', \frac{1}{8} \right\}, \quad \alpha' - \kappa \geq \frac{1}{30}, \quad \alpha' < \frac{1}{6} + \frac{2\alpha}{3}. \quad (9.28)$$

We follow Section 6.3, with $t_\alpha = N\ell_\alpha^2$, $q = N^{-1-\kappa}$, and $\varepsilon > 0$ sufficiently small depending on the margins in (9.28). The conditions (9.28) replace the choice of κ in (6.23) and the conditions (6.24), and we indicate below where each of them is used.

We begin with the Gaussian blocks. Conditionally on H , the matrix H_{t_α} has the eigenvalues of $D_y + W$, where $y_i = e^{-t_\alpha/2} \lambda_i$ and $W = (1 - e^{-t_\alpha})^{1/2} (u_i^* G u_j)_{i,j=1}^N$ for

orthonormal eigenvectors $u_1, \dots, u_N$ of H ; by the unitary invariance of the GUE, W is independent of H and has the law of $(1 - e^{-t_\alpha})^{1/2} G$. As in the proof of Lemma 6.9, Corollary 3.9 applies with $w_0 = N^\varepsilon \ell_\alpha$, and the error in (3.31) is $o(\ell_{\alpha'})$, since

$$\begin{aligned} N^\varepsilon w_0(q^{-1} + N \log N) &\leq 2N^{2\varepsilon - \alpha + \kappa} \rightarrow 0, \\ 24N^{2\varepsilon} w_0^2(q^{-1} + N \log N) &\leq CN^{4\varepsilon - 1 - 2\alpha + \kappa} = o(\ell_{\alpha'}), \end{aligned}$$

where we used $\kappa < \alpha$ in the first bound and $\kappa < 2\alpha - \alpha'$ in the last relation. For a two-element cluster $\mathcal{K} = \{i, i+1\}$, since the eigenvalues of a 2×2 Hermitian matrix differ by the Euclidean norm of the vector formed by the difference of its diagonal entries and twice the real and imaginary parts of its off-diagonal entry, the gap $\hat{g}_\mathcal{K}$ of the block $(D_y + W)_{\mathcal{K}\mathcal{K}}$ is

$$\hat{g}_\mathcal{K} = |(e^{-t_\alpha/2} g_i + W_{i+1,i+1} - W_{ii}, 2 \operatorname{Re} W_{i,i+1}, 2 \operatorname{Im} W_{i,i+1})|. \quad (9.29)$$

Next, we compute the conditional mean of the block statistic, following the proof of Lemma 6.10. The first coordinate of the vector in (9.29) minus $e^{-t_\alpha/2} g_i$ and the other two coordinates are independent centered Gaussian variables with variance $v_N^C \ell_\alpha^2$, where

$$v_N^C = \frac{2(1 - e^{-t_\alpha})}{N \ell_\alpha^2} = \frac{2(1 - e^{-t_\alpha})}{t_\alpha} \in [2 - t_\alpha, 2]$$

(compare (6.41)). The vector in (9.29) is three-dimensional because degeneracies of Hermitian matrices have codimension three. Thus $\hat{g}_\mathcal{K}/\ell_\alpha$ has the law of $|G_x|$ in Lemma 3.10 with $n = 3$, $x = g_i/\ell_\alpha$, $\varrho = e^{-t_\alpha/2}$, and $v = v_N^C$, and we apply Lemma 3.10 with $v_0 = 2$ at the scale $N^{\alpha-\alpha'} = \ell_{\alpha'}/\ell_\alpha$; its hypothesis $p > 2$ holds since $p \geq p_*$. In the notation (3.34) with $n = 3$, we have $g_2(x, 0, 0) = (4\pi)^{-3/2} e^{-x^2/4}$, and $\int_{\mathbb{R}^3} f(|y|) dy = 4\pi \int_0^\infty r^2 f(r) dr$ for radial integrands. Since $4\pi(4\pi)^{-3/2} = (2\sqrt{\pi})^{-1}$, we obtain

$$\mathbb{E} \left[f_p \left(\frac{\hat{g}_\mathcal{K}}{\ell_{\alpha'}} \right) \mid H \right] = N^{3(\alpha-\alpha')} \left[\frac{e^{-g_i^2/(4\ell_\alpha^2)}}{2\sqrt{\pi}} \int_0^\infty y^2 f_p(y) dy + o(1) f_p \left(\frac{g_i}{\ell_\alpha} \right) \right], \quad (9.30)$$

where the $o(1)$ term is deterministic and uniform in H and in $\mathcal{K}$.

Recall the block statistic $\hat{\mathcal{X}}_{\alpha'}(\varphi, p)$ of (6.33), and the set $\mathfrak{P}_2$ of two-element clusters and the quantities $g_\mathcal{K}$ and $E_\mathcal{K}$ defined before it. We sum (9.30), multiplied by $\varphi_N(E_\mathcal{K})$, over $\mathcal{K} \in \mathfrak{P}_2$, and we bound the error terms by $\mathcal{P}^C(\alpha)$. Using also $N^{3(\alpha-\alpha')} M_\alpha^C = M_{\alpha'}^C$, we find

$$\begin{aligned} \frac{1}{M_{\alpha'}^C} \mathbb{E} [\hat{\mathcal{X}}_{\alpha'}(\varphi, p) \mid H] &= \frac{\int_0^\infty y^2 f_p(y) dy}{2\sqrt{\pi} M_\alpha^C} \sum_{\mathcal{K} \in \mathfrak{P}_2} \varphi_N(E_\mathcal{K}) e^{-g_\mathcal{K}^2/(4\ell_\alpha^2)} + o_{\mathbb{P}}(1) \\ &\longrightarrow \int_{-2}^2 \varphi(E) c_2(E) dE \int_0^\infty y^2 f_p(y) dy \end{aligned}$$

in probability, which is the limit in (9.27) with α' in place of α . For the limit, we used $\mathcal{P}^C(\alpha)$ and Lemma 6.6 applied to the Gaussian test function $y \mapsto e^{-y^2/4}$, together with

$\int_0^\infty r^2 e^{-r^2/4} dr = 2\sqrt{\pi}$, and we removed the pairs that do not form a two-element cluster as explained below. The conditional variance bound is as in the proof of Lemma 6.11, since the blocks of distinct clusters are conditionally independent given H .

We now treat the pairs of eigenvalues in a common cluster with at least three elements, which appear in (6.36) and in the evaluation of the conditional mean. Here we use Proposition 9.2, which applies since $\kappa < 1/8$, in place of Proposition 3.4. Let $\delta > 0$, and let I be a bulk interval containing the closed 2δ -neighborhood of $\text{supp } \varphi$. Recall from Lemma 6.8 that every cluster has at most N^ε elements, and observe that the smallest label i of a cluster with at least three elements satisfies $\lambda_{i+2} - \lambda_i < 2e^{t_\alpha/2}q \leq 4q$, as in the proof of that lemma. Hence the number of such pairs in clusters whose smallest eigenvalue lies in I is at most $N^{2\varepsilon}$ times the number of labels i with $\lambda_i \in I$ and $\lambda_{i+2} - \lambda_i \leq 4q$. By Proposition 9.2, with probability tending to one, this is at most $N^{1-8\kappa+3\varepsilon} = o(M_{\alpha'}^{\mathbb{C}})$, since $8\kappa > 3\alpha'$; this replaces the use of Proposition 3.4 in (6.35). A cluster has diameter at most $2N^\varepsilon q$, so the pairs in clusters whose smallest eigenvalue lies outside I have midpoints at distance at least δ from $\text{supp } \varphi$, and by Lemma 5.1 they contribute at most $CN^2 \exp(-\delta^2 N^{1/5}/2) = o(1)$.

The pairs in different clusters are at distance at least $q/2$, by Corollary 3.9, so they contribute at most

$$4^p \|\varphi\|_\infty N^2 \left(\frac{\ell_{\alpha'}}{q} \right)^{2p} = 4^p \|\varphi\|_\infty M_{\alpha'}^{\mathbb{C}} N^{1+3\alpha'-2p(\alpha'-\kappa)} \leq 4^p \|\varphi\|_\infty M_{\alpha'}^{\mathbb{C}} N^{2-61/15},$$

where we used $p \geq p_* = 61$, $\alpha' - \kappa \geq 1/30$, and $1 + 3\alpha' < 2$. In the evaluation of the conditional mean, the pairs of eigenvalues of H in different clusters are at distance at least q , so, as in the proof of Lemma 6.10, they contribute at most $\|\varphi\|_\infty N^2 \exp(-q^2/(4\ell_\alpha^2)) = o(1)$, since $\kappa < \alpha$. Here $4\ell_\alpha^2$ replaces the denominator $8\ell_\alpha^2$ of the real symmetric case, because the Gaussian factor in (9.30) is $e^{-g_i^2/(4\ell_\alpha^2)}$ instead of $e^{-g_i^2/(8\ell_\alpha^2)}$. Finally, the comparison (9.4) at the scale $\ell_{\alpha'}$ and time t_α has the error $t_\alpha N^{1/2+3\alpha'+12\varepsilon} = N^{-1/2-2\alpha+3\alpha'+12\varepsilon}$, which tends to zero by the last condition in (9.28). This completes the proof of $\mathcal{P}^{\mathbb{C}}(\alpha')$.

Iteration. The conditions (9.28) hold for the triples $(\alpha, \alpha', \kappa)$ equal to

$$\left(\frac{3}{25}, \frac{17}{100}, \frac{67}{1000} \right), \quad \left(\frac{17}{100}, \frac{6}{25}, \frac{19}{200} \right), \quad \left(\frac{6}{25}, \frac{8}{25}, \frac{49}{400} \right). \quad (9.31)$$

Indeed, for these three triples, in this order, we have

$$\begin{aligned} \left(\frac{3\alpha'}{8}, \kappa, \min \left\{ \alpha, 2\alpha - \alpha', \frac{1}{8} \right\} \right) &= \left(\frac{255}{4000}, \frac{268}{4000}, \frac{280}{4000} \right), \quad \left(\frac{18}{200}, \frac{19}{200}, \frac{20}{200} \right), \quad \left(\frac{48}{400}, \frac{49}{400}, \frac{50}{400} \right), \\ \alpha' - \kappa &= \frac{103}{1000}, \quad \frac{29}{200}, \quad \frac{79}{400}, \\ \left(\alpha', \frac{1}{6} + \frac{2\alpha}{3} \right) &= \left(\frac{51}{300}, \frac{74}{300} \right), \quad \left(\frac{36}{150}, \frac{42}{150} \right), \quad \left(\frac{48}{150}, \frac{49}{150} \right). \end{aligned}$$

The minima in the first line are $2\alpha - \alpha'$, $2\alpha - \alpha'$, and $1/8$, respectively. In each triple of the first line, the numbers increase, the differences in the second line exceed $1/30$, and in each pair of the third line the first number is smaller than the second. Starting from the base case at $\alpha = 3/25 < 1/8$, three induction steps give $\mathcal{P}^C(\alpha_1)$ with $\alpha_1 = 8/25$.

9.7 The critical scale

We fix $\alpha_1 = 8/25$ and $\kappa_1 = 49/400$, so that

$$\frac{1}{4} < \alpha_1 < \frac{1}{3}, \quad \frac{3\alpha_1}{8} < \kappa_1 < \frac{1}{8}, \quad \frac{1}{15} < \kappa_1 < 2\alpha_1 - \frac{1}{3}, \quad (9.32)$$

and we set $q = N^{-1-\kappa_1}$. Recall that $t_{\alpha_1} = N\ell_{\alpha_1}^2 = N^{-1-2\alpha_1}$. Note that $\kappa_1 < \alpha_1$. We fix $\varepsilon > 0$ sufficiently small depending on the margins in (9.32); for example, we need $5\varepsilon < 8\kappa_1 - 3\alpha_1 = 1/50$. Fix a bulk interval I , a bulk interval I' whose interior contains I , and a bulk interval I'' whose interior contains I' . We also fix $x_0 \geq 1$.

The matrix $H_{t_{\alpha_1}}$. As in the induction step, conditionally on H , the matrix $H_{t_{\alpha_1}}$ has the eigenvalues of $D_y + W$, where $y_i = e^{-t_{\alpha_1}/2}\lambda_i$ and W is independent of H with the law of $(1 - e^{-t_{\alpha_1}})^{1/2}G$. We form clusters at the threshold q as in Section 6.3, and we work on the event $\{H \in \mathcal{G}_{\varepsilon/2}\} \cap \{\max_{i \leq j} |W_{ij}| \leq N^\varepsilon \ell_{\alpha_1}\}$, whose probability tends to one. On this event, Corollary 3.9 applies with $w_0 = N^\varepsilon \ell_{\alpha_1}$, since $N^{2\varepsilon} \ell_{\alpha_1} (q^{-1} + N \log N) \leq CN^{2\varepsilon - \alpha_1 + \kappa_1} \rightarrow 0$. It shows that the eigenvalues of $H_{t_{\alpha_1}}$ attached to a cluster differ from those of the corresponding diagonal block of $D_y + W$ by at most

$$CN^{4\varepsilon} \ell_{\alpha_1}^2 q^{-1} = CN^{-1-2\alpha_1+\kappa_1+4\varepsilon} = o(d_2),$$

since $\kappa_1 < 2\alpha_1 - 1/3$, and that eigenvalues attached to different clusters are at distance at least $q/2$, which is much larger than d_2 .

We next use Proposition 9.2 to control the sizes of the clusters. The smallest label i of a cluster with at least three elements satisfies $\lambda_{i+2} - \lambda_i < 2e^{t_{\alpha_1}/2}q \leq 4q$, and if the cluster has at least four elements, then also $\lambda_{i+3} - \lambda_i < 3e^{t_{\alpha_1}/2}q \leq 6q$. By Proposition 9.2, applied with $\kappa = \kappa_1 \in (1/15, 1/8)$ and with I'' in place of I , with probability tending to one no cluster whose smallest eigenvalue lies in I'' has four or more elements, and at most $N^{1-8\kappa_1+\varepsilon}$ such clusters have three elements. A cluster that contains an eigenvalue in I' has its smallest eigenvalue in I'' for $N \geq N_0$, since its diameter is at most $2N^\varepsilon q$.

Two-element clusters. Let $Q(y, E) = \sum_{j=1}^n \theta_j \varphi_j(E) f_{p_j}(y)$ with $\theta_j \geq 0$, nonnegative $\varphi_j \in C_c^\infty((-2, 2))$ with support in the interior of I' , and integers $p_j \geq p_*$, and set $p_0 = \min_j p_j$. Recall that $E_i = (\lambda_i + \lambda_{i+1})/2$. For a two-element cluster $\mathcal{K} = \{i, i+1\}$, the block gap

$\hat{g}_i = \hat{g}_{\mathcal{K}}$ is given by (9.29) with α_1 in place of α . For $E \in \mathbb{R}$, the function $\psi_E(y) = 1 - e^{-Q(y,E)}$ satisfies $0 \leq \psi_E(y) \leq Q(y,E) \leq C_Q f_{p_0}(y)$. We set

$$\bar{\psi}_i = \mathbb{E}[\psi_{E_i}(\hat{g}_i/d_2) \mid H], \quad m_Q(E) = \int_0^\infty y^2 \psi_E(y) dy. \quad (9.33)$$

By Lemma 3.10 as in (9.30), now applied to ψ_{E_i}/C_Q at the scale d_2/ℓ_{α_1} , we have

$$\bar{\psi}_i = \left(\frac{d_2}{\ell_{\alpha_1}}\right)^3 \left[\frac{e^{-g_i^2/(4\ell_{\alpha_1}^2)}}{2\sqrt{\pi}} m_Q(E_i) + o(1) f_{p_0}\left(\frac{g_i}{\ell_{\alpha_1}}\right) \right],$$

uniformly in i . In particular, $\max_i \bar{\psi}_i \leq C(d_2/\ell_{\alpha_1})^3 \rightarrow 0$.

We now sum the $\bar{\psi}_i$. The function m_Q is nonnegative and smooth, with support in the interior of I' . By $\mathcal{P}^C(\alpha_1)$, Lemma 6.6 applied with the unsmoothed weight m_Q and the test function $y \mapsto e^{-y^2/4}$, and the identities $(d_2/\ell_{\alpha_1})^3 M_{\alpha_1}^C = 1$ and $\int_0^\infty r^2 e^{-r^2/4} dr = 2\sqrt{\pi}$, we find

$$\sum_i \bar{\psi}_i \longrightarrow \int_{-2}^2 \int_0^\infty c_2(E) y^2 (1 - e^{-Q(y,E)}) dy dE \quad (9.34)$$

in probability, where the sum is over the first labels of the two-element clusters. Here the pairs that do not form a two-element cluster were removed from the statistic of $\mathcal{P}^C(\alpha_1)$ as in the induction step, using $8\kappa_1 > 3\alpha_1$ for the clusters with three elements. The blocks of distinct clusters are conditionally independent given H , so the conditional expectation of $\exp(-\sum_i Q(\hat{g}_i/d_2, E_i))$ given H is $\prod_i (1 - \bar{\psi}_i)$. Since $\max_i \bar{\psi}_i \rightarrow 0$ and $\sum_i \bar{\psi}_i = O_{\mathbb{P}}(1)$, this product converges in probability to the exponential of minus the right side of (9.34).

Three-element clusters. Let $\mathcal{K} = \{i, i+1, i+2\}$ be a cluster whose smallest eigenvalue lies in I'' , and let $B_{\mathcal{K}} = (D_y + W)_{\mathcal{K}\mathcal{K}}$. This block is a deterministic diagonal matrix plus a random matrix $W_{\mathcal{K}\mathcal{K}}$ independent of H as in Lemma 9.3, with $\sigma^2 = (1 - e^{-t_{\alpha_1}})/N \in [\ell_{\alpha_1}^2/2, \ell_{\alpha_1}^2]$. By Lemma 9.3 with $R = N^\varepsilon$, for all $\delta \in (0, \sigma]$,

$$\mathbb{P}(\text{gap}(B_{\mathcal{K}}) \leq \delta \mid H) \leq CN^{4\varepsilon} \left(\frac{\delta}{\ell_{\alpha_1}}\right)^3 + Ce^{-N^{2\varepsilon}/C}.$$

We now bound the contribution of $B_{\mathcal{K}}$. Every pair of eigenvalues of $B_{\mathcal{K}}$ is at distance at least $\text{gap}(B_{\mathcal{K}})$, and f_{p_0} is decreasing on $[0, \infty)$. Let $k \geq 0$ be the smallest integer such that $\text{gap}(B_{\mathcal{K}}) \leq 2^k d_2$. Then $f_{p_0}(\text{gap}(B_{\mathcal{K}})/d_2) \leq 4^{p_0} 2^{-2p_0 k}$, and we find

$$\begin{aligned} \mathbb{E}\left[f_{p_0}\left(\frac{\text{gap}(B_{\mathcal{K}})}{d_2}\right) \mid H\right] &\leq CN^{4\varepsilon} \left(\frac{d_2}{\ell_{\alpha_1}}\right)^3 \sum_{k \geq 0} 2^{(3-2p_0)k} + C \left(\frac{d_2}{\sigma}\right)^{2p_0} + Ce^{-N^{2\varepsilon}/C} \log N \\ &\leq CN^{4\varepsilon} \left(\frac{d_2}{\ell_{\alpha_1}}\right)^3, \end{aligned}$$

where the second term bounds the contribution of the k with $2^k d_2 > \sigma$, and the third the error terms of the $O(\log N)$ values of k with $2^k d_2 \leq \sigma$. On our event, the

eigenvalues of $H_{t_{\alpha_1}}$ attached to $\mathcal{K}$ differ from those of $B_{\mathcal{K}}$ by $o(d_2)$, so the terms in $\sum_{i < i'} Q((\lambda_{i'} - \lambda_i)/d_2, (\lambda_i + \lambda_{i'})/2)$ of the pairs of these eigenvalues are at most $Cf_{p_0}(\text{gap}(B_{\mathcal{K}})/d_2)$. Summing over the at most $N^{1-8\kappa_1+\varepsilon}$ three-element clusters, we find that their total contribution, and also the conditional probability that some three-element cluster produces a gap at most $x_0 d_2$, is bounded in conditional expectation by

$$CN^{1-8\kappa_1+5\varepsilon} \left(\frac{d_2}{\ell_{\alpha_1}} \right)^3 = CN^{1-8\kappa_1+5\varepsilon} N^{-1+3\alpha_1} = CN^{3\alpha_1-8\kappa_1+5\varepsilon} \rightarrow 0, \quad (9.35)$$

since $5\varepsilon < 8\kappa_1 - 3\alpha_1$.

The remaining pairs. With probability tending to one, every cluster containing an eigenvalue in I' has at most three elements. The pairs of eigenvalues of $H_{t_{\alpha_1}}$ attached to different clusters are at distance at least $q/2$, so, since $p_0 \geq p_*$, they contribute at most

$$CN^2 \left(\frac{2d_2}{q} \right)^{2p_0} = CN^{2-2p_0(1/3-\kappa_1)} = o(1).$$

Pairs whose midpoint lies outside I' do not contribute. For a two-element cluster $\{i, i+1\}$, the gap of $H_{t_{\alpha_1}}$ differs from $\hat{g}_i$ by $o(d_2)$, so the rational factors change by a relative amount $o(1)$, and the midpoint differs from E_i by at most $CN^\varepsilon \ell_{\alpha_1}$, as in the proof of Lemma 6.9, so the weights φ_j change by at most $CN^\varepsilon \ell_{\alpha_1}$. The conditional expectation of $\sum_i f_{p_0}(\hat{g}_i/d_2)$ over the two-element clusters meeting I' is $O_{\mathbb{P}}(1)$, by Lemma 3.10 and the upper bound (6.20) in Lemma 6.6, so these changes contribute $o_{\mathbb{P}}(1)$.

Combining these facts with the two preceding paragraphs and using the bounded convergence theorem, we conclude that, for every Q as above,

$$\begin{aligned} & \mathbb{E} \exp \left(- \sum_{i < i'} Q \left(\frac{\lambda_{i'}(H_{t_{\alpha_1}}) - \lambda_i(H_{t_{\alpha_1}})}{d_2}, \frac{\lambda_i(H_{t_{\alpha_1}}) + \lambda_{i'}(H_{t_{\alpha_1}})}{2} \right) \right) \\ & \longrightarrow \exp \left(- \int_{-2}^2 \int_0^\infty c_2(E) y^2 (1 - e^{-Q(y,E)}) dy dE \right). \end{aligned} \quad (9.36)$$

The same argument shows that the statistic

$$\sum_{i < i'} \chi \left(\frac{\lambda_i + \lambda_{i'}}{2} \right) f_{p_*} \left(\frac{\lambda_{i'} - \lambda_i}{d_2} \right), \quad \lambda_i = \lambda_i(H_{t_{\alpha_1}}),$$

is tight for every nonnegative $\chi \in C_c^\infty((-2, 2))$. Therefore, as in the proof of Lemma 6.6, using Lemma 6.2 at the scale $d_2 = \ell_{1/3}$, (9.36) also holds after replacing the weights φ_j on the left side by their smoothings $\varphi_{j,N}$, since $|e^{-a} - e^{-b}| \leq |a - b|$ for $a, b \geq 0$.

Return to H . The observable on the left side of (9.36), with smoothed weights, is of the form considered in Definition 5.4 at the scale $d_2 = \ell_{1/3}$. By (9.4) with $\alpha = 1/3$ and

$t = t_{\alpha_1}$, which is (9.3), the comparison error is $CN^{1/2-2\alpha_1+12\varepsilon} = CN^{-7/50+12\varepsilon} = o(1)$, since $\alpha_1 > 1/4$. Hence (9.36), with smoothed weights, holds with H in place of $H_{t_{\alpha_1}}$.

Next, we show that close triples are absent for H at the scale d_2 . Let $U_{d_2}^{(3)}$ be defined by (9.2) with d_2 in place of q , $p = p_*$, and a nonnegative $\chi \in C_c^\infty((-2, 2))$ with support in the interior of I' and with $\chi \geq 2$ on a neighborhood of I , and let $\eta_\chi > 0$ be such that the $2\eta_\chi$ -neighborhood of $\text{supp } \chi$ lies in I' . We first show that $U_{d_2}^{(3)}(H_{t_{\alpha_1}}) \rightarrow 0$ in probability. A triple of eigenvalues of $H_{t_{\alpha_1}}$ attached to a single three-element cluster has its term bounded by $Cf_{p_*}(\text{gap}(B_K)/d_2)$, so the total contribution of these triples tends to zero in probability by (9.35). A triple attached to at least two clusters contains two eigenvalues at distance at least $q/2$, so these triples contribute at most $CN^3(2d_2/q)^{2p_*} = o(1)$. With probability tending to one, every cluster containing an eigenvalue in I' has at most three elements, and the triples whose mean lies at distance at least η_χ from $\text{supp } \chi$ contribute at most $CN^3 \exp(-\eta_\chi^2 N^{1/5}/2) = o(1)$. Hence $U_{d_2}^{(3)}(H_{t_{\alpha_1}}) \rightarrow 0$ in probability, and $\mathbb{E}e^{-U_{d_2}^{(3)}(H_{t_{\alpha_1}})} \rightarrow 1$.

We now transfer this to H . Applying (9.7) with $r = 3$, d_2 in place of q (that is, $\kappa = 1/3$), $\theta = M = 1$, and $t = t_{\alpha_1}$, we find that $\mathbb{E}e^{-U_{d_2}^{(3)}(H)} \rightarrow 1$, so $U_{d_2}^{(3)}(H) \rightarrow 0$ in probability. If $\lambda_i \in I$ and $\lambda_{i+2} - \lambda_i \leq x_0 d_2$, then the term of $(i, i+1, i+2)$ in $U_{d_2}^{(3)}(H)$ is at least $(1 + x_0^2)^{-3p_*}$ for $N \geq N_0$. Consequently, with probability tending to one, there exists no i with $\lambda_i \in I$ and $\lambda_{i+2} - \lambda_i \leq x_0 d_2$. In particular, every pair $i < i'$ with $\lambda_i \in I$ and $\lambda_{i'} - \lambda_i \leq x_0 d_2$ is then a pair of consecutive eigenvalues.

Conclusion. We have shown that (9.36) holds for H , with smoothed weights. Taking $Q = \theta \chi f_{p_*}$ and letting $\theta \rightarrow 0$, we see that the statistics

$$\sum_{i < i'} \chi_N \left(\frac{\lambda_i + \lambda_{i'}}{2} \right) f_{p_*} \left(\frac{\lambda_{i'} - \lambda_i}{d_2} \right)$$

are tight, and Lemma 5.1 then allows us to remove the smoothing of the weights as in the proof of Lemma 6.6. Since the Laplace transform determines the law of a random vector with nonnegative coordinates, it follows that for all nonnegative $\varphi_1, \dots, \varphi_n \in C_c^\infty((-2, 2))$ and integers $p_1, \dots, p_n \geq p_*$, the vector of statistics

$$\sum_{i < i'} \varphi_j \left(\frac{\lambda_i + \lambda_{i'}}{2} \right) f_{p_j} \left(\frac{\lambda_{i'} - \lambda_i}{d_2} \right), \quad j \in [n],$$

converges jointly in law to the vector of integrals of the functions $(y, E) \mapsto \varphi_j(E) f_{p_j}(y)$ against a Poisson process Π on $[0, \infty) \times (-2, 2)$ with intensity $c_2(E) y^2 dy dE$.

We now pass to general test functions. Every continuous function f with compact support in $[0, \infty) \times (-2, 2)$ can be approximated by such combinations, with signed coefficients, up to an error $\omega f_{p_*}(y) \mathbf{1}_{I'}(E)$, by the weighted polynomial approximation in

the variable $(1 + y^2)^{-1}$ in the proof of Lemma 6.6 and the partition of unity in the proof of Proposition 4.1. Together with the tightness of the statistics $\sum_{i < i'} \chi((\lambda_i + \lambda_{i'})/2) f_{p_*}((\lambda_{i'} - \lambda_i)/d_2)$ shown above, this shows that $\sum_{i < i'} f((\lambda_{i'} - \lambda_i)/d_2, (\lambda_i + \lambda_{i'})/2)$ converges in law to $\int f d\Pi$, and hence that the point process $\sum_{i < i'} \delta_{((\lambda_{i'} - \lambda_i)/d_2, (\lambda_i + \lambda_{i'})/2)}$ converges in distribution to Π with respect to the vague topology.

Finally, we pass to consecutive eigenvalues. By the absence of close triples for H at the scale d_2 , shown above and applied to bulk intervals I containing the E -supports of the test functions, the pairs with $\lambda_{i'} - \lambda_i \leq x_0 d_2$ are pairs of consecutive eigenvalues with probability tending to one, and midpoints may be replaced by left endpoints since they differ by $O(d_2)$. This gives the convergence of the point processes (1.8). The limit (1.9) follows from Lemma 4.12, whose proof applies with Π and its intensity $c_2(E)y^2 dy dE$ in place of Ξ and ν , since $\int_0^x c_2(E)y^2 dy = \pi^2 \rho_{\text{sc}}(E)^4 x^3 / 9$.

Pair statistics of the Gaussian orthogonal ensemble

In this appendix, we prove Proposition 3.6. Throughout, we fix $\alpha \in (0, 1/4)$, $p > 1$, and a function f as in Proposition 3.6, together with the bulk interval I and the constant C_f given there.

We follow the approach of Feng, Li, and Yao [33], who treat the Gaussian symplectic ensemble in detail and sketch the orthogonal case. The correlation functions of the GOE are Pfaffians of a 2×2 matrix kernel built from Hermite functions. The mean of the statistic in (3.16) is an integral of the two-point function, and the variance is controlled by the two-, three-, and four-point functions. However, the GOE analogues of the kernel bounds in [33, Lemma 3.2] are asserted in [33, Section 7.1] without proof, and the expansion of the two-point function that we need is not stated there, so we give complete proofs. They rest on elementary estimates for Hermite functions, proved in Appendix A.1, and on the density of states, which we deduce from eigenvalue rigidity in Lemma A.7. The main step is the exact identity for the slope of the two-point function at the diagonal in Lemma A.8; together with the density of states, it determines the two-point function at separations below N^{-1} . The most technical part of the argument is the decorrelation estimate for two separated pairs of nearby points in Lemma A.9, which controls the variance.

In this appendix only, we abbreviate $\lambda_i = \lambda_i(G)$. For $\delta \in (0, 1)$, we set $I_\delta = [-2 + \delta, 2 - \delta]$. Constants $C, c > 0$ and thresholds N_0 may depend on δ, α, p, I , and f , but not on N .

A.1 Hermite functions

Let He_n denote the monic Hermite polynomials orthogonal with respect to the weight $e^{-w^2/2}$. They are determined by the generating function

$$\sum_{n=0}^{\infty} \frac{\text{He}_n(w) \zeta^n}{n!} = e^{w\zeta - \zeta^2/2}, \quad w, \zeta \in \mathbb{R}. \quad (\text{A.1})$$

For integers $n \geq 0$, we define the Hermite functions

$$h_n(w) = (2\pi)^{-1/4} (n!)^{-1/2} \text{He}_n(w) e^{-w^2/4}, \quad w \in \mathbb{R}, \quad (\text{A.2})$$

and we set $h_{-1} = 0$. The functions h_n form an orthonormal family in $L^2(\mathbb{R})$ (see [2, Section 3.2]), and $h_n(-w) = (-1)^n h_n(w)$. Differentiating (A.1) in w and in ζ , we find

$\text{He}'_n = n\text{He}_{n-1}$ and $w\text{He}_n = \text{He}_{n+1} + n\text{He}_{n-1}$. These two relations imply $\text{He}''_n - w\text{He}'_n + n\text{He}_n = 0$. We set

$$e_n(w) = n + \frac{1}{2} - \frac{w^2}{4}, \quad n \geq 0, w \in \mathbb{R}. \quad (\text{A.3})$$

In terms of the Hermite functions, the three relations above become

$$wh_n = \sqrt{n+1}h_{n+1} + \sqrt{n}h_{n-1}, \quad h'_n = \frac{\sqrt{n}h_{n-1} - \sqrt{n+1}h_{n+1}}{2}, \quad h''_n = -e_n h_n, \quad (\text{A.4})$$

for all $n \geq 0$. For $w, w' \in \mathbb{R}$ and $j \geq 0$, the first identity in (A.4) gives

$$\begin{aligned} (w - w')h_j(w)h_j(w') &= \sqrt{j+1}(h_{j+1}(w)h_j(w') - h_j(w)h_{j+1}(w')) \\ &\quad - \sqrt{j}(h_j(w)h_{j-1}(w') - h_{j-1}(w)h_j(w')). \end{aligned}$$

Summing over $0 \leq j \leq n-1$, we see that the right side telescopes, and we obtain the Christoffel–Darboux formula

$$\sum_{j=0}^{n-1} h_j(w)h_j(w') = \sqrt{n} \frac{h_n(w)h_{n-1}(w') - h_{n-1}(w)h_n(w')}{w - w'}, \quad n \geq 1, w \neq w'. \quad (\text{A.5})$$

Letting $w' \rightarrow w$ in (A.5), we also have

$$\sum_{j=0}^{n-1} h_j(w)^2 = \sqrt{n}(h'_n(w)h_{n-1}(w) - h'_{n-1}(w)h_n(w)). \quad (\text{A.6})$$

For integers $n \geq 0$ and $w \in \mathbb{R}$, we define

$$F_n(w) = \frac{1}{2} \int_{\mathbb{R}} \text{sgn}(w - w')h_n(w')dw', \quad (\text{A.7})$$

and we set $F_{-1} = 0$. Then $F'_n = h_n$. Since h'_n is integrable and $h_n(w) \rightarrow 0$ as $|w| \rightarrow \infty$, we have

$$\frac{1}{2} \int_{\mathbb{R}} \text{sgn}(w - w')h'_n(w')dw' = h_n(w), \quad w \in \mathbb{R}.$$

Inserting the second identity in (A.4) into the left side, we obtain

$$h_n = \frac{\sqrt{n}F_{n-1} - \sqrt{n+1}F_{n+1}}{2}, \quad n \geq 0. \quad (\text{A.8})$$

We will need the following estimates.

Lemma A.1. *Let $\delta \in (0, 1]$. There exists $C > 0$, depending only on δ , such that the following statements hold for every integer $n \geq 1$.*

(a) *For $|w| \leq (2 - \delta)\sqrt{n}$ and $k \in \{0, 1, 2, 3\}$, we have $|h_n^{(k)}(w)| \leq Cn^{k/2-1/4}$.*

- (b) For $|w|, |w'| \leq (2 - \delta)\sqrt{n}$, we have $|\int_w^{w'} h_n| \leq Cn^{-3/4}$.
(c) If n is even, then $C^{-1}n^{-1/4} \leq \int_{\mathbb{R}} h_n \leq Cn^{-1/4}$. If n is odd, then $|\int_0^\infty h_n| \leq Cn^{-1/4}$.
(d) For $|w| \leq (2 - \delta)\sqrt{n}$, we have $|F_n(w)| \leq Cn^{-1/4}$.

Proof. We begin with (a) for $k \in \{0, 1\}$, using the argument of Sonine and Pólya (see [68, Section 7.31]). On the interval $[0, (4n + 2)^{1/2}]$, on which $e_n > 0$, we define

$$\Theta(w) = h_n(w)^2 + \frac{h'_n(w)^2}{e_n(w)}.$$

Using $h''_n = -e_n h_n$ and $e'_n = -w/2$, we find

$$\Theta' = 2h_n h'_n + \frac{2h'_n h''_n}{e_n} - \frac{e'_n (h'_n)^2}{e_n^2} = \frac{w(h'_n)^2}{2e_n^2} \geq 0.$$

Hence Θ is nondecreasing on $[0, (4n + 2)^{1/2}]$.

Set $w_1 = (2 - \delta)\sqrt{n}$ and $w_2 = (2 - \delta/2)\sqrt{n}$. Then $w_2 < (4n + 2)^{1/2}$, and

$$e_n(w) \geq n(1 - (1 - \delta/4)^2) \geq \frac{\delta n}{4}, \quad w \in [w_1, w_2].$$

Let $0 \leq w \leq w_1$. Since Θ is nondecreasing, $\Theta(w) \leq \Theta(w')$ for all $w' \in [w_1, w_2]$. Averaging over w' , we obtain

$$\begin{aligned} \Theta(w) &\leq \frac{1}{w_2 - w_1} \int_{w_1}^{w_2} \Theta(w') dw' \leq \frac{2}{\delta\sqrt{n}} \left(\int_{\mathbb{R}} h_n^2 + \frac{4}{\delta n} \int_{\mathbb{R}} (h'_n)^2 \right) \\ &= \frac{2}{\delta\sqrt{n}} \left(1 + \frac{2n+1}{\delta n} \right) \leq \frac{8}{\delta^2\sqrt{n}}, \end{aligned}$$

where in the equality we used the identity

$$\int_{\mathbb{R}} (h'_n)^2 = \frac{2n+1}{4},$$

which follows from the second identity in (A.4) and orthonormality. Consequently,

$$h_n(w)^2 \leq \Theta(w) \leq \frac{8}{\delta^2\sqrt{n}}, \quad h'_n(w)^2 \leq e_n(w)\Theta(w) \leq \frac{8(n+1)}{\delta^2\sqrt{n}}.$$

The same bounds hold for $-w_1 \leq w \leq 0$ by the parity of h_n , which proves (a) for $k \in \{0, 1\}$.

The cases $k \in \{2, 3\}$ follow from the identity $h''_n = -e_n h_n$ and its consequence

$$h'''_n = (w/2)h_n - e_n h'_n, \tag{A.9}$$

together with $|e_n(w)| \leq n + 1$ and $|w| \leq 2\sqrt{n}$.

Next, we prove (b). We have

$$e_n(w) \geq n(1 - (1 - \delta/2)^2) \geq \frac{\delta n}{2}, \quad |w| \leq (2 - \delta)\sqrt{n}.$$

On this interval, $h_n = -h_n''/e_n$, and two integrations by parts give

$$\int_w^{w'} h_n = -\left[\frac{h_n'}{e_n} - \left(\frac{1}{e_n}\right)' h_n\right]_w^{w'} - \int_w^{w'} \left(\frac{1}{e_n}\right)'' h_n.$$

We have

$$\left(\frac{1}{e_n}\right)' = \frac{w}{2e_n^2}, \quad \left(\frac{1}{e_n}\right)'' = \frac{1}{2e_n^2} + \frac{w^2}{2e_n^3},$$

so $1/e_n$, $(1/e_n)'$, and $(1/e_n)''$ are bounded by Cn^{-1} , $Cn^{-3/2}$, and Cn^{-2} on the interval. Using (a) and $|w' - w| \leq 4\sqrt{n}$, we find

$$\left|\int_w^{w'} h_n\right| \leq C(n^{-1}n^{1/4} + n^{-3/2}n^{-1/4} + \sqrt{n}n^{-2}n^{-1/4}) \leq Cn^{-3/4}.$$

We now prove (c). In the computations below, the exchange of sums and integrals is justified by absolute convergence.

Suppose first that $n = 2l$ with $l \geq 1$. Multiplying (A.1) by $e^{-w^2/4}$ and integrating over $w \in \mathbb{R}$, we have

$$\sum_{n'=0}^{\infty} \frac{\zeta^{n'}}{n'!} \int_{\mathbb{R}} e^{-w^2/4} \text{He}_{n'}(w) dw = \int_{\mathbb{R}} e^{-(w-2\zeta)^2/4 + \zeta^2/2} dw = 2\sqrt{\pi} e^{\zeta^2/2}.$$

Comparing the coefficients of ζ^{2l} , we obtain

$$\int_{\mathbb{R}} e^{-w^2/4} \text{He}_{2l}(w) dw = \frac{2\sqrt{\pi} (2l)!}{2^l l!},$$

and therefore

$$\int_{\mathbb{R}} h_{2l} = \frac{2\sqrt{\pi}}{(2\pi)^{1/4}} \cdot \frac{\sqrt{(2l)!}}{2^l l!} = \frac{2\sqrt{\pi}}{(2\pi)^{1/4}} \left(4^{-l} \binom{2l}{l}\right)^{1/2}.$$

The claim for even n now follows from the elementary bounds

$$(4l)^{-1/2} \leq 4^{-l} \binom{2l}{l} \leq (\pi l)^{-1/2}.$$

Suppose next that $n = 2l + 1$ with $l \geq 1$. For $\zeta \in \mathbb{R}$, we have

$$\int_0^{\infty} e^{-w^2/4} e^{w\zeta - \zeta^2/2} dw = e^{\zeta^2/2} \int_{-2\zeta}^{\infty} e^{-v^2/4} dv = \sqrt{\pi} e^{\zeta^2/2} + 2\zeta \int_0^1 e^{\zeta^2(1/2 - v^2)} dv.$$

The first term on the right side is even in ζ . Let

$$j_l = \int_0^1 (1 - 2v^2)^l dv.$$

Comparing the coefficients of ζ^{2l+1} , we obtain

$$\int_0^\infty h_{2l+1} = \frac{2^{1-l} \sqrt{(2l+1)!}}{(2\pi)^{1/4} l!} j_l = \frac{2\sqrt{2l+1}}{(2\pi)^{1/4}} \left(4^{-l} \binom{2l}{l}\right)^{1/2} j_l.$$

To bound j_l , we use the inequalities

$$\begin{aligned} 0 \leq 1 - 2v^2 &\leq e^{-2v^2}, & v \in [0, 2^{-1/2}], \\ 0 \leq 2v^2 - 1 &\leq 1 - 2(1-v) \leq e^{-2(1-v)}, & v \in [2^{-1/2}, 1]. \end{aligned}$$

They give

$$|j_l| \leq \int_0^\infty e^{-2lv^2} dv + \int_0^\infty e^{-2lv} dv \leq 2l^{-1/2},$$

and therefore

$$\left| \int_0^\infty h_{2l+1} \right| \leq Cl^{1/2} l^{-1/4} l^{-1/2} \leq Cn^{-1/4}.$$

The case $n = 1$ is trivial.

Finally, we prove (d). When n is even, h_n is even, and when n is odd, h_n is odd. Therefore

$$F_n(w) = \int_0^w h_n \quad (n \text{ even}), \quad F_n(w) = \int_0^w h_n - \int_0^\infty h_n \quad (n \text{ odd}),$$

and (b) with $w' = 0$ and (c) give the claim. $\square$

A.2 The Pfaffian formula

The entries G_{ab} with $a \leq b$ are independent centered Gaussian random variables with variance $2/N$ for $a = b$ and $1/N$ for $a < b$, so G has density proportional to $\exp(-N \operatorname{Tr} A^2/4)$ with respect to Lebesgue measure on the space of $N \times N$ real symmetric matrices A . Therefore, the eigenvalues of G , in uniformly random order, have the joint density

$$\wp_N(x_1, \dots, x_N) = \frac{1}{Z_N} \prod_{1 \leq i < j \leq N} |x_i - x_j| \prod_{i=1}^N e^{-Nx_i^2/4} \quad (\text{A.10})$$

on $\mathbb{R}^N$, where Z_N is a normalizing constant (see [2, Section 2.5]; this is [33, (1.1)] for the orthogonal ensemble). For $n \in [N]$, we define the n -point correlation function

$$R_{n,N}(x_1, \dots, x_n) = \frac{N!}{(N-n)!} \int_{\mathbb{R}^{N-n}} \wp_N(x_1, \dots, x_N) dx_{n+1} \cdots dx_N. \quad (\text{A.11})$$

It is nonnegative and symmetric in its arguments. Since $\wp_N$ is symmetric, we have

$$\mathbb{E} \sum_{(i_1, \dots, i_n)} \mathcal{F}(\lambda_{i_1}, \dots, \lambda_{i_n}) = \int_{\mathbb{R}^n} \mathcal{F}(x_1, \dots, x_n) R_{n,N}(x_1, \dots, x_n) dx_1 \cdots dx_n \quad (\text{A.12})$$

for every measurable $\mathcal{F}: \mathbb{R}^n \rightarrow [0, \infty)$, where the sum runs over n -tuples of pairwise distinct indices in $[N]$. By linearity, (A.12) also holds for every bounded measurable $\mathcal{F}$ with compact support.

We recall the definition of the Pfaffian. For an antisymmetric $2n \times 2n$ matrix W , we set

$$\text{Pf} W = \sum_{\mathbf{m}} \text{sgn}(\mathbf{m}) \prod_{\{\sigma, \sigma'\} \in \mathbf{m}, \sigma < \sigma'} W_{\sigma\sigma'}, \quad (\text{A.13})$$

where the sum runs over the perfect matchings $\mathbf{m}$ of $[2n]$, and $\text{sgn}(\mathbf{m})$ is the sign of the permutation $(\sigma_1, \sigma'_1, \dots, \sigma_n, \sigma'_n)$ of $[2n]$ obtained by listing the pairs $\{\sigma_r, \sigma'_r\}$ of $\mathbf{m}$ with $\sigma_r < \sigma'_r$ (this sign does not depend on the order of the pairs). We use three properties. First, $\text{Pf}(B^T W B) = \det(B) \text{Pf}(W)$ for every $2n \times 2n$ matrix B [33, Section 2.2]. Second, let $1 \leq l < n$ and set $P_1 = [2l]$ and $P_2 = [2n] \setminus [2l]$. Let $W_{P_1 P_1}$ and $W_{P_2 P_2}$ denote the corresponding principal submatrices. Then the terms in (A.13) indexed by the matchings $\mathbf{m}$ such that every pair of $\mathbf{m}$ is a subset of P_1 or of P_2 add up to $\text{Pf}(W_{P_1 P_1}) \text{Pf}(W_{P_2 P_2})$. Third, the number of perfect matchings of $[2n]$ is $(2n-1)!!$, so that $|\text{Pf} W| \leq (2n-1)!! \max_{\sigma, \sigma'} |W_{\sigma\sigma'}|^n$.

We now define the matrix kernel of the GOE, following [33, (7.2)] up to the names of the functions.

Definition A.2. Let $N \geq 2$. For $x, y \in \mathbb{R}$, we define

$$K_N(x, y) = \sqrt{N} \sum_{j=0}^{N-1} h_j(\sqrt{N}x) h_j(\sqrt{N}y), \quad P(x, y) = \frac{N}{2} h_{N-1}(\sqrt{N}x) F_N(\sqrt{N}y),$$

and $S = K_N + P$. We define $q(x) = 0$ if N is even and

$$q(x) = \sqrt{N} h_{N-1}(\sqrt{N}x) \left(\int_{\mathbb{R}} h_{N-1} \right)^{-1}$$

if N is odd, where $\int_{\mathbb{R}} h_{N-1} > 0$ by the proof of Lemma A.1(c). We further set $Q(x) = (1/2) \int_{\mathbb{R}} \text{sgn}(x - x') q(x') dx'$ and

$$T(x, y) = S(x, y) + q(x), \quad V(x, y) = -\partial_y S(x, y),$$

$$I(x, y) = \frac{1}{2} \int_{\mathbb{R}} \text{sgn}(x - x') S(x', y) dx' - \frac{1}{2} \text{sgn}(x - y) + Q(x) - Q(y).$$

Finally, we define the 2×2 matrix

$$A(x, y) = \begin{pmatrix} I(x, y) & T(y, x) \\ -T(x, y) & -V(x, y) \end{pmatrix}.$$

We suppress the dependence on N in P , S , q , Q , T , V , l , and A . (The function K_N is the correlation kernel of the eigenvalues of the Gaussian unitary ensemble normalized so that $\mathbb{E}|G_{ab}|^2 = 1/N$ for all a, b , which is the normalization of Section 9 [2, Section 3.2], and P and q are the corrections for the orthogonal ensemble.) The functions K_N , P , S , q , T , and V are smooth, since $F'_N = h_N$. Observe that $S(\cdot, y)$ and q are integrable, so l and Q are well defined, and that $Q' = q$.

Lemma A.3. *Let $N \geq 2$ and $n \in [N]$. For all pairwise distinct $x_1, \dots, x_n \in \mathbb{R}$,*

$$R_{n,N}(x_1, \dots, x_n) = \text{Pf} [A(x_i, x_j)]_{i,j=1}^n, \quad (\text{A.14})$$

where the $2n \times 2n$ matrix on the right has the 2×2 block $A(x_i, x_j)$ in block position (i, j) .

This is the classical Pfaffian, or quaternion determinant, formula of Mehta for the correlation functions of the GOE [56]. In the present normalization, it is [33, (7.1)–(7.2)]. In the notation of [33], the matrix size is n instead of N , the functions T , V , l , q , and Q are denoted $S_n^{(1)} + \alpha_n$, $V_n^{(1)}$, $J_n^{(1)}$, α_n , and γ_n , and $A(x, y)$ is the matrix $JK_n^{(1)}(x, y)$; these symbols are used only in this paragraph. For even N , the matrix kernel is derived by Tracy and Widom [72, (9.1) and Section 10], and for odd N it is given by Adler, Forrester, Nagao, and van Moerbeke [1, (2.11) and (4.3)]. These references use the weight $e^{-x^2/2}$, which corresponds to the eigenvalues of G multiplied by $(N/2)^{1/2}$. By Lemma A.4(a) below, the matrix on the right side of (A.14) is antisymmetric.

We will use the following identities for the kernel.

Lemma A.4. *Let $N \geq 2$ and $x, y \in \mathbb{R}$.*

(a) *We have $V(x, y) = -V(y, x)$ and $l(x, y) = -l(y, x)$. In particular, $A(x, y)^T = -A(y, x)$.*

(b) *We have*

$$l(x, y) = \int_y^x T(x', y) dx' - \frac{1}{2} \text{sgn}(x - y). \quad (\text{A.15})$$

Consequently, $\partial_x l(x, y) = T(x, y)$ and $\partial_y l(x, y) = -T(y, x)$ for $x \neq y$.

(c) *We have $R_{1,N}(x) = T(x, x)$, and if $x \neq y$, then*

$$R_{2,N}(x, y) = T(x, x)T(y, y) - T(x, y)T(y, x) + l(x, y)V(x, y). \quad (\text{A.16})$$

Proof. Throughout the proof, we write $w = \sqrt{N}x$ and $w' = \sqrt{N}y$. We begin with V . The second identity in (A.4) gives

$$\sum_{j=0}^{N-1} h_j(w)h'_j(w') = \frac{1}{2} \sum_{j=1}^{N-1} \sqrt{j} (h_j(w)h_{j-1}(w') - h_{j-1}(w)h_j(w')) - \frac{\sqrt{N}}{2} h_{N-1}(w)h_N(w').$$

Using $F'_N = h_N$, we have

$$\partial_y P(x, y) = \frac{N^{3/2}}{2} h_{N-1}(w)h_N(w'),$$

and therefore

$$\begin{aligned} V(x, y) &= -N \left(\sum_{j=0}^{N-1} h_j(w) h'_j(w') + \frac{\sqrt{N}}{2} h_{N-1}(w) h_N(w') \right) \\ &= -\frac{N}{2} \sum_{j=1}^{N-1} \sqrt{j} (h_j(w) h_{j-1}(w') - h_{j-1}(w) h_j(w')), \end{aligned} \quad (\text{A.17})$$

which is antisymmetric in (x, y) .

Next, we consider I. Substituting $\sqrt{N}x'$ as the variable of integration, we find

$$\frac{1}{2} \int_{\mathbb{R}} \text{sgn}(x - x') \sqrt{N} h_j(\sqrt{N}x') dx' = F_j(w),$$

and therefore

$$\frac{1}{2} \int_{\mathbb{R}} \text{sgn}(x - x') S(x', y) dx' = \sum_{j=0}^{N-1} F_j(w) h_j(w') + \frac{\sqrt{N}}{2} F_{N-1}(w) F_N(w').$$

Inserting (A.8) for $h_j(w')$ and arguing as for (A.17), we find

$$\frac{1}{2} \int_{\mathbb{R}} \text{sgn}(x - x') S(x', y) dx' = \frac{1}{2} \sum_{j=1}^{N-1} \sqrt{j} (F_j(w) F_{j-1}(w') - F_{j-1}(w) F_j(w')). \quad (\text{A.18})$$

Let $l_0(x, y)$ denote the right side of (A.18) plus $Q(x) - Q(y)$, so that

$$l(x, y) = l_0(x, y) - \frac{1}{2} \text{sgn}(x - y).$$

The function l_0 is continuous and antisymmetric, and the function $(x, y) \mapsto \text{sgn}(x - y)$ is antisymmetric. Therefore l is antisymmetric, which completes the proof of (a), since the identity $A(x, y)^T = -A(y, x)$ follows from the antisymmetry of l and V .

We now prove (b). Since l_0 is continuous and antisymmetric, we have $l_0(y, y) = 0$. Further, the fundamental theorem of calculus and $Q' = q$ give

$$\partial_x l_0(x, y) = S(x, y) + q(x) = T(x, y), \quad x \in \mathbb{R}.$$

Hence

$$l_0(x, y) = \int_y^x T(x', y) dx',$$

which is (A.15). The formula for $\partial_x l$ follows at once, and the formula for $\partial_y l$ follows from $l(x, y) = -l(y, x)$.

It remains to prove (c). By (a), we have $l(x, x) = 0$ and $V(x, x) = 0$. Hence $A(x, x)$ has the entries $T(x, x)$ and $-T(x, x)$ off the diagonal and zeros on the diagonal, and (A.14)

with $n = 1$ gives $R_{1,N}(x) = T(x, x)$. For $n = 2$, the definition of A and (a) show that the matrix W in (A.14) is

$$W = \begin{pmatrix} 0 & T(x, x) & l(x, y) & T(y, x) \\ -T(x, x) & 0 & -T(x, y) & -V(x, y) \\ -l(x, y) & T(x, y) & 0 & T(y, y) \\ -T(y, x) & V(x, y) & -T(y, y) & 0 \end{pmatrix}.$$

Therefore

$$R_{2,N}(x, y) = W_{12}W_{34} - W_{13}W_{24} + W_{14}W_{23},$$

which is (A.16). □

A.3 Kernel estimates

For functions of two real variables, we write ∂_1 and ∂_2 for the partial derivatives in the first and second variables. We first bound the kernel K_N and its derivatives in the bulk.

Lemma A.5. *Let $\delta \in (0, 1)$. There exist $C > 0$ and N_0 such that the following statements hold for all $N \geq N_0$ and $x, y \in I_\delta$.*

- (a) *For all integers $k_1, k_2 \geq 0$ with $k_1 + k_2 \leq 3$, we have $|\partial_x^{k_1} \partial_y^{k_2} K_N(x, y)| \leq CN^{1+k_1+k_2}$.*
- (b) *If $x \neq y$, then $|K_N(x, y)| \leq C|x - y|^{-1}$ and $|\partial_y K_N(x, y)| \leq C(N|x - y|^{-1} + |x - y|^{-2})$.*
- (c) *The derivative of the function $x \mapsto K_N(x, x)$ is bounded by CN on I_δ .*
- (d) *We have*

$$\partial_y^2 K_N(x, y)|_{y=x} = -\frac{N}{3} \left(N \left(1 - \frac{x^2}{4} \right) + \frac{1}{2} \right) K_N(x, x) - \frac{N^2}{3} h_N(\sqrt{N}x) h'_{N-1}(\sqrt{N}x). \quad (\text{A.19})$$

Proof. Let $w = \sqrt{N}x$ and $w' = \sqrt{N}y$, and define

$$K^\circ(w, w') = \sum_{j=0}^{N-1} h_j(w) h_j(w'), \quad \text{so that} \quad K_N(x, y) = \sqrt{N} K^\circ(w, w').$$

We have $|w|, |w'| \leq (2 - \delta)\sqrt{N}$, and there exists N_0 such that $(2 - \delta)\sqrt{N} \leq (2 - \delta/2)\sqrt{n}$ for all $N \geq N_0$ and $n \in \{N - 1, N\}$. Hence Lemma A.1(a) with $\delta/2$ in place of δ gives

$$|h_n^{(k)}(w)| + |h_n^{(k)}(w')| \leq CN^{k/2-1/4}, \quad n \in \{N - 1, N\}, \quad k \in \{0, 1, 2, 3\}. \quad (\text{A.20})$$

We begin with (a). Using (A.6) with $n = N$ and (A.20), we obtain

$$\sum_{j=0}^{N-1} h_j(w)^2 = \sqrt{N} (h'_N(w) h_{N-1}(w) - h'_{N-1}(w) h_N(w)) \leq C\sqrt{N}.$$

Squaring the second identity in (A.4) and using the Cauchy–Schwarz inequality, we find

$$\sum_{j=0}^{N-1} h_j'(w)^2 \leq \frac{1}{2} \sum_{j=0}^{N-1} (j h_{j-1}(w)^2 + (j+1) h_{j+1}(w)^2) \leq N \sum_{j=0}^N h_j(w)^2 \leq CN^{3/2},$$

where in the last step we used (A.20) for $h_N(w)$. Next, for $j \leq N-1$, (A.4) and (A.9) give

$$\begin{aligned} |h_j''(w)| &= |e_j(w)| |h_j(w)| \leq 2N |h_j(w)|, \\ |h_j'''(w)| &\leq \frac{|w|}{2} |h_j(w)| + 2N |h_j'(w)| \leq \sqrt{N} |h_j(w)| + 2N |h_j'(w)|. \end{aligned}$$

Combining the last three bounds, we find

$$\sum_{j=0}^{N-1} h_j^{(k)}(w)^2 \leq CN^{k+1/2}, \quad k \in \{0, 1, 2, 3\}, \quad (\text{A.21})$$

and the same bound holds at w' . The Cauchy–Schwarz inequality and (A.21) now give

$$\begin{aligned} |\partial_x^{k_1} \partial_y^{k_2} K_N(x, y)| &= N^{(1+k_1+k_2)/2} \left| \sum_{j=0}^{N-1} h_j^{(k_1)}(w) h_j^{(k_2)}(w') \right| \\ &\leq N^{(1+k_1+k_2)/2} (CN^{k_1+1/2})^{1/2} (CN^{k_2+1/2})^{1/2} = CN^{1+k_1+k_2}. \end{aligned}$$

To prove (b), we write (A.5) in the form

$$K_N(x, y) = \sqrt{N} \frac{Z(w, w')}{x - y}, \quad Z(w, w') = h_N(w) h_{N-1}(w') - h_{N-1}(w) h_N(w').$$

By (A.20), we have

$$|Z(w, w')| \leq CN^{-1/2}, \quad |\partial_{w'} Z(w, w')| \leq C.$$

The bounds in (b) then follow from the last two displays and the identity

$$\partial_y K_N(x, y) = \frac{N \partial_{w'} Z(w, w')}{x - y} + \frac{\sqrt{N} Z(w, w')}{(x - y)^2}.$$

Next, we prove (c). The second identity in (A.4) and a telescoping sum give

$$\frac{d}{dx} K_N(x, x) = 2N \sum_{j=0}^{N-1} h_j(w) h_j'(w) = -N^{3/2} h_{N-1}(w) h_N(w),$$

and the right side is bounded by CN by (A.20).

It remains to prove (d), which holds for all $x \in \mathbb{R}$. We fix w and set $z(w') = Z(w, w')$, so that z is entire and $z(w) = 0$. By (A.5) and the Taylor expansion of z at w , we have

$$K^\circ(w, w') = -\sqrt{N} \frac{z(w')}{w' - w} = -\sqrt{N} \sum_{l=1}^{\infty} \frac{z^{(l)}(w)}{l!} (w' - w)^{l-1}, \quad w' \neq w.$$

Hence

$$K^\circ(w, w) = -\sqrt{N} z'(w), \quad \partial_{w'}^2 K^\circ(w, w')|_{w'=w} = -\frac{\sqrt{N}}{3} z'''(w).$$

To compute z''' , we use (A.9) with $n \in \{N-1, N\}$, together with $e_{N-1} = e_N - 1$. We obtain

$$\begin{aligned} z'''(w) &= -e_{N-1}(w)h_N(w)h'_{N-1}(w) + e_N(w)h_{N-1}(w)h'_N(w) \\ &= -e_N(w)z'(w) + h_N(w)h'_{N-1}(w). \end{aligned}$$

Combining the last two displays, we obtain

$$\partial_{w'}^2 K^\circ(w, w')|_{w'=w} = -\frac{e_N(w)}{3} K^\circ(w, w) - \frac{\sqrt{N}}{3} h_N(w)h'_{N-1}(w).$$

Finally, inserting the relations

$$\partial_y^2 K_N(x, y) = N^{3/2} \partial_w^2 K^\circ(w, w'), \quad K^\circ(w, w) = \frac{K_N(x, x)}{\sqrt{N}}, \quad e_N(\sqrt{N}x) = N\left(1 - \frac{x^2}{4}\right) + \frac{1}{2}$$

into the preceding formula for $\partial_{w'}^2 K^\circ(w, w')|_{w'=w}$, we obtain (A.19). $\square$

We now turn to the functions T , V , and l of Definition A.2.

Lemma A.6. *Let $\delta \in (0, 1)$. There exist $C > 0$ and N_0 such that the following statements hold for all $N \geq N_0$ and $x, y \in I_\delta$.*

- (a) *For all integers $k_1, k_2 \geq 0$ with $k_1 + k_2 \leq 2$, we have $|\partial_x^{k_1} \partial_y^{k_2} T(x, y)| \leq CN^{1+k_1+k_2}$ and $|\partial_x^{k_1} \partial_y^{k_2} V(x, y)| \leq CN^{2+k_1+k_2}$.*
- (b) *If $x \neq y$, then $|T(x, y)| \leq C(|x - y|^{-1} + N^{1/2})$ and $|V(x, y)| \leq C(N|x - y|^{-1} + |x - y|^{-2} + N)$.*
- (c) *We have $|l(x, y)| \leq C(1 + \log(1 + N|x - y|))$.*
- (d) *We have $|T(x, x) - K_N(x, x)| \leq CN^{1/2}$, and the derivative of the function $x \mapsto T(x, x)$ is bounded by $CN^{3/2}$ on I_δ .*
- (e) *We have*

$$\left| \partial_2 V(x, x) - \frac{N}{3} \left(N \left(1 - \frac{x^2}{4} \right) + \frac{1}{2} \right) K_N(x, x) \right| \leq CN^2. \quad (\text{A.22})$$

Proof. We write $w = \sqrt{N}x$ and $w' = \sqrt{N}y$. Let N_0 be as in the proof of Lemma A.5, so that (A.20) holds. Since $|w'| \leq (2 - \delta/2)\sqrt{N}$, Lemma A.1(d) and (c), with $\delta/2$ in place of δ , give

$$|F_N(w')| \leq CN^{-1/4}, \quad \int_{\mathbb{R}} h_{N-1} \geq cN^{-1/4} \quad \text{if } N \text{ is odd.}$$

Recall that $T = K_N + P + q$ and $V = -\partial_y K_N - \partial_y P$. We first bound the corrections P and q . For integers $k_1 \geq 0$ and $k_2 \geq 1$ with $k_1 + k_2 \leq 3$, we have

$$\begin{aligned} \partial_x^{k_1} P(x, y) &= \frac{N^{1+k_1/2}}{2} h_{N-1}^{(k_1)}(w) F_N(w'), \\ \partial_x^{k_1} \partial_y^{k_2} P(x, y) &= \frac{N^{1+(k_1+k_2)/2}}{2} h_{N-1}^{(k_1)}(w) h_N^{(k_2-1)}(w'), \end{aligned}$$

and, for odd N ,

$$q^{(k_1)}(x) = N^{(1+k_1)/2} h_{N-1}^{(k_1)}(w) \left(\int_{\mathbb{R}} h_{N-1} \right)^{-1}.$$

Therefore

$$|\partial_x^{k_1} P(x, y)| + |q^{(k_1)}(x)| \leq CN^{k_1+1/2}, \quad |\partial_x^{k_1} \partial_y^{k_2} P(x, y)| \leq CN^{k_1+k_2}, \quad (\text{A.23})$$

where we used (A.20) and the bounds on $F_N(w')$ and $\int_{\mathbb{R}} h_{N-1}$. The bounds in (a) follow from (A.23) and Lemma A.5(a), and the bounds in (b) follow from (A.23) with $k_1 + k_2 \leq 1$ and Lemma A.5(b).

To prove (c), we use (A.15) and bound the contributions of K_N , P , and q to the integral there separately. By Lemma A.5(a) and (b), we have $|K_N(x', y)| \leq C \min(N, |x' - y|^{-1})$ for $x' \in I_\delta$, so

$$\left| \int_y^x K_N(x', y) dx' \right| \leq C \int_0^{|x-y|} \min(N, v^{-1}) dv \leq C(1 + \log(1 + N|x - y|)).$$

Next, substituting $\sqrt{N}x'$ as the variable of integration, we find

$$\int_y^x P(x', y) dx' = \frac{\sqrt{N}}{2} F_N(w') \int_{w'}^w h_{N-1}, \quad \int_y^x q(x') dx' = \left(\int_{\mathbb{R}} h_{N-1} \right)^{-1} \int_{w'}^w h_{N-1},$$

where the second identity is for odd N . Since $|w|, |w'| \leq (2 - \delta/2)\sqrt{N-1}$, Lemma A.1(b) with $\delta/2$ in place of δ gives

$$\left| \int_{w'}^w h_{N-1} \right| \leq CN^{-3/4},$$

and therefore the integrals of $P(\cdot, y)$ and q from y to x are bounded in absolute value by $CN^{-1/2}$. Together with (A.15), this completes the proof of (c).

We now prove (d). We have

$$\begin{aligned} T(x, x) - K_N(x, x) &= P(x, x) + q(x), \\ \frac{d}{dx} T(x, x) &= \frac{d}{dx} K_N(x, x) + \partial_1 P(x, x) + \partial_2 P(x, x) + q'(x), \end{aligned}$$

and (A.23) and Lemma A.5(c) bound the right sides by $CN^{1/2}$ and $CN^{3/2}$, respectively.

Finally, we prove (e). We have

$$\partial_2 V(x, x) = -\partial_y^2 K_N(x, y)|_{y=x} - \partial_y^2 P(x, y)|_{y=x},$$

and the second term is bounded by CN^2 by (A.23). By (A.19), the first term differs from the main term in (A.22) by $N^2 |h_N(w)h'_{N-1}(w)|/3$, which is bounded by CN^2 by (A.20). $\square$

We now deduce the density of states from eigenvalue rigidity, using the bound on the derivative of $x \mapsto T(x, x)$ in Lemma A.6(d) to pass from intervals to points. Recall that $c(E) = (\pi^2/6)\rho_{\text{sc}}(E)^3$, and note that $1 - x^2/4 = \pi^2 \rho_{\text{sc}}(x)^2$ for $|x| \leq 2$.

Lemma A.7. *Let $\delta \in (0, 1)$. There exist $C > 0$ and N_0 such that for all $N \geq N_0$ and $x \in I_\delta$,*

$$\begin{aligned} |R_{1,N}(x) - N\rho_{\text{sc}}(x)| + |K_N(x, x) - N\rho_{\text{sc}}(x)| &\leq CN^{7/8}, \\ \left| \frac{1}{2} \partial_2 V(x, x) - N^3 c(x) \right| &\leq CN^{23/8}. \end{aligned} \tag{A.24}$$

Proof. Set $\varepsilon = 1/8$, and choose $c_* \in (0, 1/4)$ such that

$$\int_{-2}^{-2+\delta/4} \rho_{\text{sc}} > 2c_*.$$

For $N \geq c_*^{-1}$, the definition of the classical locations and the symmetry $\gamma_{N-i} = -\gamma_i$ then give

$$\gamma_i < -2 + \frac{\delta}{4} \quad (i \leq c_* N + 1), \quad \gamma_i > 2 - \frac{\delta}{4} \quad (i \geq (1 - c_*)N - 1).$$

Let $\mathcal{E}_N$ be the event that $|\lambda_i - \gamma_i| \leq N^{-1+\varepsilon}$ for all i such that $c_* N \leq i \leq (1 - c_*)N$. By Proposition 3.2(b) for G , we have $\mathbb{P}(\mathcal{E}_N^c) \leq N^{-2}$ for $N \geq N_0$.

We first show that for all $y_1 < y_2$ such that $[y_1, y_2] \subset I_{\delta/2}$,

$$\left| \mathbb{E} \# \{i : \lambda_i \in [y_1, y_2]\} - N \int_{y_1}^{y_2} \rho_{\text{sc}} \right| \leq CN^\varepsilon. \tag{A.25}$$

Suppose that $\mathcal{E}_N$ holds. Then, for $i < c_* N$, we have

$$\lambda_i \leq \lambda_{\lceil c_* N \rceil} \leq \gamma_{\lceil c_* N \rceil} + N^{-1+\varepsilon} < y_1,$$

provided $N \geq N_0$, and similarly $\lambda_i > y_2$ for $i > (1 - c_*)N$.

For the remaining indices, we have the implications

$$\begin{aligned}\lambda_i \in [y_1, y_2] &\implies \gamma_i \in [y_1 - N^{-1+\varepsilon}, y_2 + N^{-1+\varepsilon}], \\ \gamma_i \in [y_1 + N^{-1+\varepsilon}, y_2 - N^{-1+\varepsilon}] &\implies \lambda_i \in [y_1, y_2].\end{aligned}$$

Since $\rho_{\text{sc}} \leq 1/\pi$, consecutive classical locations are at distance at least π/N . Hence each of the intervals $[y_1 - N^{-1+\varepsilon}, y_1 + N^{-1+\varepsilon}]$ and $[y_2 - N^{-1+\varepsilon}, y_2 + N^{-1+\varepsilon}]$ contains at most CN^ε classical locations, and $[y_1, y_2]$ contains $N \int_{y_1}^{y_2} \rho_{\text{sc}} + O(1)$ of them. Therefore, on $\mathcal{E}_N$,

$$\#\{i : \lambda_i \in [y_1, y_2]\} = N \int_{y_1}^{y_2} \rho_{\text{sc}} + O(N^\varepsilon).$$

On $\mathcal{E}_N^c$, the left side is bounded by N , and $\text{NP}(\mathcal{E}_N^c) \leq N^{-1}$. Taking expectations, we obtain (A.25).

Now let $x \in I_\delta$, so that $[x, x + N^{-3/4}] \subset I_{\delta/2}$. Using (A.12) with $n = 1$, we have

$$\mathbb{E}\#\{i : \lambda_i \in [x, x + N^{-3/4}]\} = \int_x^{x+N^{-3/4}} R_{1,N}(x') \, dx'.$$

Note that $R_{1,N}(x') = T(x', x')$ by Lemma A.4(c), so Lemma A.6(d) on $I_{\delta/2}$ gives

$$|R_{1,N}(x') - R_{1,N}(x)| \leq CN^{3/2}N^{-3/4} = CN^{3/4}, \quad x' \in [x, x + N^{-3/4}].$$

Using also (A.25) and the fact that ρ_{sc} is Lipschitz on $I_{\delta/2}$, we find

$$\begin{aligned}|R_{1,N}(x) - N\rho_{\text{sc}}(x)| &\leq N^{3/4} \left| \int_x^{x+N^{-3/4}} R_{1,N} - N \int_x^{x+N^{-3/4}} \rho_{\text{sc}} \right| + CN^{3/4} + CN^{1/4} \\ &\leq C(N^{\varepsilon+3/4} + N^{3/4}) \leq CN^{7/8}.\end{aligned}$$

By Lemma A.6(d), we have $|K_N(x, x) - R_{1,N}(x)| \leq CN^{1/2}$, which gives the first bound in (A.24). Inserting it into (A.22) and using $1 - x^2/4 = \pi^2 \rho_{\text{sc}}(x)^2$, we obtain

$$\frac{1}{2} \partial_2 V(x, x) = \left(\frac{\pi^2}{6} N^2 \rho_{\text{sc}}(x)^2 + \frac{N}{12} \right) (N\rho_{\text{sc}}(x) + O(N^{7/8})) + O(N^2) = N^3 c(x) + O(N^{23/8}),$$

which is the second bound in (A.24). $\square$

A.4 Correlation functions

We now estimate the correlation functions. In the next lemma, part (a) determines the mean of the statistic in (3.16), and the cruder bound in part (b) will be used for the tails.

Lemma A.8. *Let $\delta \in (0, 1)$. There exist $C > 0$ and N_0 such that the following statements hold for all $N \geq N_0$.*

(a) For all $x, y \in I_\delta$ such that $0 < y - x \leq N^{-1}$,

$$\begin{aligned} |R_{2,N}(x, y) - N^3 c(x)(y - x)| &\leq C(N^{23/8}(y - x) + N^4(y - x)^2), \\ |R_{2,N}(x, y)| &\leq CN^3(y - x). \end{aligned} \quad (\text{A.26})$$

(b) For all $x, y \in I_\delta$ such that $|x - y| \geq N^{-1}$, we have $|R_{2,N}(x, y)| \leq CN^2 \log N$.

Proof. We begin with (a). We set $y_v = x + v$ for $v \in (0, y - x]$, and we define

$$r(v) = R_{2,N}(x, y_v) = T(x, x)T(y_v, y_v) - T(x, y_v)T(y_v, x) + l(x, y_v)V(x, y_v),$$

where the second equality is (A.16). By Lemma A.4(b) and the antisymmetry of l , we have

$$l(x, y') = \frac{1}{2} - \int_x^{y'} T(x', x) dx', \quad \partial_2 l(x, y') = -T(y', x), \quad y' > x,$$

so r is smooth on $(0, y - x]$. Differentiating, we find

$$\begin{aligned} r'(v) &= T(x, x)(\partial_1 T + \partial_2 T)(y_v, y_v) - \partial_2 T(x, y_v)T(y_v, x) - T(x, y_v)\partial_1 T(y_v, x) \\ &\quad - T(y_v, x)V(x, y_v) + l(x, y_v)\partial_2 V(x, y_v). \end{aligned}$$

We now let $v \downarrow 0$. Then $l(x, y_v) \rightarrow 1/2$, and $V(x, x) = 0$ by Lemma A.4(a). Since T and V are smooth, the first three terms in $r'(v)$ cancel in the limit, and the fourth term tends to $-T(x, x)V(x, x) = 0$. Hence $r(v) \rightarrow 0$ and

$$\lim_{v \downarrow 0} r'(v) = \frac{1}{2} \partial_2 V(x, x). \quad (\text{A.27})$$

Next, we bound r' and r'' . Since $v \leq N^{-1}$, Lemma A.6(c) gives $|l(x, y_v)| \leq C$. The derivatives of T and V of order at most two are bounded on I_δ^2 by Lemma A.6(a), and $\partial_2^2 l(x, y') = -\partial_1 T(y', x)$ for $y' > x$. Differentiating r' once more and inserting these bounds, we obtain

$$|r'(v)| \leq CN^3, \quad |r''(v)| \leq CN^4, \quad v \in (0, y - x].$$

For example, the last term of r' contributes $|l \partial_2^2 V| + |\partial_2 l \partial_2 V| \leq CN^4$ to r'' .

Since $r(v) \rightarrow 0$ as $v \downarrow 0$, we have

$$R_{2,N}(x, y) = \int_0^{y-x} r'(v) dv = \frac{y-x}{2} \partial_2 V(x, x) + \int_0^{y-x} \left(r'(v) - \frac{1}{2} \partial_2 V(x, x) \right) dv,$$

where the last integral is bounded by $CN^4(y - x)^2$ by (A.27) and the bound on r'' . Combining this with Lemma A.7, we obtain the first bound in (A.26). The second bound follows from the first identity in the last display and $|r'| \leq CN^3$.

We now prove (b), using (A.16) and Lemma A.6. Since $|x - y| \geq N^{-1}$, parts (a), (b), and (c) of that lemma give $|T| \leq CN$ on I_δ^2 and

$$|V(x, y)| \leq C(N^2 + N^2 + N) \leq CN^2, \quad |l(x, y)| \leq C(1 + \log(1 + 4N)) \leq C \log N,$$

where we also used $|x - y| \leq 4$. Inserting these bounds into (A.16), we obtain $|R_{2,N}(x, y)| \leq CN^2 \log N$. $\square$

To bound the higher correlation functions, we normalize the entries of the matrix in (A.14). Let $D_N = \text{diag}(1, N^{-1})$. We let B_n denote the block diagonal matrix with n diagonal blocks equal to D_N . Then $\det B_n = N^{-n}$, and the first property of the Pfaffian gives

$$R_{n,N}(x_1, \dots, x_n) = N^n \text{Pf} [D_N A(x_i, x_j) D_N]_{i,j=1}^n, \quad (\text{A.28})$$

where

$$D_N A(x, y) D_N = \begin{pmatrix} l(x, y) & N^{-1} T(y, x) \\ -N^{-1} T(x, y) & -N^{-2} V(x, y) \end{pmatrix}.$$

For the matrix in (A.28), we call row $2i - 1$ the first row of the point x_i and row $2i$ its second row, and similarly for columns. By Lemma A.6(a), all entries of $D_N A(x, y) D_N$ other than $l(x, y)$ are bounded by a constant for $x, y \in I_\delta$. For $\varrho > 0$, we set

$$\Delta_0(\varrho) = (N\varrho)^{-1} + N^{-1/2}. \quad (\text{A.29})$$

Lemma A.9. *Let $\delta \in (0, 1)$. There exist $C > 0$ and N_0 such that the following statements hold for all $N \geq N_0$.*

- (a) *For all pairwise distinct $x_1, x_2, x_3 \in I_\delta$ such that $|x_i - x_j| \leq N^{-1}$ for all i, j , we have $|R_{3,N}(x_1, x_2, x_3)| \leq CN^3$.*
- (b) *For all pairwise distinct $x_1, x_2, x_3, x_4 \in I_\delta$, we have $|R_{4,N}(x_1, x_2, x_3, x_4)| \leq CN^4(\log N)^2$.*
- (c) *Let $x_1, x_2, x_3, x_4 \in I_\delta$ with $x_1 < x_2 \leq x_1 + N^{-1}$ and $x_3 < x_4 \leq x_3 + N^{-1}$, and set $\varrho = |x_3 - x_1|$. Suppose that $\varrho \geq \max(4(x_2 - x_1 + x_4 - x_3), N^{-1})$. Then*

$$|R_{4,N}(x_1, x_2, x_3, x_4) - R_{2,N}(x_1, x_2) R_{2,N}(x_3, x_4)| \leq CN^4(\log N) \Delta_0(\varrho). \quad (\text{A.30})$$

Proof. We begin with (a). The points are pairwise within distance N^{-1} , so Lemma A.6(c) shows that the entries $l(x_i, x_j)$ are bounded. Hence all entries of the 6×6 matrix in (A.28) are bounded, and the third property of the Pfaffian gives $|R_{3,N}| \leq 15C^3 N^3$.

Next, we prove (b). The entries $l(x_i, x_j)$ of the 8×8 matrix in (A.28) are bounded by $C \log N$ by Lemma A.6(c), and all other entries are bounded. The entries $l(x_i, x_j)$ occupy the positions where a first row meets a first column. There are four first rows, so each perfect matching of $[8]$ contains at most two pairs $\{\sigma, \sigma'\}$ for which σ and σ' are first rows. Hence each term in (A.13) is bounded by $C(\log N)^2$, and $|R_{4,N}| \leq 105 CN^4(\log N)^2$.

We now prove (c), following the use of congruence transformations in [33, Section 5.2]. Let W denote the 8×8 matrix in (A.28) for the points x_1, x_2, x_3, x_4 . We define W' as the matrix obtained from W by subtracting row 1 from row 3 and row 5 from row 7, followed by the same operations on the columns. Thus row 3 of W' is the difference of the first rows of x_2 and x_1 , and row 7 is the difference of the first rows of x_4 and x_3 . We have $W' = B^T W B$ for a matrix B with $\det B = 1$, so

$$\text{Pf } W' = \text{Pf } W = N^{-4} R_{4,N}(x_1, x_2, x_3, x_4).$$

Let $P_1 = \{1, 2, 3, 4\}$ and $P_2 = \{5, 6, 7, 8\}$. The transformation does not mix P_1 and P_2 , so $W'_{P_1 P_1}$ is the congruence transform of $W_{P_1 P_1}$ by a matrix of determinant one, and (A.28) with $n = 2$ gives

$$\text{Pf } W'_{P_1 P_1} = \text{Pf } W_{P_1 P_1} = N^{-2} R_{2,N}(x_1, x_2), \quad \text{Pf } W'_{P_2 P_2} = N^{-2} R_{2,N}(x_3, x_4).$$

We next bound the entries of W' . Observe that for $x' \in [x_1, x_2]$ and $j \in \{3, 4\}$,

$$|x' - x_j| \geq \varrho - (x_2 - x_1) - (x_4 - x_3) \geq \frac{\varrho}{2},$$

and the same holds for $x' \in [x_3, x_4]$ and $j \in \{1, 2\}$. Each entry of $W'_{P_1 P_1}$ is a signed sum of at most four entries of $W_{P_1 P_1}$, and $|x_1 - x_2| \leq N^{-1}$, so the entries of $W'_{P_1 P_1}$ are bounded by a constant by Lemma A.6(a) and (c). The same holds for $W'_{P_2 P_2}$.

Now consider an entry $W'_{\sigma\sigma'}$ with $\sigma \in P_1$ and $\sigma' \in P_2$. When σ and σ' are second rows, we have $W'_{\sigma\sigma'} = -N^{-2}V(x_i, x_j)$ for some $i \in \{1, 2\}$ and $j \in \{3, 4\}$. When exactly one of σ and σ' is a first row, the entry $W'_{\sigma\sigma'}$ is of the form $\pm N^{-1}T(x_i, x_j)$ or $\pm N^{-1}T(x_j, x_i)$ with $i \in \{1, 2\}$ and $j \in \{3, 4\}$, or a difference of two such entries. By Lemma A.6(b) and $N\varrho \geq 1$,

$$\begin{aligned} N^{-1}|T(x_i, x_j)| + N^{-1}|T(x_j, x_i)| &\leq C((N\varrho)^{-1} + N^{-1/2}) \leq C\Delta_0(\varrho), \\ N^{-2}|V(x_i, x_j)| &\leq C((N\varrho)^{-1} + (N\varrho)^{-2} + N^{-1}) \leq C\Delta_0(\varrho) \end{aligned}$$

for $i \in \{1, 2\}$ and $j \in \{3, 4\}$, so $|W'_{\sigma\sigma'}| \leq C\Delta_0(\varrho)$ in these cases.

It remains to consider the first rows $\sigma \in \{1, 3\}$ and $\sigma' \in \{5, 7\}$, and there are four cases. First, the entry $W'_{15} = I(x_1, x_3)$ is not small, but it is bounded by $C \log N$ by Lemma A.6(c). Second, since $x_3 \notin [x_1, x_2]$, (A.15) gives

$$W'_{35} = I(x_2, x_3) - I(x_1, x_3) = \int_{x_1}^{x_2} T(x', x_3) dx',$$

so Lemma A.6(b) and the assumption that $x_2 - x_1 \leq N^{-1}$ give

$$|W'_{35}| \leq C(x_2 - x_1)(\varrho^{-1} + N^{1/2}) \leq C\Delta_0(\varrho).$$

Third, by the antisymmetry of l and (A.15),

$$W'_{17} = l(x_1, x_4) - l(x_1, x_3) = - \int_{x_3}^{x_4} T(x', x_1) dx',$$

which is bounded in the same way. Fourth, the same argument gives

$$W'_{37} = \int_{x_1}^{x_2} (T(x', x_4) - T(x', x_3)) dx',$$

which is also bounded by $C\Delta_0(\varrho)$.

Finally, we expand $\text{Pf} W'$ according to (A.13). By the second property of the Pfaffian, the matchings m such that every pair of m is a subset of P_1 or of P_2 contribute $\text{Pf} W'_{P_1 P_1} \text{Pf} W'_{P_2 P_2}$. Since $|P_1| = 4$ is even, every other matching contains an even number of pairs joining P_1 and P_2 , hence at least two. Among these pairs, only the pair $\{1, 5\}$, which occurs at most once, carries an entry that is not bounded by $C\Delta_0(\varrho)$. Hence the corresponding term contains at least one factor bounded by $C\Delta_0(\varrho)$, at most one factor bounded by $C \log N$, and other factors bounded by C . Using $\Delta_0(\varrho) \leq 2$, we conclude that

$$|N^{-4} R_{4,N}(x_1, x_2, x_3, x_4) - N^{-4} R_{2,N}(x_1, x_2) R_{2,N}(x_3, x_4)| \leq 105 C(\log N) \Delta_0(\varrho),$$

which is (A.30). □

A.5 Mean, variance, and tails

We now prove Proposition 3.6. Recall from (3.15) that $\ell_\alpha = N^{-1-\alpha}$ and $M_\alpha = N^{1-2\alpha} = N^3 \ell_\alpha^2$. We choose $\delta \in (0, 1/4)$ such that $I \subset I_{4\delta}$. For a bounded measurable function $g: [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$, we define

$$\mathcal{X}_\alpha(G; g) = \sum_{1 \leq i < j \leq N} g\left(\frac{\lambda_j - \lambda_i}{\ell_\alpha}, \frac{\lambda_i + \lambda_j}{2}\right), \quad \mathcal{L}(g) = \int_{-2}^2 \int_0^\infty c(E) r g(r, E) dr dE, \quad (\text{A.31})$$

where $\mathcal{L}(g)$ is defined whenever the integral converges absolutely. The statistic $\mathcal{X}_\alpha(G; g)$ extends the pair statistic (5.1) of G to functions g that are not products of a function of E and a function of r . We also set

$$g^\sharp(x, y) = g\left(\frac{|y - x|}{\ell_\alpha}, \frac{x + y}{2}\right), \quad x, y \in \mathbb{R},$$

so that

$$\mathcal{X}_\alpha(G; g) = \frac{1}{2} \sum_{i \neq j} g^\sharp(\lambda_i, \lambda_j).$$

Throughout, we consider functions g that vanish for $E \notin I$. If such an g is nonnegative or vanishes for all sufficiently large r , then (A.12) with $n = 2$ and the substitution $y = x + v$ give

$$\begin{aligned}\mathbb{E}\mathcal{X}_\alpha(G;g) &= \int_{x < y} g\left(\frac{y-x}{\ell_\alpha}, \frac{x+y}{2}\right) R_{2,N}(x,y) \, dx \, dy \\ &= \int_0^\infty \int_{\mathbb{R}} g\left(\frac{v}{\ell_\alpha}, x + \frac{v}{2}\right) R_{2,N}(x, x+v) \, dx \, dv.\end{aligned}\tag{A.32}$$

The integrand in (A.32) vanishes unless $x + v/2 \in I \subset I_{4\delta}$, and in that case $x, x+v \in I_{3\delta}$ whenever $v \leq \delta$.

The mean. Let $r_0 \geq 1$. Let g be bounded and measurable with $g(r, E) = 0$ for $r > r_0$ or $E \notin I$. We claim that

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}\mathcal{X}_\alpha(G;g)}{M_\alpha} = \mathcal{L}(g).\tag{A.33}$$

Let N be large enough that $r_0 \ell_\alpha \leq N^{-1} \leq \delta$. The integrand in (A.32) vanishes unless $v \leq r_0 \ell_\alpha$, and in that case $x, x+v \in I_{3\delta}$. Using Lemma A.8(a) with 3δ in place of δ , we obtain

$$\begin{aligned}\mathbb{E}\mathcal{X}_\alpha(G;g) &= N^3 \int_0^{r_0 \ell_\alpha} \int_{\mathbb{R}} g\left(\frac{v}{\ell_\alpha}, x + \frac{v}{2}\right) c(x) v \, dx \, dv + O\left(\int_0^{r_0 \ell_\alpha} (N^{23/8} v + N^4 v^2) \, dv\right) \\ &= N^3 \ell_\alpha^2 \int_0^{r_0} \int_I g(r, E) c\left(E - \frac{\ell_\alpha r}{2}\right) r \, dE \, dr + O(N^{23/8} \ell_\alpha^2 + N^4 \ell_\alpha^3),\end{aligned}$$

where the implicit constants depend on r_0 and $\|g\|_\infty$, and in the second line we substituted $v = \ell_\alpha r$ and $x = E - \ell_\alpha r/2$. Since the function $E \mapsto c(E)$ is uniformly continuous on $[-2, 2]$, we have $c(E - \ell_\alpha r/2) - c(E) \rightarrow 0$ uniformly in $E \in I$ and $r \in [0, r_0]$. Dividing by $M_\alpha = N^3 \ell_\alpha^2$ and using $N^{23/8} \ell_\alpha^2 = N^{-1/8} M_\alpha$ and $N^4 \ell_\alpha^3 = N^{-\alpha} M_\alpha$, we obtain (A.33).

The variance. Let r_0 and g be as in the previous paragraph. We claim that

$$\lim_{N \rightarrow \infty} \frac{\text{Var } \mathcal{X}_\alpha(G;g)}{M_\alpha^2} = 0.\tag{A.34}$$

Expanding the square of $\sum_{i \neq j} g^\sharp(\lambda_i, \lambda_j)$ according to the number of indices that the two pairs share, and using (A.12) for $n \in \{2, 3, 4\}$, we find

$$\begin{aligned}\text{Var } \mathcal{X}_\alpha(G;g) &= \frac{1}{2} \int g^\sharp(x_1, x_2)^2 R_{2,N}(x_1, x_2) + \int g^\sharp(x_1, x_2) g^\sharp(x_1, x_3) R_{3,N}(x_1, x_2, x_3) \\ &\quad + \frac{1}{4} \int g^\sharp(x_1, x_2) g^\sharp(x_3, x_4) (R_{4,N}(x_1, x_2, x_3, x_4) - R_{2,N}(x_1, x_2) R_{2,N}(x_3, x_4)),\end{aligned}\tag{A.35}$$

where the integrals are over $\mathbb{R}^2$, $\mathbb{R}^3$, and $\mathbb{R}^4$, and where we used

$$(\mathbb{E}\mathcal{X}_\alpha(G;g))^2 = \frac{1}{4} \int g^\sharp(x_1, x_2) g^\sharp(x_3, x_4) R_{2,N}(x_1, x_2) R_{2,N}(x_3, x_4) dx_1 \cdots dx_4.$$

Let N be large enough that $8r_0\ell_\alpha \leq N^{-1} \leq \delta$. We bound the three terms in (A.35) in turn, with constants depending on r_0 and $\|g\|_\infty$.

Since $R_{2,N} \geq 0$, the first term in (A.35) is bounded by $\|g\|_\infty \mathbb{E}\mathcal{X}_\alpha(G;|g|)$, which is $O(M_\alpha)$ by (A.33) applied to $|g|$. In the second term, the integrand vanishes unless $|x_2 - x_1|, |x_3 - x_1| \leq r_0\ell_\alpha$ and $x_1, x_2, x_3 \in I_{3\delta}$. The three points are then pairwise within distance $2r_0\ell_\alpha \leq N^{-1}$, and Lemma A.9(a) bounds the second term by

$$CN^3 \cdot |I_{3\delta}| (2r_0\ell_\alpha)^2 \leq CM_\alpha.$$

We now turn to the third term. The integrand is invariant under exchanging x_1 and x_2 and under exchanging x_3 and x_4 , so we may replace the factor $1/4$ by 1 and restrict the integral to the region where $x_1 < x_2$ and $x_3 < x_4$. On this region, the integrand vanishes unless $x_2 - x_1 \leq r_0\ell_\alpha$, $x_4 - x_3 \leq r_0\ell_\alpha$, and all four points lie in $I_{3\delta}$. Set $\varrho_0 = N^{-1/2}$ and $\varrho = |x_3 - x_1|$. We first consider the part of the region where $\varrho \leq \varrho_0$. There we use Lemma A.9(b) and the bound $|R_{2,N}(x_1, x_2)| \leq CN^3(x_2 - x_1) \leq CN^2$ from Lemma A.8(a). This part of the support has measure at most $|I_{3\delta}|(r_0\ell_\alpha)^2 \cdot 2\varrho_0$, so its contribution is bounded by

$$CN^4(\log N)^2 \ell_\alpha^2 \varrho_0.$$

In the remaining part, where $\varrho > \varrho_0$, we have

$$\varrho \geq N^{-1} \geq 8r_0\ell_\alpha \geq 4(x_2 - x_1 + x_4 - x_3),$$

so Lemma A.9(c) applies. The contribution of this part is bounded by

$$\begin{aligned} & CN^4(\log N)(r_0\ell_\alpha)^2 \int_{I_{3\delta}} \int_{\{x_3: \varrho_0 < |x_3 - x_1| \leq 4\varrho\}} \Delta_0(|x_3 - x_1|) dx_3 dx_1 \\ & \leq CN^4 \ell_\alpha^2 (\log N) (N^{-1} \log N + N^{-1/2}). \end{aligned}$$

Collecting the bounds and using $N^4 \ell_\alpha^2 / M_\alpha^2 = N^{-2} \ell_\alpha^{-2} = N^{2\alpha}$, we obtain

$$\frac{\text{Var } \mathcal{X}_\alpha(G;g)}{M_\alpha^2} \leq C(M_\alpha^{-1} + N^{2\alpha}(\log N)^2 N^{-1/2}) = C(N^{2\alpha-1} + (\log N)^2 N^{2\alpha-1/2}).$$

The right side tends to zero because $\alpha < 1/4$, which proves (A.34). This is the only place in the proof where the assumption $\alpha < 1/4$ is used.

The tails. Let $r_0 \geq 1$. Let g be measurable with $|g(r, E)| \leq C_f(1 + r^2)^{-p}$ for all (r, E) and $g(r, E) = 0$ for $r < r_0$ or $E \notin I$. We claim that

$$\limsup_{N \rightarrow \infty} \frac{\mathbb{E} \mathcal{X}_\alpha(G; |g|)}{M_\alpha} \leq CC_f \int_{r_0}^{\infty} r(1 + r^2)^{-p} dr. \quad (\text{A.36})$$

To prove this, we let $N \geq \delta^{-1}$ and split the integral over v in (A.32), with $|g|$ in place of g , into three ranges. In the range $r_0 \ell_\alpha \leq v \leq N^{-1}$, Lemma A.8(a) gives $|R_{2,N}(x, x + v)| \leq CN^3 v$, and the substitution $v = \ell_\alpha r$ bounds the contribution of this range by

$$CC_f |I| \int_{r_0 \ell_\alpha}^{N^{-1}} \left(1 + \frac{v^2}{\ell_\alpha^2}\right)^{-p} N^3 v dv \leq CC_f M_\alpha \int_{r_0}^{\infty} r(1 + r^2)^{-p} dr.$$

In the range $N^{-1} \leq v \leq \delta$, we use $(1 + v^2/\ell_\alpha^2)^{-p} \leq (\ell_\alpha/v)^{2p}$ and Lemma A.8(b) on $I_{3\delta}$. The contribution of this range is bounded by

$$\begin{aligned} CC_f N^2 (\log N) \int_{N^{-1}}^{\delta} \left(\frac{\ell_\alpha}{v}\right)^{2p} dv &\leq CN^{2p+1} \ell_\alpha^{2p} \log N \\ &= CN^{1-2\alpha p} \log N = CM_\alpha N^{-2\alpha(p-1)} \log N. \end{aligned}$$

The range $v > \delta$ contributes the expectation of the sum of $|g((\lambda_j - \lambda_i)/\ell_\alpha, (\lambda_i + \lambda_j)/2)|$ over the pairs $i < j$ such that $\lambda_j - \lambda_i > \delta$. There are at most N^2 such pairs, and each term is bounded by $C_f(\ell_\alpha/\delta)^{2p}$, so the contribution of this range is bounded by $CN^2 \ell_\alpha^{2p} = CN^{2-2p-2\alpha p}$. Since $p > 1$, the contributions of the last two ranges are $o(M_\alpha)$, which proves (A.36).

Conclusion. Let $r_0 \geq 1$. Set $g_1(r, E) = \mathbf{1}_{r \leq r_0} f(r, E)$ and $g_2 = f - g_1$. Then g_1 satisfies the hypotheses of (A.33) and (A.34), and g_2 satisfies the hypotheses of (A.36). By the Cauchy-Schwarz inequality, (A.33), and (A.34),

$$\mathbb{E} \left| \frac{\mathcal{X}_\alpha(G; g_1)}{M_\alpha} - \mathcal{L}(g_1) \right| \leq \left(\frac{\text{Var } \mathcal{X}_\alpha(G; g_1)}{M_\alpha^2} + \left(\frac{\mathbb{E} \mathcal{X}_\alpha(G; g_1)}{M_\alpha} - \mathcal{L}(g_1) \right)^2 \right)^{1/2} \rightarrow 0.$$

Since $\mathcal{X}_\alpha(G; f) = \mathcal{X}_\alpha(G; g_1) + \mathcal{X}_\alpha(G; g_2)$ and $\mathcal{L}(f) = \mathcal{L}(g_1) + \mathcal{L}(g_2)$, it follows from (A.36) that

$$\limsup_{N \rightarrow \infty} \mathbb{E} \left| \frac{\mathcal{X}_\alpha(G; f)}{M_\alpha} - \mathcal{L}(f) \right| \leq \limsup_{N \rightarrow \infty} \frac{\mathbb{E} \mathcal{X}_\alpha(G; |g_2|)}{M_\alpha} + |\mathcal{L}(g_2)| \leq CC_f \int_{r_0}^{\infty} r(1 + r^2)^{-p} dr,$$

where in the last inequality we also used

$$|\mathcal{L}(g_2)| \leq C_f \left(\sup_{E \in [-2, 2]} c(E) \right) |I| \int_{r_0}^{\infty} r(1 + r^2)^{-p} dr.$$

Since $p > 1$, the right side tends to zero as $r_0 \rightarrow \infty$. Hence $\mathcal{X}_\alpha(G; f)/M_\alpha \rightarrow \mathcal{L}(f)$ in L^1 , and therefore in probability by Markov's inequality. Since $\mathcal{X}_\alpha(G; f)/M_\alpha$ and $\mathcal{L}(f)$ are the two sides of (3.16), this completes the proof of Proposition 3.6.

References

- [1] Mark Adler, Peter J. Forrester, Taro Nagao, and Pierre van Moerbeke. Classical skew orthogonal polynomials and random matrices. *J. Stat. Phys.*, 99(1-2):141–170, 2000. doi: 10.1023/A:1018644606835.
- [2] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. doi: 10.1017/CBO9780511801334.
- [3] Gérard Ben Arous and Paul Bourgade. Extreme gaps between eigenvalues of random matrices. *Ann. Probab.*, 41(4):2648–2681, 2013. doi: 10.1214/11-AOP710.
- [4] Florent Benaych-Georges and Antti Knowles. Local semicircle law for Wigner matrices. In *Advanced Topics in Random Matrices*, volume 53 of *Panor. Synthèses*, pages 1–90. Soc. Math. France, Paris, 2018. Theorem numbers refer to arXiv:1601.04055v4.
- [5] Lucas Benigni and Patrick Lopatto. Fluctuations in local quantum unique ergodicity for generalized Wigner matrices. *Comm. Math. Phys.*, 391(2):401–454, 2022. doi: 10.1007/s00220-022-04314-z.
- [6] M. V. Berry and M. Tabor. Level clustering in the regular spectrum. *Proc. R. Soc. Lond. A*, 356(1686):375–394, 1977. doi: 10.1098/rspa.1977.0140.
- [7] Rabi N. Bhattacharya and R. Ranga Rao. *Normal Approximation and Asymptotic Expansions*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2010. doi: 10.1137/1.9780898719895.
- [8] O. Bohigas, M.-J. Giannoni, and C. Schmit. Characterization of chaotic quantum spectra and universality of level fluctuation laws. *Phys. Rev. Lett.*, 52(1):1–4, 1984. doi: 10.1103/PhysRevLett.52.1.
- [9] Charles Bordenave, Gábor Lugosi, and Nikita Zhivotovskiy. Noise sensitivity of the top eigenvector of a Wigner matrix. *Probab. Theory Related Fields*, 177(3-4):1103–1135, 2020. doi: 10.1007/s00440-020-00970-1.
- [10] Paul Bourgade. Extreme gaps between eigenvalues of Wigner matrices. *J. Eur. Math. Soc. (JEMS)*, 24(8):2823–2873, 2022. doi: 10.4171/JEMS/1141.
- [11] Paul Bourgade and Horng-Tzer Yau. The eigenvector moment flow and local quantum unique ergodicity. *Comm. Math. Phys.*, 350(1):231–278, 2017. doi: 10.1007/s00220-016-2627-6.

- [12] Paul Bourgade, László Erdős, Horng-Tzer Yau, and Jun Yin. Fixed energy universality for generalized Wigner matrices. *Comm. Pure Appl. Math.*, 69(10):1815–1881, 2016. doi: 10.1002/cpa.21624.
- [13] Marcelo Campos, Matthew Jenssen, Marcus Michelen, and Julian Sahasrabudhe. The least singular value of a random symmetric matrix. *Forum Math. Pi*, 12:e3, 2024. doi: 10.1017/fmp.2023.29.
- [14] Christophe Charlier. Smallest gaps of the two-dimensional Coulomb gas, 2025. Preprint, arXiv:2507.23502.
- [15] Sourav Chatterjee. A generalization of the Lindeberg principle. *Ann. Probab.*, 34(6): 2061–2076, 2006. doi: 10.1214/0091179060000000575.
- [16] Sourav Chatterjee. *Superconcentration and Related Topics*. Springer Monographs in Mathematics. Springer, Cham, 2014. doi: 10.1007/978-3-319-03886-5.
- [17] P. Deift, T. Nanda, and C. Tomei. Ordinary differential equations and the symmetric eigenvalue problem. *SIAM J. Numer. Anal.*, 20(1):1–22, 1983. doi: 10.1137/0720001.
- [18] Percy Deift. Some open problems in random matrix theory and the theory of integrable systems. In *Integrable Systems and Random Matrices*, volume 458 of *Contemp. Math.*, pages 419–430. Amer. Math. Soc., Providence, RI, 2008. doi: 10.1090/conm/458/08951.
- [19] Freeman J. Dyson. A Brownian-motion model for the eigenvalues of a random matrix. *J. Math. Phys.*, 3(6):1191–1198, 1962. doi: 10.1063/1.1703862.
- [20] Freeman J. Dyson. Statistical theory of the energy levels of complex systems. I. *J. Math. Phys.*, 3(1):140–156, 1962. doi: 10.1063/1.1703773.
- [21] Freeman J. Dyson. The threefold way. Algebraic structure of symmetry groups and ensembles in quantum mechanics. *J. Math. Phys.*, 3(6):1199–1215, 1962. doi: 10.1063/1.1703863.
- [22] László Erdős and Horng-Tzer Yau. Gap universality of generalized Wigner and β -ensembles. *J. Eur. Math. Soc. (JEMS)*, 17(8):1927–2036, 2015. doi: 10.4171/JEMS/548.
- [23] László Erdős and Horng-Tzer Yau. *A Dynamical Approach to Random Matrix Theory*, volume 28 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017. doi: 10.1090/cln/028.
- [24] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Local semicircle law and complete delocalization for Wigner random matrices. *Comm. Math. Phys.*, 287(2): 641–655, 2009. doi: 10.1007/s00220-008-0636-9.
- [25] László Erdős, Sandrine Péché, José A. Ramírez, Benjamin Schlein, and Horng-Tzer Yau. Bulk universality for Wigner matrices. *Comm. Pure Appl. Math.*, 63(7):895–925, 2010. doi: 10.1002/cpa.20317.

- [26] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Universality of random matrices and local relaxation flow. *Invent. Math.*, 185(1):75–119, 2011. doi: 10.1007/s00222-010-0302-7.
- [27] László Erdős, Horng-Tzer Yau, and Jun Yin. Universality for generalized Wigner matrices with Bernoulli distribution. *J. Comb.*, 2(1):15–81, 2011. doi: 10.4310/JOC.2011.v2.n1.a2.
- [28] László Erdős, Horng-Tzer Yau, and Jun Yin. Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields*, 154(1-2):341–407, 2012. doi: 10.1007/s00440-011-0390-3.
- [29] László Erdős, Horng-Tzer Yau, and Jun Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.*, 229(3):1435–1515, 2012. doi: 10.1016/j.aim.2011.12.010.
- [30] Renjie Feng and Dongyi Wei. Small gaps of circular β -ensemble. *Ann. Probab.*, 49(2): 997–1032, 2021. doi: 10.1214/20-AOP1468.
- [31] Renjie Feng and Dongyi Wei. Large gaps of CUE and GUE. *Ann. Probab.*, 53(5): 1668–1702, 2025. doi: 10.1214/24-AOP1747.
- [32] Renjie Feng, Gang Tian, and Dongyi Wei. Small gaps of GOE. *Geom. Funct. Anal.*, 29(6):1794–1827, 2019. doi: 10.1007/s00039-019-00520-5.
- [33] Renjie Feng, Jiaming Li, and Dong Yao. Small gaps of GSE, 2024. Preprint, arXiv:2409.03324.
- [34] Alessio Figalli and Alice Guionnet. Universality in several-matrix models via approximate transport maps. *Acta Math.*, 217(1):81–176, 2016. doi: 10.1007/s11511-016-0142-4.
- [35] Peter J. Forrester. *Log-Gases and Random Matrices*, volume 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010. doi: 10.1515/9781400835416.
- [36] Thomas Guhr, Axel Müller-Groeling, and Hans A. Weidenmüller. Random-matrix theories in quantum physics: common concepts. *Phys. Rep.*, 299(4–6):189–425, 1998. doi: 10.1016/S0370-1573(97)00088-4.
- [37] Fritz Haake, Sven Gnutzmann, and Marek Kuś. *Quantum Signatures of Chaos*. Springer Series in Synergetics. Springer, Cham, fourth edition, 2018. doi: 10.1007/978-3-319-97580-1.
- [38] R. U. Haq, A. Pandey, and O. Bohigas. Fluctuation properties of nuclear energy levels: do theory and experiment agree? *Phys. Rev. Lett.*, 48(16):1086–1089, 1982. doi: 10.1103/PhysRevLett.48.1086.
- [39] Emilie V. Haynsworth. Determination of the inertia of a partitioned Hermitian matrix. *Linear Algebra Appl.*, 1(1):73–81, 1968. doi: 10.1016/0024-3795(68)90050-5.

- [40] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, second edition, 2013. doi: 10.1017/CBO9781139020411.
- [41] Kurt Johansson. Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices. *Comm. Math. Phys.*, 215(3):683–705, 2001. doi: 10.1007/s002200000328.
- [42] Olav Kallenberg. *Random Measures, Theory and Applications*. Probability Theory and Stochastic Modelling. Springer, Cham, 2017. doi: 10.1007/978-3-319-41598-7.
- [43] Tosio Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer, Berlin, 1995. doi: 10.1007/978-3-642-66282-9.
- [44] Alexei M. Khorunzhy, Boris A. Khoruzhenko, and Leonid A. Pastur. Asymptotic properties of large random matrices with independent entries. *J. Math. Phys.*, 37(10): 5033–5060, 1996. doi: 10.1063/1.531589.
- [45] Antti Knowles and Jun Yin. Eigenvector distribution of Wigner matrices. *Probab. Theory Related Fields*, 155(3-4):543–582, 2013. doi: 10.1007/s00440-011-0407-y.
- [46] Benjamin Landon and Horng-Tzer Yau. Convergence of local statistics of Dyson Brownian motion. *Comm. Math. Phys.*, 355(3):949–1000, 2017. doi: 10.1007/s00220-017-2955-1.
- [47] Benjamin Landon, Philippe Sosoe, and Horng-Tzer Yau. Fixed energy universality of Dyson Brownian motion. *Adv. Math.*, 346:1137–1332, 2019. doi: 10.1016/j.aim.2019.02.010.
- [48] Benjamin Landon, Patrick Lopatto, and Jake Marcinek. Comparison theorem for some extremal eigenvalue statistics. *Ann. Probab.*, 48(6):2894–2919, 2020. doi: 10.1214/20-AOP1439.
- [49] G. Le Caër, C. Male, and R. Delannay. Nearest-neighbour spacing distributions of the β -Hermite ensemble of random matrices. *Phys. A*, 383(2):190–208, 2007. doi: 10.1016/j.physa.2007.04.057.
- [50] Chi-Kwong Li and Ren-Cang Li. A note on eigenvalues of perturbed Hermitian matrices. *Linear Algebra Appl.*, 395:183–190, 2005. doi: 10.1016/j.laa.2004.08.026.
- [51] Patrick Lopatto and Kyle Luh. Tail bounds for gaps between eigenvalues of sparse random matrices. *Electron. J. Probab.*, 26:1–26, 2021. doi: 10.1214/21-EJP669.
- [52] Patrick Lopatto and Matthew Meeker. Smallest gaps between eigenvalues of real Gaussian matrices. *Bernoulli*, 32(1):325–349, 2026. doi: 10.3150/25-BEJ1860.
- [53] Kyle Luh and Van Vu. Sparse random matrices have simple spectrum. *Ann. Inst. Henri Poincaré Probab. Stat.*, 56(4):2307–2328, 2020. doi: 10.1214/19-AIHP1032.
- [54] A. Lytova and L. Pastur. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.*, 37(5):1778–1840, 2009. doi: 10.1214/09-AOP452.

- [55] Roy Mathias. Quadratic residual bounds for the Hermitian eigenvalue problem. *SIAM J. Matrix Anal. Appl.*, 19(2):541–550, 1998. doi: 10.1137/S0895479896310536.
- [56] Madan Lal Mehta. *Random Matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.
- [57] Hoi Nguyen, Terence Tao, and Van Vu. Random matrices: tail bounds for gaps between eigenvalues. *Probab. Theory Related Fields*, 167(3-4):777–816, 2017. doi: 10.1007/s00440-016-0693-5.
- [58] Valentin V. Petrov. *Sums of Independent Random Variables*. Springer-Verlag, Berlin–Heidelberg–New York, 1975. doi: 10.1007/978-3-642-65809-9.
- [59] Christian W. Pfrang, Percy Deift, and Govind Menon. How long does it take to compute the eigenvalues of a random symmetric matrix? In Percy Deift and Peter Forrester, editors, *Random Matrix Theory, Interacting Particle Systems, and Integrable Systems*, volume 65 of *Math. Sci. Res. Inst. Publ.*, pages 411–442. Cambridge University Press, New York, 2014. doi: 10.1017/9781139942287.017.
- [60] Mark Rudelson and Roman Vershynin. The Littlewood–Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008. doi: 10.1016/j.aim.2008.01.010.
- [61] L. A. Shepp. A local limit theorem. *Ann. Math. Statist.*, 35(1):419–423, 1964. doi: 10.1214/aoms/1177703766.
- [62] Dai Shi and Yunjiang Jiang. Smallest gaps between eigenvalues of random matrices with complex Ginibre, Wishart and universal unitary ensembles, 2012. Preprint, arXiv:1207.4240.
- [63] Marek Smaczyński, Tomasz Tkocz, Marek Kuś, and Karol Życzkowski. Extremal spacings between eigenphases of random unitary matrices and their tensor products. *Phys. Rev. E*, 88(5):052902, 2013. doi: 10.1103/PhysRevE.88.052902.
- [64] Zeyan Song and Hanchao Wang. The eigenvalue gap of inhomogeneous symmetric discrete random matrix, 2025. Preprint, arXiv:2503.08139.
- [65] Alexander Soshnikov. Statistics of extreme spacings in determinantal random point processes. *Mosc. Math. J.*, 5(3):705–719, 2005. doi: 10.17323/1609-4514-2005-5-3-705-719.
- [66] H.-J. Stöckmann and J. Stein. “Quantum” chaos in billiards studied by microwave absorption. *Phys. Rev. Lett.*, 64(19):2215–2218, 1990. doi: 10.1103/PhysRevLett.64.2215.
- [67] Charles Stone. A local limit theorem for nonlattice multi-dimensional distribution functions. *Ann. Math. Statist.*, 36(2):546–551, 1965. doi: 10.1214/aoms/1177700165.
- [68] Gábor Szegő. *Orthogonal Polynomials*, volume XXIII of *Amer. Math. Soc. Colloq. Publ.*. Amer. Math. Soc., Providence, RI, fourth edition, 1975.

- [69] Terence Tao. The asymptotic distribution of a single eigenvalue gap of a Wigner matrix. *Probab. Theory Related Fields*, 157(1-2):81–106, 2013. doi: 10.1007/s00440-012-0450-3.
- [70] Terence Tao and Van Vu. Random matrices: universality of local eigenvalue statistics. *Acta Math.*, 206(1):127–204, 2011. doi: 10.1007/s11511-011-0061-3.
- [71] Terence Tao and Van Vu. Random matrices have simple spectrum. *Combinatorica*, 37(3):539–553, 2017. doi: 10.1007/s00493-016-3363-4.
- [72] Craig A. Tracy and Harold Widom. Correlation functions, cluster functions, and spacing distributions for random matrices. *J. Stat. Phys.*, 92(5-6):809–835, 1998. doi: 10.1023/A:1023084324803.
- [73] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, 2012. doi: 10.1007/s10208-011-9099-z.
- [74] Roman Vershynin. Invertibility of symmetric random matrices. *Random Structures Algorithms*, 44(2):135–182, 2014. doi: 10.1002/rsa.20429.
- [75] Jade P. Vinson. *Closest spacing of eigenvalues*. PhD thesis, Princeton University, 2001.
- [76] J. von Neumann and E. P. Wigner. Über das Verhalten von Eigenwerten bei adiabatischen Prozessen. *Phys. Z.*, 30:467–470, 1929.
- [77] Per von Soosten and Simone Warzel. The phase transition in the ultrametric ensemble and local stability of Dyson Brownian motion. *Electron. J. Probab.*, 23:Paper No. 70, 1–24, 2018. doi: 10.1214/18-EJP197.
- [78] Eugene P. Wigner. On the statistical distribution of the widths and spacings of nuclear resonance levels. *Proc. Cambridge Philos. Soc.*, 47(4):790–798, 1951. doi: 10.1017/S0305004100027237.
- [79] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62(3):548–564, 1955. doi: 10.2307/1970079.
- [80] Eugene P. Wigner. Results and theory of resonance absorption. In *Conference on Neutron Physics by Time-of-Flight (Gatlinburg, Tennessee, November 1–2, 1956)*, number ORNL-2309 in Oak Ridge National Laboratory Report, pages 59–70. Oak Ridge National Laboratory, Oak Ridge, TN, 1957.
- [81] Eugene P. Wigner. Random matrices in physics. *SIAM Rev.*, 9(1):1–23, 1967. doi: 10.1137/1009001.
- [82] Albert Zhang. Quantitative gap universality for Wigner matrices. *Electron. J. Probab.*, 31:1–22, 2026. doi: 10.1214/26-EJP1601. Conjecture and equation numbers refer to arXiv:2507.20442v2.