Eigenvector delocalization in quantum chaos and random matrix theory

Laura Shou

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Mathematics
Adviser: Ramon van Handel

May 2022

Abstract

This thesis consists of two parts concerning delocalization of eigenvectors: the behavior of eigenvectors associated with quantum graphs from classically ergodic interval maps, and a delocalization-localization transition in structured random matrices.

In the first part, we prove an analogue of the pointwise Weyl law for eigenvectors of famillies of unitary matrices obtained from quantization of one-dimensional interval maps. This quantization for interval maps was introduced by Pakoński *et al.* [J. Phys. A 34 9303 (2001)] as a model for quantum chaos on graphs. We allow shrinking spectral windows in the pointwise Weyl law analogue, which allows for a strengthening of the quantum ergodic theorem for these models, and also allows for construction of randomly perturbed quantizations that have approximately Gaussian eigenvectors in the semiclassical limit.

The second part is concerned with a localization-delocalization transition for structured random matrices associated with d-regular graphs. This model includes both sparse and non-sparse Gaussian matrices with $1 \ll d \leq N$ nonzero entries in each row or column, such as random band matrices, as well as various models of interest in computer science and combinatorics. For such matrices, Bandeira and van Handel [Ann. Probab. 44 2479 (2016)] showed that the norm undergoes a phase transition at $d \sim \log N$. This transition cannot in general be captured by localization or delocalization of the top eigenvectors, but we show that the transition is captured instead by a localization-delocalization transition of approximate top eigenvectors.

Acknowledgements

First and foremost I would like to thank my advisor, Ramon van Handel, for taking me as a PhD student and for everything he has taught me about mathematics and research. Ramon introduced me to probability and quantum chaos, and was always very patient, and generous with his time. I am thankful for all his guidance, insight, and support over the past several years, and am very grateful and fortunate to have had the opportunity to work with him for my PhD.

I would like to thank the rest of my thesis committee, Jeffrey Schenker as reader, and Michael Aizenman and Peter Sarnak as committee members, for their generous time and support, and for helpful and interesting mathematical discussions on random matrices, statistical physics, and quantum chaos.

I would also like to thank Jill LeClair and the rest of the math department staff for their conversations and help throughout my time in the department.

I've been fortunate to have many math grad student friends here with whom I've made many fun memories. A sample, biased towards those I spent the most time with, includes: Amina, Ben, Boya, Fabian, Jia, Jonathan, Konstantinos, Lena, Leo, Mohan, Sang Woo, Thomas, and Yangyang. This list is certainly not complete, and I would like to express my gratitude to all of my math grad student friends here for their friendship and the time we spent together. I am also thankful for my friends from other departments and universities, including Ben, Liane, Olga, Zoe, Luther, Mari, Skye, Oli, Bella, Chris, Eric, Hannah, Matt, and RJ (who are listed alphabetically in a few groupings), as well as for the people I've met at math physics conferences and through playing percussion here.

Inevitably, I have not been able to list everyone who has been involved in my life during my graduate studies, and so here I would like to extend my thanks and appreciation to those people involved in my life whom I have not already mentioned above.

Finally, I would like to thank my family for their unconditional love and support throughout my life. They have always been there for me, and they have done more than I can ever thank them for.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-2039656.

Contents

CHAPTER 1

Overview

This thesis is concerned with eigenvector properties in two different models, and in particular with whether eigenvectors are *delocalized* or *localized*. Heuristically, a delocalized eigenvector is one whose mass is spread roughly evenly throughout its coordinates, while a localized vector is one which has much of its mass concentrated on relatively few coordinates. As will be discussed further in this thesis, localization or delocalization of eigenvectors has important implications in quantum systems and condensed matter physics, as well as applications in combinatorics and computer science.

This thesis is comprised of two parts. Part 1 is adapted from [S21] `arxiv.org/abs/2110.15301`. It is concerned with quantum chaos and the eigenvectors of unitary matrices that are obtained by quantizing classically ergodic 1D interval maps. These matrices numerically display quantum chaotic behavior, in particular appearing to have eigenvalues and eigenvectors like those of circular unitary ensemble (CUE) matrices, despite that they can be non-random and have a very simple and sparse structure. Motivated by these observations, we will study the eigenvectors by proving a *pointwise Weyl law*, which has implications for quantum ergodicity as well as for constructing random quantizations with approximately Gaussian eigenvectors.

Part 2 of this thesis is joint work with Ramon van Handel, and involves a localization-delocalization transition for structured random matrices. These are $N \times N$ symmetric matrices X_N with d iid (modulo symmetry) Gaussian entries in each row and column, and zeros everywhere else. Equivalently, one starts with a non-random d-regular graph on N vertices, and constructs the matrix X_N by placing iid standard Gaussian variables on the nonzero entries in the adjacency matrix of the graph, modulo symmetry. For such matrices, Bandeira and van Handel showed in [BvH16] that the norm undergoes a phase transition at $d \sim \log N$.

We will show that while this transition is in general not captured by the localization or delocalization of the top eigenvectors, it is instead captured by a localization-delocalization transition of *approximate* top eigenvectors, where by approximate top eigenvector we will mean a unit vector v with $\|X_N v\|_2$ close to the maximum possible value $\|X_N\|$.

Both parts of this thesis utilize projection matrix estimates to obtain eigenvector properties. In Part 1, we approximate the projection matrix of a unitary matrix U_n using a Fourier series, which allows us to relate properties of powers of U_n to properties of the projection matrix. In Part 2, we approximate the projection matrix of a real symmetric matrix using the Poisson kernel and resolvents, which allows us to use a local semicircle law to prove the projection matrix estimates.

The estimates on the projection matrix entries provide information about the structure of the subspace spanned by the corresponding eigenvectors. If one takes a unit vector chosen uniformly at random (according to Haar measure) from this subspace, then for large dimensions, Gaussian concentration ensures that it looks like a multivariate Gaussian whose covariance matrix is just the orthogonal projection matrix times $1/n$. We will use this in Part 1 to construct random quantizations with approximately Gaussian eigenvectors, and in Part 2 to show the existence of a delocalized approximate top eigenvector.

1.1. Quantum chaos on graphs

In Part 1 of this thesis, we will consider a quantization method for certain ergodic piecewise-linear 1D interval maps $S : [0,1] \to [0,1]$, introduced by Pakoński, Życzkowski, and Kuś in [PZK01] as a model for quantum chaos on graphs. Precise conditions for these interval maps will be described in Section 2.2, but for simplicity, one can consider just the doubling map, $S(x) = 2x \pmod 1$. The quantization method associates to S a family of unitary matrices U_n, where U_n is a size $n \times n$ matrix, and $n \in \mathbb{N}$ is taken in a subset of allowable dimensions. These unitary matrices describe quantum dynamics on a directed graph, and are considered "quantizations" in the sense they satisfy a classical-quantum correspondence principle (Egorov theorem) as the dimension $n \to \infty$. For the doubling map, for $n \in 2\mathbb{Z}$, one can take the quantizations to be the $n \times n$ matrices

$$U_n = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 & & & & \\ & & 1 & -1 & & \\ & & & & \ddots & \\ & & & & & 1 \;\; -1 \\ 1 & 1 & & & & \\ & & 1 & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \;\; 1 \end{pmatrix}.$$

For large n, surprisingly this non-random unitary matrix U_n tends to have level spacings that numerically look Wigner–Dyson, as well as eigenvector coordinates that numerically look Gaussian (Figures 2.1 and 2.2), despite its simple, sparse structure uncharacteristic of a typical CUE Haar unitary matrix. This behavior is however consistent with major open conjectures in quantum chaos, that quantum systems corresponding to classically chaotic ones should exhibit random matrix ensemble spectral statistics (BGS conjecture [BGS84]) and have eigenvectors that behave like Gaussian random waves [Ber77] in the semiclassical limit.

Motivated by the above, we will study the eigenvectors of such unitary quantizations U_n constructed from allowable interval maps by proving a *pointwise Weyl law*, which consists of estimates on the diagonal elements of spectral projection matrices. Because we allow shrinking spectral windows, this will let us construct randomly perturbed quantizations with eigenvectors that look Gaussian, and also obtain a strengthening of the quantum ergodic theorem for these models.

For n in a set of allowable dimensions, and given the $n \times n$ unitary matrices U_n obtained as a quantization of an appropriate interval map, denote the eigenvalues and eigenvectors of U_n by $(e^{i\theta^{(n,j)}})_j$ and $(\psi^{(n,j)})_j$ respectively. We will prove the following for spectral projections onto shrinking arcs on the unit circle, which will be stated more precisely in Part 1 as Theorem 2.1. Note although the spectral windows $I(n)$ may shrink, they are not allowed to shrink too fast, and in particular they will generally need to satisfy a condition like $|I(n)| \log n \to \infty$, related to an Ehrenfest time.

Theorem 1.1 (projection matrix estimates/pointwise Weyl law for U_n)**.** *Let $(I(n))$ be a sequence of intervals in $\mathbb{R}/(2\pi\mathbb{Z})$ that is allowed to shrink at a specific rate, and let $P^{I(n)}$*

be the orthogonal projection onto $\mathrm{span}\{\psi^{(n,j)} : \theta^{(n,j)} \in I(n)\}$*. Then for at least* $n(1-o(1))$ *coordinates* $x \in [n]$,

$$(P^{I(n)})_{xx} \equiv \sum_{j:\theta^{(n,j)}\in I(n)} |\psi_x^{(n,j)}|^2 = \frac{|I(n)|}{2\pi}(1+o(1)), \tag{1.1.1}$$

as $n \to \infty$*, for allowable dimensions* $n \in \mathbb{N}$.

This also implies a Weyl law analogue, which counts the number of eigenvalues in a bin $I(n)$,

$$\#\{j : \theta^{(n,j)} \in I(n)\} = n\frac{|I(n)|}{2\pi}(1+o(1)).$$

We will use the projection matrix estimates (1.1.1) to construct small random perturbations of the original matrices U_n by randomly rotating eigenvectors within each shrinking bin $I(n)$. This produces a family of matrices $V_n(\beta)$ with approximately Gaussian coordinates in the semiclassical limit $n \to \infty$. These matrices $V_n(\beta)$ still satisfy a classical-quantum correspondence principle, and so in this sense can still be considered a quantization for the classical ergodic dynamics. This provides an example of a quantization, in this sense, of a classically ergodic system, whose eigenvectors have approximately Gaussian coordinate statistics.

Theorem 1.2. *The random matrices* $V_n(\beta)$ *satisfy a classical-quantum correspondence property, and with high probability, have empirical coordinate distributions that look Gaussian as* $n \to \infty$*.*

We also use the projection matrix estimates directly to prove a stronger version of the quantum ergodic theorem in this model. As will be explained more precisely in Part 1, a quantum ergodic theorem ensures equidistribution of eigenvector coordinates for a limiting density one set of eigenvectors. It however allows for an exceptional limiting density zero set of eigenvectors that may not equidistribute over their coordinates. Using (1.1.1), we will strengthen the quantum ergodic theorem proved for this model in [BKS07], to hold over a limiting density one set within the shrinking sets $I(n)$, which by the Weyl law contain only a limiting density zero set of eigenvectors. This ensures that the original set of exceptional eigenvectors that may not equidistribute cannot accumulate too strongly in one region $I(n)$ of the unit circle.

Theorem 1.3. *Quantum ergodicity holds in a limiting density 1 set within shrinking bins.*

1.2. Structured random matrices

In Part 2 of this thesis, we consider symmetric random matrices X_N obtained from (non-random) d-regular graphs. These matrices can be defined via $(X_N)_{ij} = \delta_{i\sim j} g_{ij}$, where g_{ij} are iid standard Gaussian variables modulo symmetry, and $i \sim j$ indicates that nodes i and j are connected by an edge in the graph. Since the Gaussian entries are allowed to be arranged in any such fixed (non-random) structure, these matrices are nonhomogeneous random matrices. One notable example included in this model is 1D random band matrices, which are of particular interest in mathematical physics in connection with random Schrödinger operators.

We are interested in identifying the phase transition at $d \sim \log N$ using delocalization properties of approximate top eigenvectors. Intuitively, for small enough d, one expects the top eigenvectors to localize on large outliers, while for very large d, one expects them to delocalize across many coordinates like the eigenvectors of Gaussian orthogonal ensemble (GOE) matrices. However, as we will see, this transition is in general not captured by the localization or delocalization of the top eigenvectors, but rather by that of approximate top eigenvectors.

As there are many different notions and properties of delocalization, we will define precisely the notion we use in Section 6.1. Roughly speaking, a delocalized vector will be one that does not have a constant fraction of the mass accumulate on just $o(N)$ coordinates, and otherwise the vector will be called localized. For $d \ll \log N$, we will prove localization in this sense using Gaussian concentration and suprema bounds. For $d \gg \log N$, we will use projection matrix estimates to guarantee that we can find a delocalized approximate top eigenvector. Specifically, for a sequence $\varepsilon_N \to 0$, we will look at the orthogonal projection of the $N \times N$ matrix $X_N/\sqrt{d}$ onto the interval $[2-\varepsilon_N, b]$ for a fixed $b > 2$, which will be the projection onto the top roughly $\mathcal{O}(N\varepsilon_N^{3/2}) = o(N)$ eigenvectors. The following theorem will appear in Part 2 as Theorem 9.8.

Theorem 1.4 (projection matrix estimates)**.** *Let $d \gg \log N$ and let $P_{[a,b]}$ denote the projection matrix of $X_N/\sqrt{d}$ onto the interval $[a,b]$. Fix $b > 2$. Then there is a sequence $\varepsilon_N \to 0$ so that with probability at least $1 - o(1)$, the matrix $P_{[2-\varepsilon_N,b]}$ has diagonal elements*

$$(P_{[2-\varepsilon_N,b]})_{xx} = \frac{2}{3\pi}\varepsilon_N^{3/2}(1 + o(1)).$$

Considering random rotations within the subspace, the estimates on $(P_{[2-\varepsilon_N,b]})_{xx}$ can be used to describe the expected ℓ^q norms of a randomly chosen vector for large N. In particular, by considering large q, we can infer the existence of a vector in the subspace with good delocalization properties. Combining with the localization statement, this results in the following theorem, which is the main result of Part 2, and will be stated precisely as Theorem 6.2.

Theorem 1.5. *There is a localization-delocalization transition of approximate top eigenvectors at $d \sim \log N$. Informally, with high probability, for $d \ll \log N$, all top eigenvectors and $(1-\varepsilon)$-approximate top eigenvectors are localized, while for $d \gg \log N$, there exists a delocalized $(1-o(1))$-approximate top eigenvector.*

Part 1

Pointwise Weyl law for graphs from quantized interval maps

CHAPTER 2

Introduction, set-up, and main results

2.1. Introduction

In quantum systems, the eigenvectors and spectrum of the Hamiltonian capture the physical behavior of quantum particles in the system. The eigenfunction ψ defines a probability density $|\psi|^2$, which describes how likely the particle is to be found in a certain region. Of particular interest is determining whether eigenfunctions are *localized* in one area of space, or *delocalized* and spread throughout the system. The former is associated with insulating behavior, while the latter is associated with transport and metal-like behavior.

One specific example of generally delocalized eigenfunctions comes from systems that are classically ergodic, such as ergodic billiards or geodesic flow on negatively curved compact Riemannian manifolds. The correspondence principle from quantum mechanics suggests this classical ergodic behavior should manifest itself in the associated quantum system in the high-energy, semiclassical limit. For geodesic flow φ_t on manifolds, the associated quantum Hamiltonian is the Laplacian, and the correspondence is given by the quantum ergodic theorem of Shnirelman–Zelditch–de-Verdière [Shn74, Zel87, dV85]. For φ_t ergodic, this guarantees, in the large eigenvalue limit, a density 1 subsequence of Laplace eigenfunctions that equidistribute in all of phase space.

One also expects the spectrum of quantum Hamiltonians associated with classically chaotic systems to look like that of a random matrix ensemble, a relationship first utilized for heavy nuclei by Wigner in the 1950s, and conjectured to hold for any sufficiently chaotic[1] system by Bohigas, Giannoni, and Schmit [BGS84]. In view of this BGS conjecture, simpler systems such as quantum graphs have been used to investigate quantum chaotic

[1] We will not address the ergodic hierarchy or definition of chaotic systems here, but refer the reader to the textbook [CM06], e.g. Appendix C. We also note that in the special case of arithmetic hyperbolic surfaces there are exceptions (counterexamples) to the BGS conjecture, see [LS94, BGGS97].

behavior of both spectral and eigenvector statistics. While quantum graphs have long been used as models of idealized one-dimensional structures in physics, their use as simplified models for studying complex phenomena such as Anderson localization and quantum chaos is more recent to the last several decades [BK10].

The first evidence for quantum chaotic behavior in quantum graphs was given by Kottos and Smilansky [KS97, KS99], who showed numerically that the spectral statistics of certain families of quantum graphs behave like those of a random matrix ensemble. Further results regarding convergence of spectral statistics to those of random matrix theory include [Tan01, BSW02, BSW03, GA04, GA05], among others.

In this part, we look at unitary operators on sequences of graphs obtained from piecewise linear interval maps as constructed in [PZK01]. Given a (Lebesgue) measure-preserving map $S : [0,1] \to [0,1]$ satisfying a number of conditions described in Section 2.2, one obtains a sequence of graphs by partitioning $[0,1]$ into n equal atoms, and defining a Markov transition matrix P_n based on where S sends each atom. A *quantization* of the classical map S will be a family of $n \times n$ unitary matrices that recover the classical dynamics in the limit as an effective semiclassical parameter, in this case the reciprocal of the dimension, $1/n$, tends to zero. The quantization method used in [PZK01, BKS07] applies to *unistochastic* Markov matrices P_n, which are matrices P_n for which there is a unitary matrix U_n with the entrywise relation[2] $|(U_n)_{xy}|^2 = (P_n)_{xy}$. The matrices U_n are a quantization of the classical dynamics described by P_n, in the sense that they satisfy a correspondence principle (Egorov theorem, [BKS07]) that relates unitary evolution under U_n to the map S as $n \to \infty$. Physically, these unitary matrices are related to wave propagation and scattering in the graphs.

As investigated in [PZK01, Tan00, Tan01], if the graphs correspond to classically chaotic systems, then the spectral properties of these matrices U_n appear to behave like those of CUE random matrices as $n \to \infty$. As for eigenvector statistics, quantum ergodicity for these graphs with classically ergodic S was proved by Berkolaiko, Keating, and Smilansky

[2] Note such a relation does not uniquely define U_n if it exists, as one can always add additional phases without changing unitarity or the entrywise relation $|(U_n)_{xy}|^2 = (P_n)_{xy}$. For example, given any $\Phi \in [0, 2\pi)^n$ and defining the diagonal matrix $e^{i\Phi} := \mathrm{diag}(e^{i\Phi_1}, \ldots, e^{i\Phi_n})$, then $e^{i\Phi} U_n$ is also unitary and satisfies the same entrywise norm-squared relation.

in [BKS07]. They showed that in the large dimension limit, nearly all eigenvectors of U_n equidistribute over their coordinates: for sequences of allowable dimensions n, there is a sequence of sets $\Lambda_n \subseteq [n] := \{1, \ldots, n\}$ with $\lim_{n\to\infty} \frac{\#\Lambda_n}{n} = 1$ so that for all sequences $(j_n)_n$ with $j_n \in \Lambda_n$, and appropriate quantum observables $O_n(\phi)$,

$$\lim_{n\to\infty} \langle \psi^{(n,j_n)}, O_n(\phi)\psi^{(n,j_n)} \rangle = \int_0^1 \phi(x)\, dx, \tag{2.1.1}$$

where $\psi^{(n,j)}$ is the jth eigenvector of U_n. This is the analogue for these graphs of the Shnirelman–Zelditch–de-Verdière quantum ergodic theorem, which was originally stated for ergodic flows on compact Riemannian manifolds. Quantum ergodicity has also been extended to other settings such as torus maps [BD96, KR00, KR01, MO05, Zel97] and other graphs [AS19, AL15, Ana17]; see also [Ana18] for an overview and additional references.

In addition to the equidistribution from the quantum ergodic theorem, eigenfunctions from a classically ergodic system are expected to follow Berry's random wave conjecture [Ber77], which asserts that the eigenfunctions should behave like Gaussian random waves in the large eigenvalue limit. For graphs, instead of the large eigenvalue limit, one considers as usual the large dimension limit. In this limit, [GKP08, GKP10] used supersymmetry methods to study the eigenfunction statistics for quantum graphs in view of the random wave conjecture.

In the specific discrete models from interval maps that we consider, one expects that the empirical distribution of the coordinates $\{\psi_x^{(n,j)}\}_{x=1}^n$ of an eigenvector of U_n should behave like a random complex Gaussian $N_{\mathbb{C}}(0, \frac{1}{n})$ for most eigenvectors. This is consistent with both the random matrix ensemble behavior and the random wave conjecture. As an example, take the simplest allowable interval map, the doubling map (drawn in Figure 2.3),

$$T(x) = \begin{cases} 2x, & 0 \le x < \frac{1}{2} \\ 2x - 1, & \frac{1}{2} \le x \le 1 \end{cases},$$

which is ergodic. For $n \in 2\mathbb{N}$, the Markov matrices P_n along with a particularly simple unitary quantization U_n, are

$$P_n = \frac{1}{2}\begin{pmatrix} 1\ 1 & & & \\ & 1\ 1 & & \\ & & \ddots & \\ & & & 1\ 1 \\ 1\ 1 & & & \\ & 1\ 1 & & \\ & & \ddots & \\ & & & 1\ 1 \end{pmatrix}, \qquad U_n = \frac{1}{\sqrt{2}}\begin{pmatrix} 1\ -1 & & & \\ & 1\ -1 & & \\ & & \ddots & \\ & & & 1\ -1 \\ 1\ \ 1 & & & \\ & 1\ \ 1 & & \\ & & \ddots & \\ & & & 1\ \ 1 \end{pmatrix}. \tag{2.1.2}$$

Numerically, for large n not a power of 2, the eigenvalues of the U_n above appear to have CUE-like level statistics (Figure 2.1), despite the U_n being non-random and having a simple, sparse structure. We note that we must exclude the special case $n = 2^K$, as for these dimensions, the spectrum of this particular U_n is degenerate, cf. Section 4.2. However, with a different choice of quantization, the level spacings still appear to look CUE as in Figure 2.1. See also [PZK01, Tan00, Tan01] for additional spectral statistics.

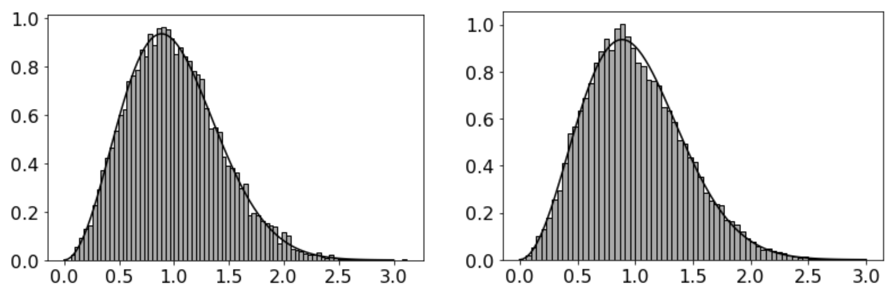

FIGURE 2.1. Left: Level spacings for the eigenvalues of U_n in (2.1.2), for $n = 25\,000$. Right: Level spacings for the eigenvalues of the matrix $e^{i\Phi}U_n$, for $n = 2^{14} = 16\,384$ and $e^{i\Phi} = \operatorname{diag}(e^{i\Phi_1}, \dots, e^{i\Phi_n})$, for a randomly chosen $\Phi \in [0, 2\pi)^n$. For both plots, the histogram is of the angle differences of the eigenvalues scaled by $\frac{n}{2\pi}$, and the solid curve is the Wigner GUE surmise[3] $p(s) = \frac{32}{\pi^2}s^2 e^{-4s^2/\pi}$.

Additionally, numerically the vast majority of eigenvectors of these U_n have coordinates that look like a complex Gaussian $N_{\mathbb{C}}(0, \frac{1}{n})$. Typical histograms for the coordinates of an eigenvector of U_n from (2.1.2) are shown in Figure 2.2.

While we do not prove Gaussian behavior for the eigenvectors of these U_n (except in a very special case where the eigenspaces end up highly degenerate, see Section 4.2), as an

[3] While the Wigner surmise is not the exact level spacing for GUE or CUE matrices, it is generally a quite adequate approximation for this type of purpose. For the actual level spacing distributions and comparison with Wigner's surmise, see [Meh04, HGK18].

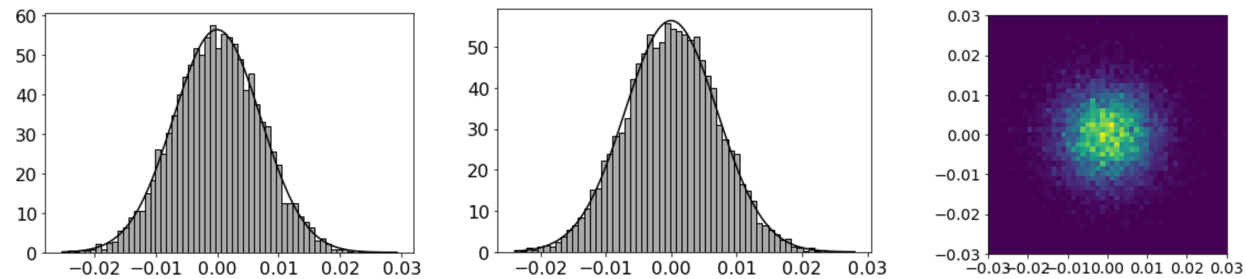

FIGURE 2.2. Plots for a randomly chosen eigenvector ψ (this one with eigenvalue $-0.3061126 + 0.9519953i$) for $n = 10\,000$ and U_n in (2.1.2). Left: Histogram of the values $(\mathfrak{Re}\,\psi_x)_{x=1}^{10\,000}$ plotted against the pdf of the real Gaussian $N(0, \frac{1}{20\,000})$. Center: Histogram of the values $(\mathfrak{Im}\,\psi_x)_{x=1}^{10\,000}$ plotted against the pdf of $N(0, \frac{1}{20\,000})$. Right: 2D histogram in $\mathbb{C}$ of the coordinates $(\psi_x)_{x=1}^{10\,000}$. Since this is fairly spherically symmetric, the overall choice of phase for the eigenvector does not significantly impact the shape of the other plots.

application of our main result we will prove that for allowable S, there are many random quantizations V_n with approximately Gaussian eigenvector coordinates. These quantizations are not quantizations in the strict sense of $|(V_n)_{xy}|^2 = (P_n)_{xy}$ from [PZK01, BKS07], but they will satisfy $|(V_n)_{xy}|^2 = (P_n)_{xy} + o(1)$ as well as an Egorov theorem, so they are still quantizations of S in the sense that they recover the classical dynamics in the semiclassical limit $n \to \infty$.

Our main result in this part of the thesis is Theorem 2.1, an analogue of the pointwise Weyl law for the eigenvectors of the matrices U_n under shrinking spectral windows, which will have implications for quantum ergodicity and for constructing random perturbations of U_n with the desired Gaussian eigenvector behavior. Traditionally, a pointwise Weyl law gives the leading order asymptotics of the spectral projection kernel $\mathbb{1}_{(-\infty,t]}(-\Delta + V)(x,x)$, for x in M a compact Riemannian manifold. For the unitary matrices U_n, we look at the spectral projection onto arcs on the unit circle, $P_I = \sum_{j:\theta^{(n,j)} \in I} |\psi^{(n,j)}\rangle\langle\psi^{(n,j)}|$ where $I \subseteq \mathbb{R}/(2\pi\mathbb{Z})$. Then a pointwise Weyl law analogue would be a statement of the form $\sum_{j:\theta^{(n,j)} \in I(n)} |\psi_x^{(n,j)}|^2 = \frac{|I(n)|}{2\pi}(1 + o(1))$ for $n \to \infty$ and appropriate intervals $I(n)$. We will show this holds for sequences of intervals $I(n)$ shrinking at certain rates, and for at least $n(1 - o(|I(n)|))$ coordinates x. The coordinates for which this statement may not hold correspond to short periodic orbits in the graphs corresponding to S.

We then present two applications of this pointwise Weyl law. The first is a strengthening of the quantum ergodic theorem to apply to sets of eigenvectors in bins $\{\psi^{(n,j)} : \theta^{(n,j)} \in I(n)\}$ with shrinking $I(n)$. The second concerns random perturbations of the matrix U_n to produce a family of random matrices $V_n(\beta^{[n]})$ whose eigenvectors have the approximately Gaussian $N_{\mathbb{C}}(0, \frac{1}{n})$ eigenvector statistics. These eigenvectors will also tend to satisfy a version of quantum unique ergodicity (QUE), a notion introduced by Rudnick and Sarnak in [RS94], and where all eigenvectors are considered in the limit (2.1.1), rather than just those in a sequence of limiting density one sets.

2.2. Set-up

Here we state the assumptions on the map S and matrices P_n. Let $S : [0,1] \to [0,1]$ be a piecewise-linear map that satisfies the following conditions:

(i) S is (Lebesgue) measure-preserving, $\mu(A) = \mu(S^{-1}(A))$ for any measurable set A.

(ii) There exists a partition $\mathcal{M}_0$ of $[0,1]$ into M_0 equal intervals (called atoms) $A_1, \ldots, A_{M_0}$, with S linear on each atom A_j.

(iii) The endpoints $\mathcal{E}_0 = \mathcal{E}(\mathcal{M}_0)$ of the atoms have left and right limits satisfying $\lim_{x\to e_0^\pm} S(x) \in \mathcal{E}_0$ for $e_0 \in \mathcal{E}_0$. This means the linear segments in S begin and end in the grid $\mathcal{E}_0 \times \mathcal{E}_0$. With (i) and (ii), this ensures the slope of S on each atom must be an integer. For convenience, also assume $S(e_0)$ takes one of the values of these one-sided limits.

(iv) The absolute value of the slope of S on each atom is at least two, i.e. the slope is never ± 1.

Conditions (i), (ii), and (iii) are essentially the same as in [PZK01, BKS07]. Condition (iv) is there instead of the ergodicity assumption. It allows for some non-ergodic S such as those corresponding to block matrices of various ergodic maps. Two examples of allowable ergodic S are the doubling map and the "four legs map" shown in Figure 2.3. In general, conditions for ergodicity of S would follow from results on piecewise expanding Markov maps, see for example Chapter III in the textbook [Mañ87].

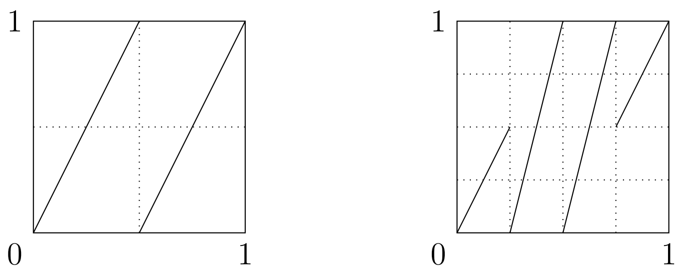

Figure 2.3. The doubling map (left) and "four legs map" (right). For the doubling map $M_0 = L_0 = 2$, while for the four legs map $M_0 = L_0 = 4$.

For $n \in M_0\mathbb{Z}$, partition $[0,1]$ into n equal atoms, $E_x = (\frac{x-1}{n}, \frac{x}{n})$ for $x = 1, \ldots, n$, and define the corresponding $n \times n$ Markov transition matrix P_n by

$$(P_n)_{xy} = \begin{cases} 0, & S(E_x) \cap E_y = \emptyset \\ \frac{1}{|S'(z)|}, & S(E_x) \cap E_y \neq \emptyset, \text{ any } z \in E_x \end{cases}. \tag{2.2.1}$$

The matrix P_n looks at where S sends an atom E_x, and assigns a uniform probability $\frac{1}{|S'(z)|}$ to each atom E_y that S can reach from E_x. To generate the family of corresponding unitary matrices U_n as done in [PZK01, BKS07], it is required that P_n be *unistochastic*, so that there are unitary matrices U_n with the entrywise relation $|(U_n)_{xy}|^2 = (P_n)_{xy}$. In general, characterizing which bistochastic matrices are unistochastic is difficult; however see [PZK01, ZSKS03, BKS07] for some conditions and examples.

Let L_0 be the least common multiple of the slopes in S, and let $\widetilde{K}(n)$ be the largest power of L_0 that divides n/M_0, so $n = M_0 L_0^{\widetilde{K}(n)} r$ and r does not contain any factors of L_0. The purpose of $\widetilde{K}(n)$ will be to keep track of how many powers of S we can take, while still ensuring S^ℓ behaves nicely with the partition into n atoms.

2.3. Main result and applications

With the above definitions, we state the main result:

Theorem 2.1 (pointwise Weyl law analogue)**.** *Let $S : [0,1] \to [0,1]$ satisfy assumptions (i)–(iv). Consider a sequence $(n_k)_k$ so that $\widetilde{K}(n_k) \to \infty$, and suppose each $n_k \times n_k$ Markov matrix P_{n_k} is unistochastic with corresponding unitary matrix U_{n_k}. Let $(I(n_k))$ be a sequence* *of intervals in* $\mathbb{R}/(2\pi\mathbb{Z})$ *satisfying*

$$|I(n_k)|\widetilde{K}(n_k) \to \infty, \quad \text{as } k \to \infty. \tag{2.3.1}$$

Then denoting the eigenvalues and eigenvectors of U_{n_k} *by* $(e^{i\theta^{(n_k,j)}})_j$ *and* $(\psi^{(n_k,j)})_j$ *respectively, there is a sequence of subsets* $G_{n_k} \subseteq [1:n_k]$ *with sizes* $\#G_{n_k} = n_k(1 - o(|I(n_k)|))$ *so that for all* $x \in G_{n_k}$*,*

$$\sum_{j:\theta^{(n_k,j)} \in I(n_k)} |\psi_x^{(n_k,j)}|^2 = \frac{|I(n_k)|}{2\pi}(1 + o(1)), \quad \text{as } k \to \infty, \tag{2.3.2}$$

where the error term $o(1)$ *depends only on* n_k*,* $|I(n_k)|$*, and* $\#G_{n_k}$*, and is independent of* $x \in G_{n_k}$*. Additionally,* G_{n_k} *can be chosen independent of* $I(n_k)$ *or* $|I(n_k)|$*.*

Remark 2.3.1. (i) The coordinates x that we exclude from G_{n_k} correspond to those with short periodic orbits in the graphs associated to P_{n_k} and U_{n_k}. This is reminiscent of the relationship between geodesic loops and the size of the remainder in the Weyl law [DG75, Ivr80] or pointwise Weyl law [Saf88, SZ02, CG20], in the usual setting on manifolds. For the sequences of coordinates that we exclude, we do not expect the leading order approximation to necessarily be $\frac{|I(n_k)|}{2\pi}$ in general, see Section 5.1.

(ii) The condition (2.3.1) that $|I(n_k)|$ does not shrink too fast appears from error terms from only considering powers of U_{n_k} up to an Ehrenfest time $\widetilde{K}(n_k) \sim \log n_k$. This time is a common obstruction in semiclassical problems, and even in these discrete models, our analysis does not go beyond this time. If the lengths $|I(n_k)|$ are larger than required to satisfy (2.3.1), then more precise remainder terms than just $o(1)$ are obtained from the proof.

The proof details of Theorem 2.1 will be specific to our discrete case, where we have sparse matrices U_{n_k} and can analyze matrix powers and paths in finite graphs. We will start by just taking a smooth approximation of the indicator function of the interval $I(n_k)$, and estimating the left side of (2.3.2) by a Fourier series in terms of powers of U_{n_k}. However, the properties of S ensure that we understand powers of U_{n_k} well up to time $\widetilde{K}(n_k)$. This allows us to identify and exclude the few coordinates x that have short loops before a set cut-off time. Using properties of the powers of U_{n_k} again, the remaining coordinates will then produce small enough Fourier coefficients that (2.3.2) holds.

Summing (2.3.2) over all x (separating $x \in G_{n_k}$ from $x \notin G_{n_k}$) produces a Weyl law analogue that counts the number of eigenvalues in a bin.

Corollary 2.2 (Weyl law analogue). *Let S, $(n_k)_k$, U_{n_k}, and $I(n_k)$ be as in Theorem 2.1, including* (2.3.1). *Then as $k \to \infty$,*

$$\#\{j : \theta^{(n_k,j)} \in I(n_k)\} = n_k \frac{|I(n_k)|}{2\pi}(1 + o(1)), \tag{2.3.3}$$

where the remainder term depends on $|I(n_k)|$ but is independent of the particular location of $I(n_k)$.

In the following subsections, we discuss implications of Theorem 2.1 on eigenvectors of U_n. We present the two applications, the first a strengthening of the quantum ergodic theorem for this model, and the second a construction of random perturbations of U_n with approximately Gaussian eigenvectors. For the first application, using Theorem 2.1 with shrinking intervals $|I(n_k)|$, rather than the usual local Weyl law, in the standard proof of quantum ergodicity naturally produces a stronger quantum ergodicity statement. For the second, we take random unitary rotations of bins of eigenvectors, and apply results on the distribution of random projections from [DF84, CM08, Mec09] to show the resulting eigenvectors have approximately Gaussian value statistics.

2.3.1. Application to quantum ergodicity in bins. To state a quantum ergodic theorem, we first define quantum observables as in [BKS07], as discretized versions of a classical observable $h \in L^2([0,1])$. Given $n \in \mathbb{N}$ and $h \in L^2([0,1])$, define its quantization $O_n(h)$ to be the $n \times n$ diagonal matrix with entries

$$O_n(h)_{xx} = \frac{1}{|E_x|} \int_{E_x} h(z)\, dz = n \int_{E_x} h(z)\, dz. \tag{2.3.4}$$

Note that $\frac{1}{n}\operatorname{tr} O_n(h) = \int_0^1 h$, the analogue of the local Weyl law. Quantum ergodicity for this model, as proved in [BKS07], states there is a sequence of sets $\Lambda_{n_k} \subseteq [1 : n_k]$ with $\lim_{n_k\to\infty} \frac{\#\Lambda_{n_k}}{n_k} = 1$ such that for all sequences $(j_{n_k})_k$, $j_{n_k} \in \Lambda_{n_k}$ and $h \in C([0,1])$,

$$\lim_{k\to\infty} \langle \psi^{(n_k,j_{n_k})}, O_{n_k}(h)\psi^{(n_k,j_{n_k})}\rangle = \int_0^1 h(x)\,dx. \tag{2.3.5}$$

This is equivalent to the decay of the quantum variance,

$$V_{n_k} := \frac{1}{n_k}\sum_{j=1}^{n_k} \left| \langle \psi^{(n_k,j)}, O_{n_k}(h)\psi^{(n_k,j)}\rangle - \int_0^1 h(x)\,dx \right|^2 \to 0,$$

as $k \to \infty$. Using Theorem 2.1 and an Egorov property from [BKS07], we will prove the following concerning quantum ergodicity in bins.

Theorem 2.3 (Quantum ergodicity in bins)**.** *Let S satisfy (i)–(iv) and also be ergodic. Let $(n_k)_k$, U_{n_k}, and $I(n_k)$ be as in Theorem 2.1, including* (2.3.1)*. Then for any Lipschitz $h : [0,1] \to \mathbb{C}$,*

$$\frac{1}{\#\{j : \theta^{(n_k,j)} \in I(n_k)\}} \sum_{j:\theta^{(n_k,j)}\in I(n_k)} \left| \langle \psi^{(n_k,j)}, O_{n_k}(h)\psi^{(n_k,j)}\rangle - \int_0^1 h(x)\,dx \right|^2 \to 0, \tag{2.3.6}$$

as $k \to \infty$.

This decay of the quantum variance in a bin implies there is a sequence of sets $\Lambda_{n_k} \subseteq \{j : \theta^{(n_k,j)} \in I(n_k)\}$ with $\frac{\#\Lambda_{n_k}}{\#\{j:\theta^{(n_k,j)}\in I(n_k)\}} \to 1$ such that (2.3.5) holds for all sequences $(j_{n_k})_k$, $j_{n_k} \in \Lambda_{n_k}$ and continuous $h : [0,1] \to \mathbb{C}$. Since we allow $|I(n)| \to 0$, the bin sizes are $o(n)$ by the Weyl law analogue, and so Theorem 2.3 guarantees that the density 0 subsequence excluded from the original quantum ergodic theorem cannot accumulate too strongly in one region $I(n)$ of the unit circle.

2.3.2. Application to random quantizations with Gaussian eigenvectors. The second application of Theorem 2.1 will be to construct random perturbations of U_n with eigenvectors that look approximately Gaussian. To construct the random perturbations, we first use results on low-dimensional projections from [DF84, Mec09, CM08] combined with estimates on the matrix entries of the spectral projection matrix to prove:

Theorem 2.4 (Gaussian approximate eigenvectors)**.** *Let S, $(n_k)_k$, U_{n_k}, and $I(n_k)$ be as in Theorem 2.1, in particular assume* (2.3.1) *holds. Denote the eigenvalues and eigenvectors of U_{n_k} by $(e^{i\theta^{(n_k,j)}})_j$ and $(\psi^{(n_k,j)})_j$ respectively. Then letting $\phi^{(n_k)}$ be a unit vector chosen randomly according to Lebesgue measure from* $\operatorname{span}(\psi^{(n_k,j)} : \theta^{(n_k,j)} \in I(n_k))$*, the empirical distribution $\mu^{(n_k)}$ of the scaled coordinates*

$$\sqrt{n_k}\phi_1^{(n_k)}, \sqrt{n_k}\phi_2^{(n_k)}, \ldots, \sqrt{n_k}\phi_{n_k}^{(n_k)},$$

converges weakly in probability to the standard complex Gaussian $N_{\mathbb{C}}(0,1)$ as $k \to \infty$. In fact, for any $f : \mathbb{C} \to \mathbb{C}$ bounded Lipschitz and $\varepsilon > 0$, there is $k_0 > 0$ so that for $k \geq k_0$,

$$\mathbb{P}\left[\left|\int f(x)\, d\mu^{(n_k)}(x) - \mathbb{E}f(Z)\right| > \varepsilon\right] \leq 6 \exp\left(-\frac{\varepsilon^2 n_k |I(n_k)|}{2^8 \pi \|f\|_{\mathrm{Lip}}^2}\right), \tag{2.3.7}$$

where $\|f\|_{\mathrm{Lip}} := \sup_{x\neq y} \frac{|f(x)-f(y)|}{|x-y|}$, and $Z \sim N_{\mathbb{C}}(0,1)$.

Then we construct random perturbations $V_{n_k}(\beta^{[n_k]})$ of U_{n_k} by binning the eigenvalues of U_{n_k} and randomly rotating the eigenvectors within each bin. This idea of rotating small sets of eigenvectors was used in different models in [Zel14, Van97, Map13, CG18] to construct random orthonormal bases with quantum ergodic or quantum unique ergodic properties. In our setting, Theorem 2.4 will additionally show that the coordinates of these randomly rotated eigenvectors look approximately Gaussian. The matrices $V_{n_k}(\beta^{[n_k]})$ also satisfy the entrywise relations $|(V_{n_k}(\beta^{[n_k]}))_{xy}|^2 = (P_{n_k})_{xy} + o(1)$ as well as a weaker Egorov property relating them to the classical dynamics, so that they can be viewed as a quantization of the classical map S. Thus while we do not prove approximate Gaussian behavior for the quantizations U_{n_k} with $|(U_{n_k})_{xy}|^2 = (P_{n_k})_{xy}$, we prove it for the family of random matrices $V_{n_k}(\beta^{[n_k]})$, which are alternative quantizations of the original classical dynamics of S.

Note also that S is not required to be ergodic here. In particular, we can take the direct sum of two ergodic maps S_1 and S_2, whose resulting block matrix U_{n_k} has eigenvectors localized on just half of the coordinates. Then U_{n_k} will not have equidistributed or Gaussian eigenvectors, though the randomly perturbed matrices $V_{n_k}(\beta^{[n_k]})$ still will.

Theorem 2.5 (Random quantizations with Gaussian eigenvectors)**.** *Let S satisfy (i)–(iv), and let $(n_k)_k$ be a sequence with $\widetilde{K}(n_k) \to \infty$ and with each Markov matrix P_{n_k} unistochastic. Then there exists a family of random unitary matrices $V_{n_k}(\beta^{[n_k]})$ in some probability spaces $(\Omega_{n_k}, \mathbb{P}_{n_k})$ with the following properties:*

(a) $V_{n_k}(\beta^{[n_k]})$ is a small perturbation of U_{n_k}, as in $\sup_{\beta^{[n_k]}} \|V_{n_k}(\beta^{[n_k]}) - U_{n_k}\| = o(1)$. Additionally, for every $\beta^{[n_k]}$, V_{n_k} satisfies an Egorov property; for Lipschitz $h : [0,1] \to \mathbb{C}$,

$$\|V_{n_k} O_{n_k}(h) V_{n_k}^{-1} - O_{n_k}(h \circ S)\| = o(1) \cdot \|h\|_{\mathrm{Lip}}.$$

(b) (Gaussian coordinates). There is a sequence of sets $\Pi_{n_k} \subseteq \Omega_{n_k}$ with $\mathbb{P}[\Pi_{n_k}] \to 1$ with the following property: Let $(\widetilde{V}_{n_k})_k$ be a sequence of matrices with $\widetilde{V}_{n_k} \in \Pi_{n_k}$, and let $\widetilde{\phi}^{[n_k,j]}$ be the jth eigenvector of $\widetilde{V}_{n_k}$, and $\mu^{[n_k,j]} = \frac{1}{n_k} \sum_{x=1}^{n_k} \delta_{\sqrt{n_k}\widetilde{\phi}_x^{[n_k,j]}}$ the empirical distribution of the scaled coordinates of $\widetilde{\phi}^{[n_k,j]}$. Then for every sequence $(j_{n_k})_k$ with $j_{n_k} \in [1 : n_k]$, the sequence $(\mu^{[n_k, j_{n_k}]})_k$ converges weakly to $N_{\mathbb{C}}(0,1)$ as $k \to \infty$.

(c) (QUE). There is a sequence of sets $\Gamma_{n_k} \subseteq \Omega_{n_k}$ with $\mathbb{P}[\Gamma_{n_k}] \to 1$ such that for any sequence of matrices $(\widetilde{V}_{n_k})_k$ with $\widetilde{V}_{n_k} \in \Gamma_{n_k}$, the eigenvectors $\widetilde{\phi}^{[n_k,j]}$ of $\widetilde{V}_{n_k}$ equidistribute over their coordinates. That is, for any sequence $(j_{n_k})_k$ with $j_{n_k} \in [1 : n_k]$, and any $h \in C([0,1])$,

$$\lim_{k\to\infty} \langle \widetilde{\phi}^{[n_k, j_{n_k}]}, O_{n_k}(h) \widetilde{\phi}^{[n_k, j_{n_k}]} \rangle = \int_0^1 h(x)\, dx. \tag{2.3.8}$$

(d) For every $\beta^{[n_k]}$, the spectrum of $V_{n_k}(\beta^{[n_k]})$ is non-degenerate.

(e) The matrix elements of $V_{n_k}(\beta^{[n_k]})$ satisfy $\sup_{\beta^{[n_k]}} \max_{x,y} \big| |V_{n_k}(\beta^{[n_k]})_{xy}|^2 - (P_{n_k})_{xy} \big| \to 0$ as $k \to \infty$.

2.3.3. The doubling map. Finally, in Sections 4.1 and 4.2 we study the case when S is the doubling map on $\mathbb{R}/\mathbb{Z}$ and the specific quantization U_n is the orthogonal one in (2.1.2). We study this case using similar arguments as in the general case, but with stronger estimates from analyzing binary trees and bit shifts specific to the doubling map. Theorem 2.1 will hold with any sequence of even $n \in 2\mathbb{N}$, not just those with $\widetilde{K}(n) \to \infty$. Additionally, when $n = 2^K$, the spectrum of this specific quantization U_n is degenerate with multiplicities asymptotically $\frac{2^K}{4K}$, and most every eigenbasis looks Gaussian (Theorem 4.6).

2.3.4. Outline. Chapter 3 contains the main proofs: Section 3.1 contains some lemmas concerning properties of the map S and the corresponding Markov matrices P_n. Section 3.2 is the proof of the pointwise Weyl law analogue, Theorem 2.1. The first application, Theorem 2.3 on quantum ergodicity in bins, is proved in Section 3.3. Section 3.4 covers the second application on random perturbations of U_n with approximately Gaussian eigenvectors. Chapter 4 deals with the specific map the doubling map, especially with the degenerate case of dimension a power of two. Chapter 5 contains additional remarks.

CHAPTER 3

Proof of the main results

3.1. Properties of the map S and matrices P_n

In this section we gather some results about the relationship between the map S and the Markov matrices P_n. The following lemma contains properties from [BKS07] and [PZK01], stated here for a specific condition involving $\widetilde{K}(n)$.

Lemma 3.1 (powers of S, [BKS07, PZK01])**.** *Assume (i)–(iii) and let the partition size be $n \in M_0\mathbb{Z}$ with atoms $E_1, \ldots, E_n$. Then for $1 \leq \ell \leq \widetilde{K}(n) + 1$,*

(a) S^ℓ is linear with integer slope on each atom E_x, and for endpoints $e \in \mathcal{E}$, the right and left limits satisfy $\lim_{y \to e^\pm} S^\ell(y) \in \mathcal{E}$.

(b) If $S^\ell(E_x) \cap E_y \neq \emptyset$, then $S^\ell(E_x) \supset E_y$. In fact $S^\ell(E_x)$ is a union of several adjacent atoms and some endpoints.

(c) $(P_n)_{x\tau_1}(P_n)_{\tau_1\tau_2} \cdots (P_n)_{\tau_{\ell-1}y} \neq 0$ iff there exists $z \in E_x$ with $S^\ell(z) \in E_y$ and $S^j(z) \in E_{\tau_j}$ for $j = 1, \ldots, \ell - 1$.

(d) If $S^\ell(E_x) \cap E_y = \emptyset$ then $(P_n^\ell)_{xy} = 0$. If $S^\ell(E_x) \cap E_y \neq \emptyset$, then there is a unique sequence $\tau = (\tau_1, \tau_2, \ldots, \tau_{\ell-1})$ such that $(P_n)_{x\tau_1}(P_n)_{\tau_1\tau_2} \cdots (P_n)_{\tau_{\ell-1}y} \neq 0$.

The condition here with $\widetilde{K}(n)$ can be more restrictive than needed in [BKS07], but is a concrete example of allowable powers ℓ and dimensions n. For completeness with these concrete conditions, we include most of the proofs below.

Proof. (a) Both parts are done recursively. For example, if $S^{\ell-1}$ is linear on atoms E_x, then for the first part of (a), it suffices to show for each x, $S^{\ell-1}(E_x) \subseteq A_j$ for one of the atoms A_j of the "base" partition $\mathcal{M}_0$ (depending on x), since then composition with S shows $S^\ell = S \circ S^{\ell-1}$ is linear on E_x. The inclusion $S^{\ell-1}(E_x) \subseteq A_j$ holds for any $\ell - 1 \leq \widetilde{K}(n)$, essentially because this image must avoid all endpoints $e_0 \in \mathcal{M}_0$.

Since L_0 is the least common multiple of the slopes of S, then $L_0^{\ell-1}$ is a multiple of the slopes of $S^{\ell-1}$, and so the preimages $S^{-(\ell-1)}(e_0)$ of a "base" partition endpoint $e_0 \in \mathcal{E}_0$ must live in the endpoints,[1] not interior, of the size $M_0 \cdot L_0^{\ell-1}$ partition.

(b) follows from (a) since the linear segments in S^ℓ start at points in $\mathcal{E}$ and have integer slopes.

(c) The ($\Leftarrow$) direction is immediate from the definition of P_n. The ($\Rightarrow$) direction follows from the relations

$$S(E_x) \supset E_{\tau_1}, \quad S(E_{\tau_1}) \supset E_{\tau_2}, \quad \dots, \quad S(E_{\tau_{\ell-1}}) \supset E_y$$

and working backwards, taking $z_{\ell-1} \in E_{\tau_{\ell-1}}$ with $S(z_{\ell-1}) \in E_y$, and then $z_j \in E_{\tau_j}$ with $S(z_j) = z_{j+1}$.

(d) The first part follows from the above inclusions as well; note that if $(P_n)_{x\tau_1}(P_n)_{\tau_1\tau_2}\cdots(P_n)_{\tau_{\ell-1}y} \neq 0$, then $S^\ell(E_x) \supset S^{\ell-1}(E_{\tau_1}) \supset \cdots \supset E_y$. The unique path part is Lemma 2 from [BKS07]: the proof is to suppose there are $z_1, z_2 \in E_x$ with $S^\ell(z_1), S^\ell(z_2) \in E_y$ but with $S^r(z_1) \in E_1$ and $S^r(z_2) \in E_2$ for some $1 < r < \ell$ and $E_1 \neq E_2$. Then pick $w \in E_y$, and by part (b), then there is $w_1 \in E_1$ with $S^{\ell-r}(w_1) = w$ and $w_2 \in E_2$ with $S^{\ell-r}(w_2) = w$. Again by (b) then there are $v_1 \neq v_2$ in E_x with $S^r(v_1) = w_1$ and $S^r(v_2) = w_2$. But then $S^\ell(v_1) = S^\ell(v_2) = w$ which contradicts S^ℓ being linear (with nonzero slope since S is measure-preserving) and injective on E_x.

□

The next lemma shows how $\widetilde{K}(n)$ is used to ensure that small powers of P_n^ℓ interact nicely with the partition of $[0, 1]$ into n atoms.

[1]If $S^{-(\ell-2)}(e_0) \subset \frac{1}{M_0 L_0^{\ell-2}}\mathbb{Z}$ and $y \in S^{-1}(S^{-(\ell-2)}(e_0))$, then $S(y) = my + b \in \frac{1}{M_0 L_0^{\ell-2}}\mathbb{Z}$ for some $m | L_0$ and $b \in \frac{1}{M_0}\mathbb{Z}$, so $y \in \frac{1}{M_0 L_0^{\ell-2} m} \subseteq \frac{1}{M_0 L_0^{\ell-1}}\mathbb{Z}$.

Lemma 3.2 (powers of P_n)**.** *Assume (i)–(iii) and let* $1 \le \ell \le \widetilde{K}(n)+1$*. Then*

$$(3.1.1) \qquad (P_n^\ell)_{xy} = \begin{cases} 0, & S^\ell(E_x) \cap E_y = \emptyset \\ \frac{1}{|(S^\ell)'(z)|}, & S^\ell(E_x) \cap E_y \neq \emptyset, \ \text{any } z \in E_x \end{cases} .$$

That is, for $1 \le \ell \le \widetilde{K}(n)+1$*, we can compute* P_n^ℓ *by drawing* S^ℓ *and applying the same procedure we used to define* P_n *from* S*.*

Proof. From Lemma 3.1(a), partitioning $[0,1]$ into $M_0 \cdot L_0^{\ell-1}$ equal atoms ensures S^ℓ is linear on each atom. Since $1 \le \ell \le \widetilde{K}(n)+1$, then $M_0 \cdot L_0^{\ell-1}$ divides n so for these ℓ, the map S^ℓ is linear on each atom of the size n partition and the value $\frac{1}{|(S^\ell)'(z)|}$ is the same for any $z \in E_x$. The matrix elements of P_n^ℓ are $(P_n^\ell)_{xy} = \sum_{\tau: x \to y} (P_n)_{x\tau_1}(P_n)_{\tau_1\tau_2} \cdots (P_n)_{\tau_{\ell-1}y}$. By Lemma 3.1(d), for $1 \le \ell \le \widetilde{K}(n)+1$ and fixed x, y, this sum over τ collapses to either zero or just a single term $(P_n)_{x\tau_1} \cdots (P_n)_{\tau_{\ell-1}y}$. If this is nonzero, then by the definition of P_n,

$$(P_n^\ell)_{xy} = \frac{1}{|S'(E_x)||S'(E_{\tau_1})||S'(E_{\tau_2})| \cdots |S'(E_{\tau_{\ell-1}})|}.$$

By Lemma 3.1(c), there exists $z \in E_x$ with $S(z) \in E_{\tau_1}, S^2(z) \in E_{\tau_2}, \ldots, S^\ell(z) \in E_y$, so

$$(P_n^\ell)_{xy} = \frac{1}{|S'(z)S'(S(z)) \cdots S'(S^{\ell-1}(z))|} = \frac{1}{|(S^\ell)'(z)|}.$$

□

The following lemma demonstrates the sparseness of the matrices P_n^ℓ for times before $\widetilde{K}(n)$. Essentially, this is because for these times, the nonzero entries of the matrix P_n^ℓ are placed by drawing S^ℓ and an $n \times n$ grid in $[0,1]^2$, and then placing a nonzero entry in each position in the grid that S^ℓ passes through. As n increases, the grid becomes finer and the (one-dimensional) graph of S^ℓ in $[0,1]^2$ cannot pass through a very large fraction of the boxes.

Lemma 3.3 (number of nonzero entries)**.** *Assume (i)–(iv) and let* $1 \le \ell \le \widetilde{K}(n)+1$*. Then the diagonal of* P_n^ℓ *contains at most* $2M_0L_0^{\ell-1}$ *nonzero entries, and in total* P_n^ℓ *has at most*

$n \cdot s_{\max}^{\ell}$ nonzero entries, where $s_{\max}$ is the maximum of the absolute values of the slopes in S.

Proof. Pick an atom E_x. Since the maximum slope magnitude of S is $s_{\max}$, the interval $S^{\ell}(E_x)$ has length at most $s_{\max}^{\ell} \cdot |E_x|$ and intersects at most $s_{\max}^{\ell}$ atoms E_y. Thus by Lemma 3.2 the xth row of P_n^{ℓ} has at most $s_{\max}^{\ell}$ nonzero entries, so in total P_n^{ℓ} has at most $n \cdot s_{\max}^{\ell}$ nonzero entries.

Also by Lemma 3.2, nonzero diagonal elements of P_n^{ℓ} occur exactly when $S^{\ell}(E_x) \cap E_x \neq \emptyset$. Let $Q \subset [0,1] \times [0,1]$ be the diagonal chain of squares $Q = \bigcup_{x=0}^{n-1} (\frac{x}{n}, \frac{x+1}{n}) \times (\frac{x}{n}, \frac{x+1}{n})$, so that the nonzero diagonal elements $(P_n^{\ell})_{xx}$ occur exactly when S^{ℓ} intersects the xth square $(\frac{x}{n}, \frac{x+1}{n}) \times (\frac{x}{n}, \frac{x+1}{n})$ (Figure 3.1).

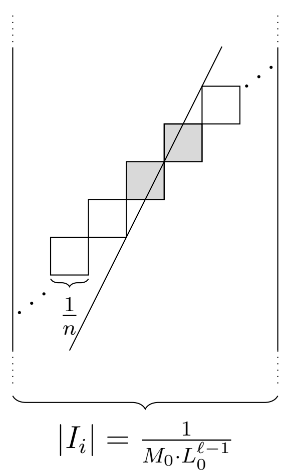

FIGURE 3.1. A line of slope 2 intersecting two (open) boxes in the diagonal Q.

Choose an interval $I_i := (\frac{i}{M_0 L_0^{\ell-1}}, \frac{i+1}{M_0 L_0^{\ell-1}})$, an atom of the partition into $M_0 L_0^{\ell-1}$ atoms. This is the coarsest partition for S^{ℓ} for which we can guarantee by Lemma 3.1(a) that S^{ℓ} is linear on each atom. Since $\ell \leq \widetilde{K}(n) + 1$, the partition into n atoms E_x is a refinement of this one. If the slope of S^{ℓ} is negative on I_i, then S^{ℓ} can intersect at most one box in the diagonal Q. If the slope of S^{ℓ} is positive and at least two on I_i, then it can intersect at most two boxes in Q (see Figure 3.1): Consider the slope one lines $t \pm \frac{1}{n}$ in $[0,1]^2$, which bound a parallelogram $R \supset Q$. Project the line segment $S^{\ell}(I_i) \cap R$ onto the x-axis. If S^{ℓ} on I_i has slope $m > 1$, then one can compute this projection is an interval of length $\leq \frac{1}{n}\frac{2}{m-1}$. For

$m \geq 3$, this bound is $\leq \frac{1}{n}$ so $S^\ell(I_i) \cap R$ can intersect at most two $\frac{1}{n} \times \frac{1}{n}$ boxes in Q. For $m = 2$, the length can be $\frac{2}{n}$, but $S^\ell(I_i) \cap R$ can still only intersect at most two boxes in Q, by using that $S^\ell(\frac{j}{n}) \in \frac{1}{n}\mathbb{Z}$ since n is a multiple of $M_0 L_0^{\ell-1}$.

Then in total since there are $M_0 L_0^{\ell-1}$ intervals $I_0, \ldots, I_{M_0 L_0^{\ell-1}-1}$, there are at most $2M_0 L_0^{\ell-1}$ nonzero entries on the diagonal of P_n^ℓ. □

Remark 3.1.1. Although the above argument works for slope -1, we do not allow slope -1 in S since powers of S could then have segments with slope $+1$.

3.2. Proof of Theorem 2.1 pointwise Weyl law

In this section we prove Theorem 2.1 using a Fourier series approximation of the projection matrix, and properties of the Markov matrix P_n and quantization U_n to identify potentially bad coordinates x. We first make some remarks about the proof and statement. For notational convenience, we will use n instead of n_k.

Remark 3.2.1. Let $r : \mathbb{N} \to \mathbb{N}$ be any function such that $r(m) < m$, like $r(m) = \lfloor m/2 \rfloor$ or $\lfloor \log m \rfloor$. This is a cut-off function that determines which Fourier coefficients to examine for bad coordinates with short loops.

(i) To show (2.3.2), we will show we can choose G_n (not depending on $I(n)$) so that $\#G_n \geq n - \frac{2M_0}{L_0 - 1} L_0^{r(\widetilde{K}(n))}$, and for $x \in G_n$ that

$$\left| \sum_{j:\theta^{(n,j)} \in I(n)} |\psi_x^{(n,j)}|^2 - \frac{|I(n)|}{2\pi} \right| \leq \frac{|I(n)|}{2\pi} \left[2\pi |I(n)|^{-1} \widetilde{K}(n)^{-1} + (1 + 2\pi |I(n)|^{-1} \widetilde{K}(n)^{-1}) \cdot 6 \cdot 2^{-r(\widetilde{K}(n))/2} \right]. \tag{3.2.1}$$

(ii) To ensure the right side of (3.2.1) is $o(|I(n)|)$ and $\#G_n = n(1 - o(|I(n)|))$, choose r so that

$$r(\widetilde{K}(n)) \to \infty, \quad \widetilde{K}(n) - r(\widetilde{K}(n)) - \log_{L_0} \frac{1}{|I(n)|} \to \infty, \quad \text{as } \widetilde{K}(n) \to \infty. \tag{3.2.2}$$

Since $|I(n)| \cdot \widetilde{K}(n) \to \infty$ by (2.3.1), then eventually $\frac{1}{|I(n)|} < \widetilde{K}(n)$, so condition (3.2.2) is always met if

$$\widetilde{K}(n) - r(\widetilde{K}(n)) - \log_{L_0} \widetilde{K}(n) \to \infty. \tag{3.2.3}$$

For example, $r(\widetilde{K}(n)) = \lfloor \widetilde{K}(n)/2 \rfloor$ or $r(\widetilde{K}(n)) = \lfloor \log \widetilde{K}(n) \rfloor$ always satisfy the conditions on r.

(iii) We are interested in sequences of intervals $I(n)$ where $|I(n)| \to 0$. For the proof, we will assume that $|I(n)|$ is bounded away from 2π. If $|I(n)|$ is near 2π, apply the Theorem to the complement $I(n)^c$ or to a larger interval around $I(n)^c$ that satisfies (2.3.1), to conclude

$$\left| \sum_{j:\theta^{(n,j)} \in I(n)} |\psi_x^{(n,j)}|^2 - \frac{|I(n)|}{2\pi} \right| = \left| \sum_{j:\theta^{(n,j)} \in I(n)^c} |\psi_x^{(n,j)}|^2 - \frac{|I(n)^c|}{2\pi} \right| \to 0,$$

as $|I(n)^c| \to 0$.

3.2.1. Fourier series approximation. Let $p_{I(n)}$ be the function on the unit circle in $\mathbb{C}$ defined by $p_{I(n)}(e^{it}) := \chi_{I(n)}(t)$, so that $p_{I(n)}(U)$ is the projection

$$P_{I(n)} := p_{I(n)}(U) = \sum_{j:\theta^{(n,j)} \in I(n)} |\psi^{(n,j)}\rangle\langle\psi^{(n,j)}|.$$

The sum $\sum_{j:\theta^{(n,j)} \in I(n)} |\psi_x^{(n,j)}|^2$ is the (x,x) coordinate of the projection matrix $P_{I(n)}$. To approximate $P_{I(n)}$ by a polynomial in powers of U_n, we approximate the indicator function $\chi_{I(n)}$ by trigonometric polynomials.

These particular polynomials are based on an entire function $B(z)$ introduced by Beurling, which satisfies $\operatorname{sgn}(x) \le B(x)$ for $x \in \mathbb{R}$, and $\int_{\mathbb{R}} (B(x) - \operatorname{sgn}(x))\, dx = 1$. The function $B(z)$ also satisfies an extremal property; it minimizes the L^1 difference $\int_{\mathbb{R}} (f(x) - \operatorname{sgn}(x))\, dx$ over entire functions f of exponential type 2π with $f(x) \ge \operatorname{sgn}(x)$ for $x \in \mathbb{R}$. By the Paley–Wiener theorem, exponential of type 2π means that the Fourier transform of $B(z)$ is supported in $[-2\pi, 2\pi]$. Selberg later used this function $B(z)$ to produce majorants and minorants of the characteristic function χ_I of an interval I, with compactly supported Fourier transform.

Theorem 3.4 (Beurling–Selberg function)**.** *Let $I \subset \mathbb{R}$ be a finite interval and $\delta > 0$. Then there are functions $g_{I,\delta}^{(+)}$ and $g_{I,\delta}^{(-)}$ such that*

(i) $g_{I,\delta}^{(-)}(x) \le \chi_I(x) \le g_{I,\delta}^{(+)}(x)$ for all $x \in \mathbb{R}$.
(ii) The Fourier transforms $\widehat{g_{I,\delta}^{(+)}}$ and $\widehat{g_{I,\delta}^{(-)}}$ are compactly supported in $[-\delta, \delta]$.
(iii) $\int_{\mathbb{R}} (g_{I,\delta}^{(+)}(x) - \chi_I(x))\, dx = 2\pi\delta^{-1}$ and $\int_{\mathbb{R}} (\chi_I(x) - g_{I,\delta}^{(-)}(x))\, dx = 2\pi\delta^{-1}$.

For references on Beurling and Selberg functions, see [Sel91, Chapter 45 §20], [Mon94], or [Vaa85]. For $I \subset \mathbb{R}/(2\pi\mathbb{Z})$ with $|I| < 2\pi$, to take 2π-periodic functions, define

$$G_{I,\delta}^{(+)}(x) = \sum_{j\in\mathbb{Z}} g_{I,\delta}^{(+)}(x - 2\pi j), \qquad G_{I,\delta}^{(-)}(x) = \sum_{j\in\mathbb{Z}} g_{I,\delta}^{(-)}(x - 2\pi j),$$

whose Fourier series coefficients agree with the Fourier transform of $g_{I,\delta}^{(+)}$ or $g_{I,\delta}^{(-)}$ at integers,

$$\widehat{G_{I,\delta}^{(+)}}(k) = \widehat{g_{I,\delta}^{(+)}}(k), \qquad \widehat{G_{I,\delta}^{(-)}}(k) = \widehat{g_{I,\delta}^{(-)}}(k). \tag{3.2.4}$$

Thus also using property (iii),

$$G_{I,\delta}^{(\pm)}(x) = \frac{|I| \pm 2\pi\delta^{-1}}{2\pi} + \sum_{\ell=1}^{\lfloor\delta\rfloor} \left(\widehat{g_{I,\delta}^{(\pm)}}(\ell) e^{i\ell x} + \widehat{g_{I,\delta}^{(\pm)}}(-\ell) e^{-i\ell x} \right). \tag{3.2.5}$$

These are trigonometric polynomials, sometimes called *Selberg polynomials*, that approximate χ_I well from above or below.

3.2.2. Projection matrix estimates. Take $\delta = \widetilde{K}(n)$, and define the functions on the unit circle in $\mathbb{C}$,

$$F_{I(n),\widetilde{K}(n)}^{(\pm)}(e^{it}) := G_{I(n),\widetilde{K}(n)}^{(\pm)}(t).$$

Recall we also defined $p_{I(n)}(e^{it}) = \chi_{I(n)}(t)$ and the projection $P_{I(n)} = p_{I(n)}(U_n)$, so that by the spectral theorem,

$$F^{(-)}_{I(n),\widetilde{K}(n)}(U_n)_{xx} \leq (P_{I(n)})_{xx} \leq F^{(+)}_{I(n),\widetilde{K}(n)}(U_n)_{xx}. \tag{3.2.6}$$

By (3.2.5) and the spectral theorem again,

$$F^{(\pm)}_{I(n),\widetilde{K}(n)}(U_n) = \frac{|I(n)|}{2\pi}(1 \pm 2\pi|I(n)|^{-1}\widetilde{K}(n)^{-1})\,\mathrm{Id} + \\ + \sum_{\ell=1}^{\widetilde{K}(n)} \left(\widehat{g^{(\pm)}_{I(n),\widetilde{K}(n)}}(\ell) U_n^{\ell} + + \widehat{g^{(\pm)}_{I(n),\widetilde{K}(n)}}(-\ell) U_n^{-\ell} \right). \tag{3.2.7}$$

The identity term $\frac{|I(n)|}{2\pi}(1 \pm 2\pi|I(n)|^{-1}\widetilde{K}(n)^{-1})\,\mathrm{Id}$ has the values we want already since $|I(n)|^{-1}\widetilde{K}(n)^{-1} \to 0$ by (2.3.1), so to show (3.2.1) we want to show the rest of the terms are small. Since

$$|\widehat{g^{(\pm)}_{I,\delta}}(\ell)| \leq \frac{1}{2\pi}\int_{\mathbb{R}} |g^{(\pm)}_{I,\delta}(x)|\,dx \leq \frac{1}{2\pi}(|I| + 2\pi\delta^{-1}), \tag{3.2.8}$$

then for any $x, y \in [n]$, the (x, y) element of the non-identity terms can be bounded as

$$\left| \sum_{\ell=1}^{\widetilde{K}(n)} \left(\widehat{g^{(\pm)}_{I(n),\widetilde{K}(n)}}(\ell)(U_n^{\ell})_{xy} + \widehat{g^{(\pm)}_{I(n),\widetilde{K}(n)}}(-\ell)(U_n^{-\ell})_{xy} \right) \right| \leq \\ \leq \frac{|I(n)|}{2\pi}(1 + 2\pi|I(n)|^{-1}\widetilde{K}(n)^{-1}) \sum_{\ell=1}^{\widetilde{K}(n)} (|(U_n^{\ell})_{xy}| + |(U_n^{\ell})_{yx}|). \tag{3.2.9}$$

3.2.3. Removing potentially bad points. Here we use properties of U_n and P_n from Section 3.1 to remove coordinates x where (3.2.9) may be large. For $1 \leq \ell \leq \widetilde{K}(n) + 1$, by Lemma 3.1(d) there is at most one path of length ℓ from a given x to itself (or to another y), so

$$\left|(U_n^{\ell})_{xx}\right| = \left| \sum_{\tau : x \xrightarrow{\ell} x} (U_n)_{\tau_0\tau_1} \cdots (U_n)_{\tau_{\ell-1}\tau_\ell} \right| = \left|(U_n)_{x\tau_1}(U_n)_{\tau_1\tau_2}\cdots(U_n)_{\tau_{\ell-1}x}\right| = ((P_n^{\ell})_{xx})^{1/2}.$$

Since all slopes of S are at least 2 in absolute value, then all the slopes of S^ℓ are at least 2^ℓ in absolute value, so by Lemma 3.2, $|(U_n^\ell)_{xx}| \le 2^{-\ell/2}$ for all $x \in [n]$. In order to make the sum $2\sum_{\ell=1}^{\widetilde{K}(n)} |(U_n^\ell)_{xx}|$ in (3.2.9) small then, we only need to be concerned with smaller ℓ, since $|(U_n^\ell)_{xx}|$ decays exponentially in ℓ. As we will see, by Lemma 3.3, for small ℓ, $(U_n^\ell)_{xx} = 0$ for most coordinates x, so we can pick a cut-off for small ℓ and just throw out any coordinates x where $(U_n^\ell)_{xx} \neq 0$ below this cut-off.

Let $r : \mathbb{N} \to \mathbb{N}$ satisfy $r(k) < k$ and (3.2.2); this will determine the cut-off for which ℓ are "small". Define the set of potentially bad coordinates as

$$B_n := \{x \in [n] : (U_n^\ell)_{xx} \neq 0 \text{ for some } \ell \in [1 : r(\widetilde{K}(n))]\}. \tag{3.2.10}$$

For $\ell \le \widetilde{K}(n) + 1$, by Lemma 3.3, the diagonal of U_n^ℓ contains at most $2 \cdot M_0 \cdot L_0^{\ell-1}$ nonzero entries, so there are not many bad points,

$$\#B_n \le 2M_0 \sum_{\ell=1}^{r(\widetilde{K}(n))} L_0^{\ell-1} = \frac{2M_0}{L_0 - 1}(L_0^{r(\widetilde{K}(n))} - 1) = o(n|I(n)|), \tag{3.2.11}$$

using assumption (3.2.2) for the last equality. For $x \in G_n := [n] \setminus B_n$, then

$$\begin{aligned} \sum_{\ell=1}^{\widetilde{K}(n)} |(U_n^\ell)_{xx}| + |(U_n^{-\ell})_{xx}| &= 2 \sum_{\ell=r(\widetilde{K}(n))+1}^{\widetilde{K}(n)} |(U_n^\ell)_{xx}| \\ &\le 2 \sum_{\ell=r(\widetilde{K}(n))+1}^{\infty} 2^{-\ell/2} = 2(1+\sqrt{2}) \cdot 2^{-r(\widetilde{K}(n))/2}. \end{aligned} \tag{3.2.12}$$

Then for $x \in G_n$,

$$\begin{aligned} &\left| (P_{I(n)})_{xx} - \frac{|I(n)|}{2\pi} \right| \\ &\quad \le \frac{|I(n)|}{2\pi} \left[2\pi |I(n)|^{-1} \widetilde{K}(n)^{-1} + (1 + 2\pi |I(n)|^{-1} \widetilde{K}(n)^{-1}) \cdot 6 \cdot 2^{-r(\widetilde{K}(n))/2} \right] \\ &\quad = o(|I(n)|), \end{aligned} \tag{3.2.13}$$

since $|I(n)|\widetilde{K}(n) \to \infty$ by (2.3.1). By (3.2.11), $\#G_n \ge n(1 - o(|I(n)|))$. □

3.3. Quantum ergodicity in bins

In this section we prove Theorem 2.3 concerning quantum ergodicity in bins $\{j : \theta^{(n_k,j)} \in I(n_k)\}$, following the standard proof of quantum ergodicity that uses the Egorov property.

Theorem 3.5 (Egorov property, [BKS07])**.** *Suppose S satisfies conditions (i)–(iv) and has a corresponding $n \times n$ unitary matrix U_n with eigenvectors $(\psi^{(n,j)})_{j=1}^n$. Let $O_n(h)$ be the quantum observable corresponding to $h : [0,1] \to \mathbb{C}$. If h is Lipschitz continuous on each image $S(E_x)$, and $n \in M_0 L_0 \mathbb{Z}$, then*

$$\|U_n O_n(h) U_n^{-1} - O_{n_k}(h \circ S)\| \le \frac{1}{2} L_0^2 M_0 \cdot \frac{\max_{x\in[n]} \|h\|_{\mathrm{Lip}(S(E_x))}}{n}, \tag{3.3.1}$$

where the norm on the left side is the operator norm.

If $t \le \widetilde{K}(n) + 1$, then by the same recursive argument as in Lemma 3.1(a), S^{t-1} is linear on each $S(E_x)$, so $h \circ S^{t-1}$ is Lipschitz on $S(E_x)$ with Lipschitz constant $\le \|h\|_{\mathrm{Lip}} L_0^{t-1}$. Then iterating (3.3.1) t times yields,

$$\begin{aligned}
&\|U_n^t O_n(h) U_n^{-t} - O_n(h \circ S^t)\| \\
&\qquad \le \sum_{r=1}^{t} \|U_n^{t-r}(U_n O_n(h \circ S^{r-1}) U_n^{-1}) U_n^{-(t-r)} - U_n^{t-r} O_n(h \circ S^r) U_n^{-(t-r)}\| \\
&\qquad \le \sum_{r=1}^{t} \|U_n O_n(h \circ S^{r-1}) U_n^{-1} - O_n(h \circ S^r)\| \\
&\qquad \le \sum_{r=1}^{t} \frac{L_0^2 M_0 \|h\|_{\mathrm{Lip}} L_0^{r-1}}{2n} \le \frac{C_S \|h\|_{\mathrm{Lip}} \cdot L_0^t}{n}. && (3.3.2)
\end{aligned}$$

If say $t \le \frac{\widetilde{K}(n)}{2}$, then $L_0^t \ll n$, so the error bound is small, and the Egorov property (3.3.2) relates the quantum dynamics $U_n^t O_n(h) U_n^{-t}$ to the classical dynamics $h \circ S^t$ for t well before the Ehrenfest time $T_E := \widetilde{K}(n) \lesssim \log n$.

3.3.1. Proof of Theorem 2.3. Since $O_{n_k}(h) - (\int_0^1 h) \cdot \mathrm{Id} = O_{n_k}(h - \int_0^1 h)$, wlog assume $\int_0^1 h = 0$ and define the quantum variance for a fixed bin $I(n_k)$,

$$V_{n_k} := \frac{1}{\#\{j : \theta^{(n_k,j)} \in I(n_k)\}} \sum_{j:\theta^{(n_k,j)} \in I(n_k)} \left|\langle \psi^{(n_k,j)}, O_{n_k}(h)\psi^{(n_k,j)}\rangle\right|^2, \tag{3.3.3}$$

which we will show tends to zero as $k \to \infty$. For a function $g : [0,1] \to \mathbb{C}$, define $[g]_T := \frac{1}{T}\sum_{t=0}^{T-1} g \circ S^t$. Using that $\psi^{(n_k,j)}$ are eigenvectors of U_{n_k}, followed by the Egorov property and averaging over t (stopping before $\frac{\widetilde{K}(n_k)}{2}$),

$$\begin{aligned}
\langle \psi^{(n_k,j)}, O_{n_k}(h)\psi^{(n_k,j)}\rangle &= \langle \psi^{(n_k,j)}, (U_{n_k}^*)^t O_{n_k}(h) U_{n_k}^t \psi^{(n_k,j)}\rangle \\
&= \langle \psi^{(n_k,j)}, O_{n_k}(h \circ S^t)\psi^{(n_k,j)}\rangle + \mathcal{O}\left(\frac{\|h\|_{\mathrm{Lip}} \cdot L_0^t}{n_k}\right) \\
&= \langle \psi^{(n_k,j)}, O_{n_k}([h]_T)\psi^{(n_k,j)}\rangle + \mathcal{O}\left(\frac{\|h\|_{\mathrm{Lip}} \cdot L_0^T}{T n_k}\right).
\end{aligned}$$

Then by Cauchy-Schwarz,

$$\begin{aligned}
|\langle \psi^{(n_k,j)}, O_{n_k}(h)\psi^{(n_k,j)}\rangle|^2 &\le |\langle \psi^{(n_k,j)}, O_{n_k}([h]_T)\psi^{(n_k,j)}\rangle|^2 + \mathcal{O}_h\left(\frac{L_0^T}{T n_k}\right) \\
&\le \langle \psi^{(n_k,j)}, O_{n_k}([h]_T^*) O_{n_k}([h]_T)\psi^{(n_k,j)}\rangle + \mathcal{O}_h\left(\frac{L_0^T}{T n_k}\right). \qquad (3.3.4)
\end{aligned}$$

For this quantization method, just a sup norm bound shows

$$\begin{aligned}
|O_{n_k}(ab)_{xx} - O_{n_k}(a)_{xx} O_{n_k}(b)_{xx}| &= n \left| \int_{E_x} \left(a - \frac{1}{|E_x|}\int_{E_x} a\right)\left(b - \frac{1}{|E_x|}\int_{E_x} b\right)\right| \\
&\le \frac{\|a\|_{\mathrm{Lip}(E_x)} \|b\|_{\mathrm{Lip}(E_x)}}{n^2},
\end{aligned}$$

so that

$$\|O_{n_k}(ab) - O_{n_k}(a) O_{n_k}(b)\| \le \max_{x \in [1:n_k]} \frac{\|a\|_{\mathrm{Lip}(E_x)} \|b\|_{\mathrm{Lip}(E_x)}}{n_k^2}. \tag{3.3.5}$$

Taking $T = \lfloor \frac{\widetilde{K}(n_k)}{2} \rfloor$, then for $t \le T$, S^t is linear on every E_x so that $\| \frac{1}{T} \sum_{t=0}^{T-1} h \circ S^t \|_{\mathrm{Lip}(E_x)} \le \frac{1}{T} \sum_{t=0}^{T-1} \|h\|_{\mathrm{Lip}} s_{\max}^t = \mathcal{O}_h(\frac{s_{\max}^T}{T})$, and

$$|\langle \psi^{(n_k,j)}, O_{n_k}(h)\psi^{(n_k,j)}\rangle|^2 \le \langle \psi^{(n_k,j)}, O_{n_k}(|[h]_T|^2)\psi^{(n_k,j)}\rangle + \mathcal{O}_h\left(\frac{L_0^T}{Tn_k}\right). \tag{3.3.6}$$

Applying the above and Theorem 2.1 yields

$$\begin{aligned}
&\frac{1}{\#\{j : \theta^{(n_k,j)} \in I(n_k)\}} \sum_{j:\theta^{(n_k,j)} \in I(n_k)} \left|\langle \psi^{(n_k,j)}, O_{n_k}(h)\psi^{(n_k,j)}\rangle\right|^2 \\
&\quad \le \frac{2\pi}{n_k|I(n_k)|(1+o(1))} \sum_{j:\theta^{(n_k,j)} \in I(n_k)} \langle \psi^{(n_k,j)}, O_{n_k}(|[h]_T|^2)\psi^{(n_k,j)}\rangle + \mathcal{O}_h\left(\frac{L_0^T}{Tn_k}\right) \\
&\quad \le \frac{2\pi(1+o(1))}{n_k|I(n_k)|} \left(\sum_{x\in G_{n_k}} \sum_{j:\theta^{(n_k,j)} \in I(n_k)} |\psi_x^{(n_k,j)}|^2 O_{n_k}(|[h]_T|^2)_{xx} + \sum_{x\in B_{n_k}} \|h\|_\infty^2 \right) + o(1) \\
&\quad \le (1+o(1)) \cdot \int_0^1 \left| \frac{1}{T} \sum_{t=0}^{T-1} h(S^t(y)) \right|^2 dy + \frac{C \cdot L_0^{r(\widetilde{K})} \|h\|_\infty^2}{n_k|I(n_k)|} + o(1) \xrightarrow{k\to\infty} 0,
\end{aligned}$$

using the L^2 ergodic theorem as $T = \lfloor \frac{\widetilde{K}(n_k)}{2} \rfloor \to \infty$. □

The passage from decay of the quantum variance (2.3.6) to the density one statement is by the usual method (for details see for example Theorem 15.5 in the textbook [Zwo12]). To start, by Chebyshev–Markov with $\varepsilon = V_{n_k}^{1/4}$, Theorem 2.3 implies for a single Lipschitz function h, there is the sequence of sets $\Lambda_{n_k}(h) \subseteq \{j : \theta^{(n_k,j)} \in I(n_k)\}$ with

$$\frac{\#\Lambda_{n_k}(h)}{\#\{j : \theta^{(n_k,j)} \in I(n_k)\}} \to 1, \tag{3.3.7}$$

such that for all sequences $(j_{n_k})_k$ with $j_{n_k} \in \Lambda_{n_k}(h)$,

$$\lim_{k\to\infty} \langle \psi^{(n_k,j_{n_k})}, O_{n_k}(h)\psi^{(n_k,j_{n_k})}\rangle = \int_0^1 h(x)\,dx. \tag{3.3.8}$$

For a countable set of Lipschitz functions $(h_\ell)_\ell$, since finite intersections of sets Λ_{n_k} satisfying (3.3.7) also satisfy (3.3.7), we can assume $\Lambda_{n_k}(h_{\ell+1}) \subseteq \Lambda_{n_k}(h_\ell)$ for all n_k. Then for each h_ℓ, let $N(\ell) > 0$ be large enough so that for $n_k \geq N(\ell)$,

$$\frac{\#\Lambda_{n_k}(h_\ell)}{\#\{j : \theta^{(n_k,j)} \in I(n_k)\}} \geq 1 - \frac{1}{\ell}. \tag{3.3.9}$$

Take $N(\ell)$ increasing in ℓ and let $\Lambda^\infty_{n_k} := \Lambda_{n_k}(h_\ell)$ for $N(\ell) \leq n_k < N(\ell+1)$, so (3.3.8) holds for sequences in $\Lambda^\infty_{n_k}$ and h_ℓ in the countable set. Then take $(h_\ell)_\ell$ to be a countable set of Lipschitz functions that are dense in $(C([0,1]), \|\cdot\|_\infty)$, so that for any $h \in C([0,1])$,

$$\begin{aligned}
&\left|\langle \psi^{(n_k,j_{n_k})}, O_{n_k}(h)\psi^{(n_k,j_{n_k})}\rangle - \int_0^1 h\right| \\
&\qquad\leq \left|\langle \psi^{(n_k,j_{n_k})}, O_{n_k}(h-h_\ell)\psi^{(n_k,j_{n_k})}\rangle\right| + \\
&\qquad\quad + \left|\langle \psi^{(n_k,j_{n_k})}, O_{n_k}(h_\ell)\psi^{(n_k,j_{n_k})}\rangle - \int_0^1 h_\ell\right| + \left|\int_0^1 (h_\ell - h)\right|.
\end{aligned}$$

The terms on the right side are bounded by $\|h - h_\ell\|_\infty$ or are $o(1)$ as $k \to \infty$.

3.4. Random Gaussian eigenvectors

In this section we prove Theorems 2.4 and 2.5 on random unitary rotations of bins of eigenvectors. To analyze the statistics of the rotated eigenvectors, we look at their coordinate values, which can be expressed as one-dimensional random projections. The behavior of low-dimensional projections of high-dimensional vectors has been well-studied since the 1970s for its applications in analyzing large data sets; see for example the survey [Hub85] for an overview of the early history and motivation of "projection pursuit" methods. The marginals of high-dimensional random vectors are often known to look approximately Gaussian, with precise conditions first proved by Diaconis and Freedman in [DF84].

3.4.1. Random projections and bases. For Theorem 2.4, we are interested in the coordinate values of a random unit vector in the span of $V := \{\psi^{(n,j)} : \theta^{(n,j)} \in I(n)\}$. Let M_V be the $n \times (\dim V)$ matrix whose columns are the basis $(\psi^{(j)})$ in V. Then $P_{I(n)} := M_V M_V^*$ is the projection onto this space, and a random unit vector ϕ in the span is chosen according to $\omega \sim N_{\mathbb{C}}(0, P_{I(n)})/\|N_{\mathbb{C}}(0, P_{I(n)})\|_2 \sim M_V u$ for $u \sim \mathrm{Unif}(\mathbb{S}^{\dim V - 1}_{\mathbb{C}})$. The coordinates

$\phi_1 = \langle \phi_1, e_x \rangle, \phi_2, \dots, \phi_n$ are

$$\langle u, M_V^* e_1 \rangle, \langle u, M_V^* e_2 \rangle, \dots, \langle u, M_V^* e_n \rangle$$

which is a 1-dimensional projection in the direction $u \in \mathbb{C}^{\dim V}$ of the data set $\{M_V^* e_1, \dots, M_V^* e_n\} \subset \mathbb{C}^{\dim V}$. Since $\sum_{x=1}^n |\phi_x|^2 = 1$, we use the scaled data set $\sqrt{n}\{M_V^* e_1, \dots, M_V^* e_n\}$. The following theorem due to Meckes [Mec09] and Chatterjee and Meckes [CM08] is a quantitative version of the theorem from [DF84].

Theorem 3.6 (Complex version of Theorem 2 in [Mec09]). *Let $\{x_j\}_{j=1}^n$ be deterministic vectors in $\mathbb{C}^d$. Define $\sigma^2 = \frac{1}{nd}\sum_{i=1}^n |x_i|^2$ and suppose*

$$\frac{1}{n}\sum_{i=1}^n \left| \frac{|x_i|^2}{\sigma^2} - d \right| \leq A \tag{3.4.1}$$

$$\sup_{\theta \in \mathbb{S}_{\mathbb{C}}^{d-1}} \frac{1}{n}\sum_{i=1}^n |\langle \theta, x_i \rangle|^2 \leq B. \tag{3.4.2}$$

For a point $\theta \in \mathbb{S}_{\mathbb{C}}^{d-1} \subset \mathbb{C}^d$, define the measure $\mu_\theta^{(n)} := \frac{1}{n}\sum_{j=1}^n \delta_{\langle \theta, x_j \rangle}$ on $\mathbb{C}$. Then for $\theta \sim \mathrm{Unif}(\mathbb{S}_{\mathbb{C}^{d-1}})$, any bounded Lipschitz $f : \mathbb{C} \to \mathbb{C}$ with Lipschitz constant $L = \|f\|_{\mathrm{Lip}} := \sup_{x \neq y} \frac{|f(x)-f(y)|}{|x-y|}$, and $\varepsilon > \frac{2L(A+3)}{d-1}$, there is the quantitative bound

$$\mathbb{P}\left[\left| \int f(x)\, d\mu_\theta^{(n)}(x) - \mathbb{E}f(\sigma Z) \right| > \varepsilon \right] \leq 6 \exp\Big(- \frac{\varepsilon^2 d}{2^6 L^2 B} \Big), \tag{3.4.3}$$

where $Z \sim N_{\mathbb{C}}(0,1)$. In particular, if $A = o(d)$ and $B = o(d)$ and $\sigma^2 = 1$, then $\mu^{(n)}$ converges weakly in probability to $N_{\mathbb{C}}(0,1)$.

Proof. The proof is the same as the real version in [Mec09], except that the (multi-dimensional) Theorem 3.7 written below from [CM08] replaces the single-variable version. The proof idea from [Mec09] is to let $F(\theta) := \frac{1}{n}\sum_{x=1}^n f(\langle \theta, x_i \rangle)$ and write

$$\mathbb{P}\left[|F(\theta) - \mathbb{E}f(Z)| > \varepsilon \right] \leq \mathbb{P}\left[|F(\theta) - \mathbb{E}F(\theta)| > \varepsilon - |\mathbb{E}F(\theta) - \mathbb{E}f(Z)| \right].$$

Then one uses Theorem 3.7, a generalization of Stein's method of exchangeable pairs for abstract normal approximation, to bound $|\mathbb{E}F(\theta) - \mathbb{E}f(Z)|$ with $V = \langle \theta, x_I \rangle$ where $I \sim$

Unif$[n]$, and then one can apply Gaussian concentration (Lemma 3.8, cf. Section 7.1) to F which is $(L\sqrt{B})$-Lipschitz. $\square$

Theorem 3.7 (Theorem 2.5 for $\mathbb{C}$ in [CM08]). *Let W be a $\mathbb{C}$-valued random variable and for each $\varepsilon > 0$ let W_ε be a random vector such that $\mathcal{L}(W) = \mathcal{L}(W_\varepsilon)$, with the property that $\lim_{\varepsilon\to 0} W_\varepsilon = W$ almost surely. Suppose there is a function $\lambda(\varepsilon)$ and measurable Γ, Λ such that as $\varepsilon \to 0$,*

(i) $\frac{1}{\lambda(\varepsilon)}\mathbb{E}[(W_\varepsilon - W)|W] \xrightarrow{L^1} -W$.

(ii) $\frac{1}{2\lambda(\varepsilon)}\mathbb{E}[|W_\varepsilon - W|^2|W] \xrightarrow{L^1} 1 + \mathbb{E}[\Gamma|W]$.

(iii) $\frac{1}{2\lambda(\varepsilon)}\mathbb{E}[(W_\varepsilon - W)^2|W] \xrightarrow{L^1} \mathbb{E}[\Lambda|W]$.

(iv) $\frac{1}{\lambda(\varepsilon)}\mathbb{E}|W_\varepsilon - W|^3 \to 0$.

Then letting $Z \sim N_{\mathbb{C}}(0,1)$,

$$d_{\mathrm{Wass}}(W,Z) \le \mathbb{E}|\Gamma| + \mathbb{E}|\Lambda|, \tag{3.4.4}$$

where d_{Wass} is the Wasserstein distance $d_{\mathrm{Wass}}(W,Z) = \sup_{\|g\|_{\mathrm{Lip}}\le 1} |\mathbb{E}g(X) - \mathbb{E}g(Z)|$.

Lemma 3.8 (Gaussian concentration on the complex sphere). *Let $F : \mathbb{C}^d \to \mathbb{C}$ be L-Lipschitz and $\theta \sim \mathrm{Unif}(\mathbb{S}^{d-1}_{\mathbb{C}})$. Then*

$$\mathbb{P}[|F(\theta) - \mathbb{E}F(\theta)| \ge t] \le 6\exp\left(-\frac{t^2 d}{16L^2}\right). \tag{3.4.5}$$

Applying Theorem 3.6 to our case immediately yields the following.

Theorem 3.9 (Complex projection version of Theorem 2 in [Mec09]). *Let $P^{(\nu)}$ be an $n \times n$ self-adjoint projection matrix onto a d-dimensional subspace $V^{(\nu)}$ of $\mathbb{C}^n$, and suppose*

$$\sum_{x=1}^{n} \left| \|P^{(\nu)} e_x\|_2^2 - \frac{d}{n} \right| \le A. \tag{3.4.6}$$

Let $\omega = (\omega_1, \ldots, \omega_n)$ be chosen uniformly at random from the $(d-1)$-dimensional sphere $\mathbb{S}(V^{(\nu)}) := \{v \in V^{(\nu)} : \|v\| = 1\}$, and define the empirical distribution $\widetilde{\mu}^{(\nu)}_\omega$ of the coordinates *of* ω *scaled by* $\sqrt{n}$*,*

$$\widetilde{\mu}_\omega^{(\nu)} := \frac{1}{n}\sum_{x=1}^{n} \delta_{\sqrt{n}\omega_x}.$$

Then for $f : \mathbb{C} \to \mathbb{C}$ *bounded Lipschitz and* $\varepsilon > \frac{2\|f\|_{\mathrm{Lip}}(A+3)}{d-1}$*,*

$$\mathbb{P}\left[\left|\int f(x)\, d\mu_\omega^{(\nu)}(x) - \mathbb{E}f(Z)\right| > \varepsilon\right] \leq 6\exp\Big(-\frac{\varepsilon^2 d}{2^6\|f\|_{\mathrm{Lip}}^2}\Big), \tag{3.4.7}$$

where $Z \sim N_{\mathbb{C}}(0,1)$*.*

Remark 3.4.1. As will be shown by (3.4.13) and (3.4.11), Theorem 2.1 shows $A = o(d)$ with $d = \frac{n_k|I(n_k)|}{2\pi}(1+o(1))$, so the above theorem proves Theorem 2.4.

Proof. Let $v_1, \ldots, v_d$ be an orthonormal basis for $V^{(\nu)}$, and let M_V be the $n \times d$ matrix with those vectors as columns. Then $P^{(\nu)} = M_V M_V^*$, and

$$\|M_V^* e_x\|_{\mathbb{C}^d}^2 = \langle e_x, M_V M_V^* e_x\rangle_{\mathbb{C}^n}^2 = \|P^{(\nu)} e_x\|_{\mathbb{C}^n}^2.$$

Apply Theorem 3.6 to the data set $\sqrt{n}M_V^* e_1, \sqrt{n}M_V^* e_2, \ldots, \sqrt{n}M_V^* e_n$ in $\mathbb{C}^d$. We can take $B = 1$ since for any $\theta \in \mathbb{S}_{\mathbb{C}}^{d-1}$,

$$\frac{1}{n}\sum_{x=1}^{n} |\langle\theta, \sqrt{n}M_V^* e_x\rangle|^2 = \sum_{x=1}^{n}\langle M_V\theta, e_x\rangle\langle e_x, M_V\theta\rangle = \|M_V\theta\|_{\mathbb{C}^n}^2 = \langle\theta, M_V^* M_V\theta\rangle_{\mathbb{C}^d} = \|\theta\|_{\mathbb{C}^d}^2.$$

If θ is uniform on $\mathbb{S}_{\mathbb{C}}^{d-1} \subset \mathbb{C}^d$, then $M_V\theta$ is uniform on $\mathbb{S}(V^{(\nu)})$, so $\frac{1}{n}\sum_{j=1}^{n}\delta_{\langle\theta, \sqrt{n}M_V^* e_x\rangle} \sim \frac{1}{n}\sum_{j=1}^{n}\delta_{\sqrt{n}\omega_x}$. ☐

Theorem 3.9 provides a bound for the probability that a single randomly chosen vector does not look Gaussian. Because the quantitative bound (3.4.7) decays quickly, a simple union bound gives a bound on finding an entire orthonormal basis that looks Gaussian (Corollary 3.11 below). This family of random orthonormal bases will then be used to construct the unitary matrices $V_{n_k}(\beta^{[n_k]})$ in Theorem 2.5.

Lemma 3.10 (union bound for random ONB)**.** *Let* $B \subset \mathbb{S}_{\mathbb{C}}^{d-1}$ *and let* σ *be surface measure on* $\mathbb{S}_{\mathbb{C}}^{d-1}$ *normalized so* $\sigma(\mathbb{S}_{\mathbb{C}}^{d-1}) = 1$*. Then a random orthonormal basis of* $\mathbb{C}^d$ *(chosen from Haar measure) avoids* B *with probability at least* $1 - d\sigma(B)$*.*

Proof. Let μ be normalized Haar measure on $U(d)$. Then for any $x \in \mathbb{S}_{\mathbb{C}}^{d-1}$, $\sigma(A) = \mu(g \in U(d) : g(x) \in A)$. By union bound, letting $\{e_j\}$ be the standard basis,

$$\begin{aligned}\mu(\{g \in U(d) : g(e_j) \in B \text{ for some } j \in [1:d]\}) &\leq d \cdot \mu(\{g \in U(d) : g(e_1) \in B\}) \\ &= d \cdot \sigma(B),\end{aligned}$$

so $\mu(\{g \in U(d) : \forall j \in [1:d],\ g(e_j) \notin B\}) \geq 1 - d\sigma(B)$. $\square$

Corollary 3.11 (Random Gaussian basis)**.** *Let $\mathbb{C}^n = V^{[1]} \oplus \cdots \oplus V^{[\kappa]}$, and let $P^{[\ell]}$ be the orthogonal projection onto the subspace $V^{[\ell]}$. Suppose there is A and $d_1, \ldots, d_\kappa \in \mathbb{R}_+$ so that*

$$\sum_{x=1}^{n} \left| \|P^{[\ell]} e_x\|_2^2 - \frac{d_\ell}{n} \right| \leq A, \quad \forall \ell \in [1:\kappa]. \tag{3.4.8}$$

Choose a random orthonormal basis $(\phi^{[j]})_j$ for $\mathbb{C}^n$ by choosing a random orthonormal basis from each $V^{[\ell]}$ (according to Haar measure), and let

$$\mu^{[j]} := \frac{1}{n} \sum_{x=1}^{n} \delta_{\sqrt{n}\phi_x^{[j]}},$$

the empirical distribution for the jth basis vector's coordinates. Then for any $f : \mathbb{C} \to \mathbb{C}$ bounded and Lipschitz and $\varepsilon > \frac{2\|f\|_{\mathrm{Lip}}(2A+3)}{(\min d_\ell - A) - 1}$,

$$\mathbb{P}\left[\max_{j \in [n]} \left| \int f(x)\, d\mu^{[j]}(x) - \mathbb{E} f(Z) \right| > \varepsilon \right] \leq 6n \exp\left(-\frac{\varepsilon^2 (\min d_\ell - A)}{2^6 \|f\|_{\mathrm{Lip}}^2} \right), \tag{3.4.9}$$

where $Z \sim N_{\mathbb{C}}(0,1)$.

Proof. The numbers d_ℓ need not be the dimensions of $V^{[\ell]}$, but since

$$\left| \frac{\dim V^{[\ell]}}{n} - \frac{d_\ell}{n} \right| = \left| \frac{1}{n} \sum_{x=1}^{n} \left(\|P^{[\ell]} e_x\|_2^2 - \frac{d_\ell}{n} \right) \right| \leq \frac{1}{n} A, \tag{3.4.10}$$

then

$$\sum_{x=1}^{n} \left| \|P^{[\ell]} e_x\|_2^2 - \frac{\dim V^{[\ell]}}{n} \right| \leq 2A. \tag{3.4.11}$$

Then Theorem 3.9 implies for $\varepsilon > \frac{2\|f\|_{\mathrm{Lip}}(2A+3)}{\min d_\ell - A - 1} \geq \frac{2\|f\|_{\mathrm{Lip}}(2A+3)}{\min \dim V^{[j]} - 1}$, that $R_\ell(f) := \{\omega \in \mathbb{S}(V^{[\ell]}) : \left|\int f(x)\, d\mu_\omega(x) - \mathbb{E} f(Z)\right| > \varepsilon\}$ has small measure $\leq 6\exp(-\frac{\varepsilon^2 \dim V^{[\ell]}}{2^6\|f\|_{\mathrm{Lip}}^2})$. By Lemma 3.10, a random orthonormal basis for $V^{[\ell]}$ avoids $R_\ell(f)$ with probability at least $1 - \dim V^{[\ell]} \cdot 6\exp(-\frac{\varepsilon^2 \dim V^{[\ell]}}{2^6\|f\|_{\mathrm{Lip}}})$. Thus letting $I_\ell \subset [n]$ be the set of indices j corresponding to $V^{[\ell]}$,

$$\mathbb{P}\Big[\max_{j\in[n]}\Big|\int f(x)\, d\mu^{[j]}(x) - \mathbb{E}f(Z)\Big| > \varepsilon\Big]$$

$$\leq \sum_{\ell=1}^{\kappa} \mathbb{P}\left[\left|\int f(x)\, d\mu^{[j]}(x) - \mathbb{E}f(Z)\right| > \varepsilon \text{ for some } j \in I_\ell\right]$$

$$\leq \sum_{\ell=1}^{\kappa} \dim V^{[\ell]} \cdot 6\exp\left(-\frac{\varepsilon^2 \dim V^{[\ell]}}{2^6\|f\|_{\mathrm{Lip}}}\right)$$

$$\leq 6n\exp\left(-\frac{\varepsilon^2 \min\dim V^{[\ell]}}{2^6\|f\|_{\mathrm{Lip}}}\right) \leq 6n\exp\left(-\frac{\varepsilon^2(\min d_\ell - A)}{2^6\|f\|_{\mathrm{Lip}}}\right).$$

□

3.4.2. Proof of Theorem 2.5. Choose $\kappa(n_k) \in \mathbb{N}$ so that if we divide $[0, 2\pi]$ up into $\kappa(n_k)$ equal sized intervals $I_1(n_k), \dots, I_{\kappa(n_k)}(n_k)$, then (2.3.1) holds for $|I(n_k)| = \frac{2\pi}{\kappa(n_k)}$. Let $\psi^{(n_k,j)}$ be the jth eigenvector of U_{n_k}. Like the method used in [CG18], construct $V_{n_k}(\beta^{[n_k]})$ by taking a random unitary rotation (according to Haar measure) of the eigenvectors $\{\psi^{(n_k,j)} : \theta^{(n_k,j)} \in I_\ell(n_k)\}$ within each interval. Then perturb any degenerate eigenvalues to be simple, while still keeping them in the same bin. Denote the resulting eigenvectors of $V_{n_k}(\beta^{[n_k]})$ by $\phi^{[n_k,j]}_{(\beta)}$.

(a) Let $\widetilde{U}_{n_k}$ be the perturbation of U_{n_k} obtained by reassigning all eigenvalues in the same bin $I_\ell(n_k)$ to a single value $e^{i\Theta_\ell}$ in the bin. Then

$$\|(U_{n_k} - \widetilde{U}_{n_k})v\|_2^2 = \sum_{j=1}^{n} |e^{i\theta^{(j)}} - e^{i\Theta_\ell(j)}|^2 |\langle \psi^{(j)}, v\rangle|^2 \leq C\frac{(2\pi)^2}{\kappa(n_k)^2}\|v\|_2^2,$$

since the reassigned eigenvalues are still in the same bin. Also, $\|\widetilde{U}_{n_k} - V_{n_k}(\beta^{[n_k]})\| \leq C\frac{2\pi}{\kappa(n_k)}$ by the same computation, since $\widetilde{U}_{n_k}$ has degenerate eigenspaces that can be rotated to match the eigenvectors of $V_{n_k}(\beta^{[n_k]})$. Thus for any random $V_{n_k}(\beta^{[n_k]})$, $\|U_{n_k} - V_{n_k}(\beta^{[n_k]})\| \leq C\frac{2\pi}{\kappa(n_k)} = o(1)$. The Egorov property for U_{n_k}, Theorem 3.5, then implies the weaker Egorov property for $V_{n_k}(\beta^{[n_k]})$, since if A and B are unitary, then

$$\|AMA^{-1} - BMB^{-1}\| = \|(A-B)MA^{-1} + BM(A^{-1} - B^{-1}) \leq 2\|A-B\|\|M\|,$$

and this also holds if we replace M with $M - c\cdot\mathrm{Id}$ for any $c \in \mathbb{C}$ like $c = \int_0^1 h$ or $c = h(0)$.

(b) To show Gaussian behavior, we first show there is $\varepsilon(n_k) \to 0$ so that for any bounded Lipschitz $f : \mathbb{C} \to \mathbb{C}$, as $k \to \infty$,

(3.4.12)
$$\mathbb{P}\left[\max_{j\in[1:n_k]}\left|\int f(x)\, d\mu_\beta^{[n_k,j]}(x) - \mathbb{E}f(Z)\right| > \|f\|_{\mathrm{Lip}}\varepsilon(n_k)\right] \leq 6n_k \exp(-Cn_k^{1/2}|I(n_k)|^{1/2}),$$

where $Z \sim N_{\mathbb{C}}(0,1)$. A density argument followed by tightness will then complete the proof of (b).

To show (3.4.12), note that for any $W^{[\ell]} = \mathrm{span}\{\psi^{(n_k,j)} : \theta^{(j)} \in I_\ell(n_k)\}$ and $P^{[\ell]}$ the orthogonal projection onto $W^{[\ell]}$, the pointwise Weyl law Theorem 2.1 implies

$$\sum_{x=1}^{n}\left|\|P^{[\ell]}e_x\|_2^2 - \frac{|I(n_k)|}{2\pi}\right| \leq \sum_{x\in G_{n_k}} \frac{|I(n_k)|}{2\pi}o(1) + \sum_{x\in B_{n_k}} 2 = o(n_k|I_\ell(n_k)|), \tag{3.4.13}$$

so the quantity A in Corollary 3.11 can be taken to be $o(n_k|I_\ell(n_k)|)$. Let $\mu_{(\beta)}^{[n_k,j]}$ be the coordinate distribution of the jth eigenvector $\phi_{(\beta)}^{[n_k,j]}$ of $V_{n_k}(\beta^{[n_k]})$. Then applying Corollary 3.11 with all $d_\ell = \frac{n_k|I(n_k)|}{2\pi}$ and

$$\varepsilon(n_k) = \max\left(\frac{4A+6}{(d_\ell - A) - 1}, \frac{1}{(n_k|I(n_k)|)^{1/4}}\right) \to 0, \tag{3.4.14}$$

this yields for $Z \sim N_{\mathbb{C}}(0,1)$,

$$\begin{aligned}
\mathbb{P}\Big[\max_{j\in[1:n_k]}\Big|\int_{\mathbb{C}} f(x)\, d\mu_{(\beta)}^{[n_k,j]}(x) - \mathbb{E}f(Z)\Big| > \|f\|_{\mathrm{Lip}}\varepsilon(n_k)\Big]& \\
\leq 6n_k \exp\left(-\frac{\varepsilon(n_k)^2(n_k|I(n_k)| - 2\pi A)}{2^7\pi}\right)& \\
\leq 6n_k \exp\left(-\frac{n_k^{1/2}|I(n_k)|^{1/2}(1 - \frac{2\pi A}{n_k|I(n_k)|})}{2^7\pi}\right).& \qquad (3.4.15)
\end{aligned}$$

Now let $(f_\ell)_\ell$ be a countable set of Lipschitz functions with compact support that are dense in $C_c(\mathbb{C})$, and set

$$\Pi_{n_k} = \Big\{ V_{n_k}(\beta^{[n_k]}) : \forall \ell \in [1:n_k], j \in [1:n_k], \\ \Big| \int_{\mathbb{C}} f_\ell(x)\, d\mu_{(\beta)}^{[n_k,j]} - \mathbb{E} f_\ell(Z) \Big| \le \|f_\ell\|_{\mathrm{Lip}} \varepsilon(n_k) \Big\}. \tag{3.4.16}$$

Then

$$\mathbb{P}[\Pi_{n_k}^c] \le 6 n_k^2 \exp(-C n_k^{1/2} |I(n_k)|^{1/2}) \to 0, \tag{3.4.17}$$

since by (2.3.1), $n_k^{1/2}|I(n_k)|^{1/2} \gg 2 \log n_k$. For a sequence of matrices $(\widetilde{V}_{n_k})_k$ with $\widetilde{V}_{n_k} \in \Pi_{n_k}$, let $\widetilde{\mu}^{[n_k,j]}$ be the scaled coordinate distribution of the jth eigenvector $\widetilde{\phi}^{[n_k,j]}$ of $\widetilde{V}_{n_k}$. By definition of Π_{n_k}, we know for any f_ℓ that $\int f_\ell \, d\mu^{[n_k,j_{n_k}]} \to \mathbb{E} f_\ell(Z)$ as $k \to \infty$, for any sequence $(j_{n_k})_k$ with $j_{n_k} \in [1:n_k]$. Denseness of $(f_\ell)_\ell$ shows that this holds for all $f \in C_c(\mathbb{C})$ as well. Then $(\widetilde{\mu}^{[n_k,j_{n_k}]})_k$ is tight, and with the vague convergence we get weak convergence of $\widetilde{\mu}^{[n_k,j_{n_k}]}$ to $N_{\mathbb{C}}(0,1)$.

(c) To show QUE, like in (b), we first show there is $\varepsilon(n_k) \to 0$ so that for any bounded Lipschitz $h : [0,1] \to \mathbb{C}$, as $k \to \infty$,

$$\mathbb{P}\left[\max_{j \in [1:n_k]} \left| \langle \phi_{(\beta)}^{[n_k,j]}, O_{n_k}(h) \phi_{(\beta)}^{[n_k,j]} \rangle - \int_0^1 h(x)\, dx \right| > \|h\|_\infty \varepsilon(n_k) \right] \\ \le C n_k \exp(-c n_k^{1/2} |I(n_k)|^{1/2}). \tag{3.4.18}$$

This is done by the same argument presented in [CG18] using the Hanson–Wright inequality [RV13]. After proving (3.4.18), part (c) follows from density like in (b).

For $W^{[\ell]}$ with dimension d, let $M_{W^{[\ell]}}$ be an $n \times d$ matrix whose d columns form an orthonormal basis for $W^{[\ell]}$. Then $\phi^{[n_k,j]}$ chosen randomly from $\mathbb{S}(W^{[\ell]})$ is distributed like $M_{W^{[\ell]}} u$ for $u \sim \mathrm{Unif}(\mathbb{S}_{\mathbb{C}}^{d-1})$, and

$$\langle \phi^{[n_k,j]}, O_{n_k}(h) \phi^{[n_k,j]} \rangle_{\mathbb{C}^n} \sim \langle u, (M_{W^{[\ell]}}^* O_{n_k}(h) M_{W^{[\ell]}}) u \rangle_{\mathbb{C}^d}. \tag{3.4.19}$$

The Hanson–Wright inequality combined with subgaussian concentration on the norm $\|N_{\mathbb{C}}(0,I_d)\|_2$ shows that $\langle u, (M^*_{W^{[\ell]}} O_{n_k}(h) M_{W^{[\ell]}})u\rangle$ concentrates around its mean $\frac{1}{d}\operatorname{tr}(M^*_{W^{[\ell]}} O_{n_k}(h) M_{W^{[\ell]}})$ (see [CG18], Theorem 4.1 for details), which by the pointwise Wey law Theorem 2.1 is $\int_0^1 h + R(n_k,h)$ with $|R(n_k,h)| \le R(n_k)\|h\|_\infty$, some $R(n_k)\to 0$. In particular, for any $\varepsilon > 2|R(n_k,h)|$,

$$\mathbb{P}\left[\left|\langle \phi^{[n_k,j]}_{(\beta)}, O_{n_k}(h)\phi^{[n_k,j]}_{(\beta)}\rangle - \int_0^1 h\right| > \varepsilon\right] \le C_1 \exp\left(-C_2 \min\left(\frac{\varepsilon^2}{4\|h\|_\infty^2}, \frac{\varepsilon}{2\|h\|_\infty}\right)\cdot d\right). \tag{3.4.20}$$

Then taking $\varepsilon(n_k) = \max\left(2R(n_k), \frac{1}{(n_k|I(n_k)|)^{1/4}}\right)$ and applying a union bound like with Lemma 3.10 yields (3.4.18), using that eventually $\min(\varepsilon(n_k)^2, \varepsilon(n_k)) = \varepsilon(n_k)^2$.

Next, taking $(h_\ell)_\ell$ to be a countable dense set of Lipschitz functions in $C([0,1])$, let

$$\begin{aligned}\Gamma_{n_k} = \Bigg\{ & V_{n_k}(\beta^{[n_k]}) : \forall \ell \in [1:n_k], j \in [1:n_k], \\ & \left|\langle \phi^{[n_k,j]}_{(\beta)}, O_{n_k}(h_\ell)\phi^{[n_k,j]}_{(\beta)}\rangle - \int_0^1 h_\ell(x)\,dx\right| \le \|h_\ell\|_\infty \varepsilon(n_k)\Bigg\}.\end{aligned} \tag{3.4.21}$$

Then $\mathbb{P}[\Gamma^c_{n_k}] \le C_1 n_k^2 \exp(-C_2 n_k^{1/2}|I(n_k)|^{1/2}) \to 0$, and denseness shows that for sequences $(\widetilde{V}_{n_k})_k$ with $\widetilde{V}_{n_k} \in \Gamma_{n_k}$ and with eigenvectors denoted by $\widetilde{\phi}^{[n_k,j]}$, that $\langle \widetilde{\phi}^{[n_k,j_{n_k}]}, O_{n_k}(h)\widetilde{\phi}^{[n_k,j_{n_k}]}\rangle \to \int_0^1 h$ for all $f \in C([0,1])$ as well.

(d) To make the spectrum simple, we simply perturbed any degenerate eigenvalues while keeping them in the same bin.

(e) This follows from $\|U_{n_k} - V_{n_k}(\beta^{[n_k]})\| \le C \cdot \frac{2\pi}{\kappa(n_k)}$, since for any matrices U and V with entries $|\cdot| \le 1$,

$$\left||V_{xy}|^2 - |U_{xy}|^2\right| \le 2\,||V_{xy}| - |U_{xy}|| \le 2|V_{xy} - U_{xy}| \le 2\|V-U\|.$$

□

CHAPTER 4

The doubling map

4.1. The doubling map with any even n

Recall if $S : [0,1] \to [0,1]$ is the doubling map, then for any $n \in 2\mathbb{N}$, the $n \times n$ Markov matrix P_n along with a specific quantization U_n can be taken as in (2.1.2). For general maps S, in Theorem 2.1, we restricted to dimensions $n \in M_0\mathbb{Z}$ with $\widetilde{K}(n) \to \infty$. This ensured that enough powers of P_n behaved nicely with the partitions (Lemma 3.2). For the doubling map, $M_0 = 2$ and we can take all $n \in 2\mathbb{Z}$, not just those with the largest power of two dividing n tending to infinity. The statement is as follows (note that the quantization U_n does not have to be the orthogonal one in (2.1.2)).

Theorem 4.1 (pointwise Weyl law analogue for the doubling map)**.** *For $n \in 2\mathbb{N}$, let P_n be the matrix in (2.1.2), and let U_n satisfy $|(U_n)_{xy}|^2 = (P_n)_{xy}$. Denote the eigenvalues and eigenvectors of U_n by $(e^{i\theta_j^{(n)}})_j$ and $(\psi^{(n,j)})_j$ respectively for $j \in [n]$. Let $(I(n))$ be a sequence of intervals in $\mathbb{R}/(2\pi\mathbb{Z})$ satisfying*

$$|I(n)| \log n \to \infty, \quad \text{as } n \to \infty. \tag{4.1.1}$$

Then there is a sequence of subsets $G_n \subseteq [n]$ with sizes $\#G_n = n(1 - o(|I(n)|))$ so that for all $x \in G_n$,

$$\sum_{j:\theta^{(n,j)} \in I(n)} |\psi_x^{(n,j)}|^2 = \frac{|I(n)|}{2\pi}(1 + o(1)), \tag{4.1.2}$$

where the error term $o(1)$ depends only on n, $|I(n)|$, and $\#G_n$, and is independent of $x \in G_n$. Additionally, G_{n_k} can be chosen independent of $I(n_k)$ or $|I(n_k)|$.

The proof is the same as Theorem 2.1, except that Lemma 3.3, which bounds the number of nonzero entries on the diagonal of P_n^ℓ, is proved differently. To analyze the matrix powers

P_n^ℓ, instead of viewing them in terms of S^ℓ, we count paths of length ℓ in the directed graph associated with the Markov matrix P_n. The proof of Theorem 4.1 then follows from the following lemma and by replacing all instances of $\widetilde{K}(n)+1$ by $K := \lfloor \log_2 n \rfloor$ in the proof of Theorem 2.1.

Lemma 4.2 (number of nonzero entries for the doubling map)**.** *For $n \in 2\mathbb{N}$, let P_n be as in* (2.1.2) *and let $1 \le \ell \le K$. Consider the directed graph with n nodes $1, 2, \ldots, n$, whose adjacency matrix is $2P_n$. Then:*

(i) For any coordinates x, y, there is at most one path of length ℓ from x to y in the graph.

(ii) The diagonal of P_n^ℓ contains at most $2 \cdot 2^\ell$ nonzero entries.

(iii) In total, P_n^ℓ has exactly $n \cdot 2^\ell$ nonzero entries.

Proof. All possible paths starting at a node x and of length ℓ can be represented as paths in a binary tree of height ℓ with root node x. (Figure 4.1.) The nodes $1, 2, \ldots, n$ of the graph may be listed multiple times in the binary tree. If we always put the descendant $2x - 1$ on the left and put $2x$ on the right, then the list of nodes in each row of the tree will be consecutive increasing in $\mathbb{Z}/n\mathbb{Z}$. Thus if $\ell \le K := \lfloor \log_2 n \rfloor$, the ℓth row of the tree will contain $2^\ell \le n$ nodes, so that for any two nodes x and y, there is at most one path of length ℓ from x to y, proving part (i).

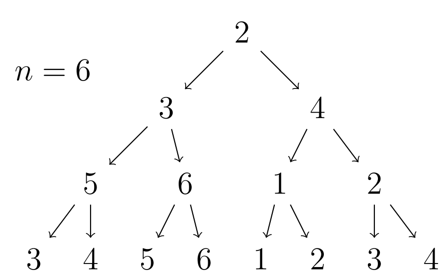

FIGURE 4.1. Start of a binary tree for $n = 6$ ($K = 2$). This tree describes all paths of length 3 that start at node 2.

Applying part (i), the total number of nonzero entries on the diagonal of P_n^ℓ is the total number of paths of length $\ell \le K$ with the same start and end point x. Similarly, the total number of nonzero entries in P_n^ℓ is the total number of length ℓ paths from any x to any y. The collection of all paths of length ℓ can be represented by the paths in a forest of n binary trees each of depth ℓ, one tree for each possible starting node $x \in [n]$. (Figure 4.2.) The ℓth row contains $2^\ell \cdot n$ numbers, showing part (iii).

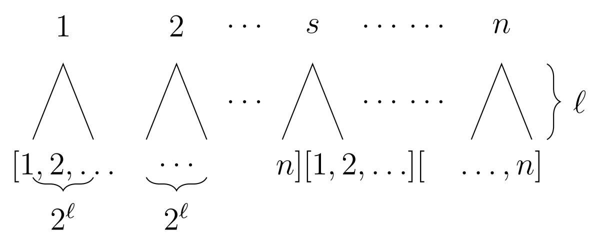

FIGURE 4.2. All paths of length ℓ as paths in a forest.

These $2^\ell \cdot n$ numbers at the bottom of the forest are 2^ℓ copies of the sequence $(1, 2, \ldots, n)$. To show (ii), we will show that for each copy C_j of $(1, 2, \ldots, n)$, there can be at most two paths with the same start and end point that end in this copy.

Let $F(j)$ be the set of starting nodes that have descendants in the jth copy C_j of $(1, 2, \ldots, n)$. (The last node in $F(j)$ may overlap with the first node in $F(j+1)$.) Consider just the paths that start in $F(j)$ and end in C_j, and suppose there is a length ℓ path $x \to x$. We claim that only either $x-1$ or $x+1$ in $F(j)$ can also have a loop of length ℓ. (See Figure 4.3.)

Let $L^\ell(x)$ be the left-most descendant of x in C_j, and let $R^\ell(x)$ be the right-most descendant of x in C_j.

(a) If $L^\ell(x) < x < R^\ell(x)$, then no other $y \in F(j)$ has a path $y \to y$. (Use $L^\ell(x+1) \geq x+2$ and $R^\ell(x-1) \leq x-2$, and then continue for the rest of $F(j)$ using $L^\ell(y+1) \geq L^\ell(y)+2$ and $R^\ell(y-1) \leq R^\ell(y)-2$.)
(b) Similarly, if $L^\ell(x) = x$, then only also $x-1$ has a path $x-1 \to x-1$.
(c) If $R^\ell(x) = x$, then only also $x+1$ has a path $x+1 \to x+1$.

Thus in total there are at most $2 \cdot 2^\ell$ paths of length ℓ that start and end at the same point, proving part (ii). □

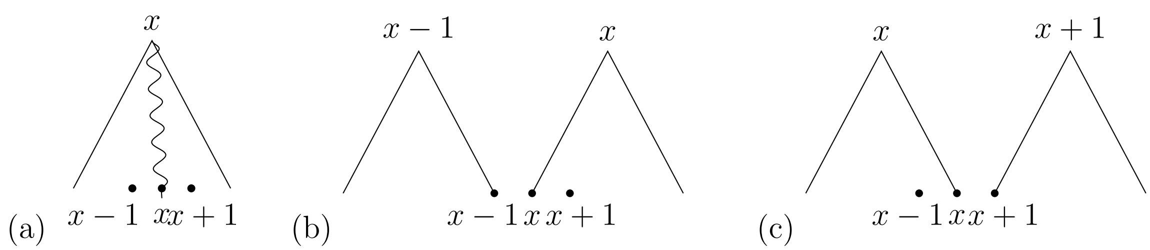

FIGURE 4.3. The possible cases if there is a loop $x \to x$.

4.2. The doubling map with $n = 2^K$

When $n = 2^K$, the corresponding graphs from the doubling map are the de Bruijn graphs on two symbols. Orbits in these graphs have been studied in the context of quantum chaos in for example [Tan00, Ler05, GO13, HH20]. In these dimensions, the particular matrices U_n from (2.1.2), despite coming from the doubling map, exhibit some behavior like that of integrable systems. Any choice of eigenbasis still satisfies the quantum ergodic theorem since the doubling map is ergodic, but the eigenvalues of U_n in these dimensions are degenerate and evenly spaced in the unit circle. As a result of the degeneracy, we will be able to show that random eigenbases look approximately Gaussian. This will follow from properties of the spectral projection matrix of an eigenspace combined with the results on random projections used in Section 3.4.

We start by showing the eigenvalues of U_n are $4K$th roots of 1 if K is even, and $4K$th roots of -1 if K is odd.

Proposition 4.3 (Repeating powers of U_n)**.** *Let U_n be as in* (2.1.2) *with $n = 2^K$. Then*

(a) $U_n^{4K} = (-1)^K I$.

(b) $U_n^r = (-1)^K (U_n^{4K-r})^T$*, for* $1 \le r \le 4K-1$*. More generally,* $U_n^r = (-1)^{Kw}(U_n^{4Kw-r})^T$*, for* $1 \le r \le 4Kw - 1$.

Proof. Part (b) follows from (a) and unitarity (orthogonality) of U_n. For part (a), view the doubling map on $[0,1]$ as the left bit shift on a sequence $\{0,1\}^{\mathbb{N}}$ corresponding to the binary expansion of $x \in [0,1]$. If we partition $[0,1]$ into 2^K atoms $E_i = (\frac{i}{2^K}, \frac{i+1}{2^K})$, $i = 0, \dots, 2^K - 1$, then we can identify atom E_i with the length K bit string corresponding to the binary expansion of $\frac{i}{2^K}$. Then $z \in E_i$ iff the first K digits of its binary expansion match the length K bit string for E_i. The Markov matrix P_n then takes an atom indexed by $i = (i_1, \ldots, i_K)$ and sends it to the atoms indexed by $(i_2, \ldots, i_{K-1}, 0)$ and $(i_2, \ldots, i_{K-1}, 1)$, the result of the left bit shift. Thus for $1 \leq \ell \leq K$, there is at most one length ℓ path from $i = (i_1, \ldots, i_K)$ to $j = (j_1, \ldots, j_K)$, which is described by the sequence $(i_1, \ldots, i_\ell, j_1, \ldots, j_K)$ and requires $i_{\ell+1}, i_{\ell+2}, \cdots, i_{\ell+(K-\ell)} = j_1, j_2, \ldots, j_{K-\ell}$. Note this recovers Lemma 3.1(d).

Now considering the signs in U_{2^K} and viewing the indices i, j as length K bit strings, if there is an edge $i \to j$, then

$$(U_{2^K})_{ij} = 2^{-1/2} \begin{cases} -1, & i_1 = 0, j_K = 1 \\ 1, & \text{else} \end{cases}.$$

Thus if there is a length K path $\tau : i \to j$, then

$$(U_{2^K}^K)_{ij} = (U_{2^K})_{i\tau_1}(U_{2^K})_{\tau_1\tau_2} \cdots (U_{2^K})_{\tau_{K-1}j} = 2^{-K/2} \prod_{m=1}^{K} (-1)^{(1-i_m)j_m}, \tag{4.2.1}$$

since $(\tau_m)_1 = i_m$ and $(\tau_m)_K = j_m$. This is the structure of a tensor product,

$$U_{2^K}^K = 2^{-K/2} \bigotimes_{m=1}^{K} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \tag{4.2.2}$$

so that

$$U_{2^K}^{2K} = \bigotimes_{m=1}^{K} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad U_{2^K}^{4K} = \bigotimes_{m=1}^{K} \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = (-1)^K I_{2^K}. \tag{4.2.3}$$

□

Remark 4.2.1. Proposition 4.3 can also be proved (although with significantly more effort) by analyzing paths in the corresponding de Bruijn graph. As in Section 4.1, possible paths can be described using trees, but the edges in the trees carry a sign to keep track of the negative signs in the matrix U_{2^K}.

Since the eigenvalues of U_{2^K} are $4K$-th roots of 1 or -1, instead of eigenvectors from eigenvalues in an interval $I(n) \subseteq \mathbb{R}/(2\pi\mathbb{Z})$ like in Theorem 2.1, we are just interested in all the eigenvectors from a single eigenspace. A stronger version of Theorem 2.1 for this specific case controls the spectral projection onto a single eigenspace.

Theorem 4.4 (Eigenspace projection when $n = 2^K$). *For $n = 2^K$, let U_n be as in (2.1.2), and let $P^{(n,j)}$ be the projection onto its jth eigenspace. Let $r(K) : \mathbb{N} \to \mathbb{N}$ be any function satisfying $r(K) < K$, $r(K) \to \infty$, and $K - r(K) - \log_2 K \to \infty$ as $K \to \infty$. Then there are sets $G_K \subseteq [1 : 2^K]$ and $GP_K \subseteq [1 : 2^K]^2$ with*

$$\#G_K \geq 2^K \left(1 - \frac{4}{2^{K-r(K)}}\right) = 2^K \left(1 - o\Big(\frac{1}{4K}\Big)\right) \tag{4.2.4}$$

$$\#GP_K \geq (2^K)^2 \left(1 - \frac{8}{2^{K-r(K)}}\right) = (2^K)^2 \left(1 - o\Big(\frac{1}{4K}\Big)\right), \tag{4.2.5}$$

such that the following hold as $K \to \infty$.

(a) For $x \in G_K$ and any j,

$$\left| \|P^{(n,j)} e_x\|_2^2 - \frac{1}{4K} \right| \leq \frac{1}{4K} \cdot 10 \cdot 2^{-r(K)/2}. \tag{4.2.6}$$

(b) For pairs $(x, y) \in GP_K$ and any j,

$$|\langle e_y, P^{(n,j)} e_x \rangle| \leq \frac{10 \cdot 2^{-r(K)/2}}{4K}. \tag{4.2.7}$$

Using (4.2.6) and summing over all x, we also obtain:

Corollary 4.5 (Eigenspace degeneracy). *The degeneracy of each eigenspace of U_{2^K} is $\frac{2^K}{4K}(1 + o(1))$.*

Returning to eigenvectors, Theorem 4.4(a) applied to Corollary 3.11 shows that taking a random basis within each eigenspace produces approximately Gaussian eigenvectors.

Theorem 4.6 (Gaussian eigenvectors when $n = 2^K$). *For $K \in \mathbb{N}$, let $(\psi^{(2^K,j)})_{j=1}^{2^K}$ be a random ONB of eigenvectors chosen according to Haar measure in each eigenspace of U_n.*

Let $\mu^{(2^K,j)} := \frac{1}{2^K}\sum_{x=1}^{2^K}\delta_{\sqrt{2^K}\psi_x^{(2^K,j)}}$ *be the empirical distribution of the scaled coordinates of* $\psi^{(2^K,j)}$*. Then there is* $\varepsilon(K)\to 0$ *so that for any bounded Lipschitz* $f:\mathbb{C}\to\mathbb{C}$ *with* $\|f\|_{\mathrm{Lip}}\le 1$*, as* $K\to\infty$*,*

$$\mathbb{P}\left[\max_{j\in[1:2^K]}\left|\int f(x)\,d\mu^{(2^K,j)}(x)-\mathbb{E}f(Z)\right|>\varepsilon(K)\right]\to 0, \tag{4.2.8}$$

where $Z\sim N_{\mathbb{C}}(0,1)$*. In particular, each* $\mu^{(2^K,j)}$ *converges weakly in probability to* $N_{\mathbb{C}}(0,1)$ *as* $K\to\infty$*.*

Proof (of Theorem 4.6). Theorem 4.4 shows we can take $d_\ell=\frac{n}{4K}$, and A in Corollary 3.11 to be $o(\frac{n}{4K})$. Then similar to Subsection 3.4.2, take

$$\varepsilon(K)=\max\left(\frac{4A+6}{(d_\ell-A)-1},\frac{K^{1/4}}{n^{1/4}}\right)\to 0, \tag{4.2.9}$$

and note that $6n\exp\left(-\frac{n^{1/2}}{K^{1/2}}\frac{(1-o(1))}{2^8}\right)\to 0.$ □

The rest of this section is the proof of Theorem 4.4.

4.2.1. Polynomial for eigenspace projection. Instead of using trigonometric polynomials to approximate the spectral projection matrix like in the proof of Theorem 2.1, we use a polynomial with zeros at $4K$-th roots of 1 or -1 to get exact formulas. Let U_n be as in (2.1.2) with $n=2^K$. First consider K even, so $U_n^{4K}=I$ and the eigenvalues of U_n are $4K$-th roots of unity. Since $\frac{x^{4K}-1}{x-1}=1+x+x^2+\cdots+x^{4K-1}$ is zero at all $4K$-th roots of unity except for $x=1$, the polynomial

$$p_{K,j}(x)=1+\sum_{\ell=1}^{4K-1}(e^{-2\pi ij/(4K)})^\ell x^\ell \tag{4.2.10}$$

is zero at all $4K$-th roots of unity except for $e^{2\pi ij/(4K)}$, where it takes the value $4K$. Writing

$$U_n=\sum_{\alpha=0}^{4K-1}e^{2\pi i\alpha/(4N)}\sum_{\lambda=e^{2\pi i\alpha/(4K)}}|\psi^{(\lambda)}\rangle\langle\psi^{(\lambda)}|,$$

the spectral projection onto the eigenspace of $e^{2\pi ij/(4K)}$ is

$$P^{(n,j)} = \sum_{\lambda=e^{2\pi ij/(4K)}} |\psi^{(\lambda)}\rangle\langle\psi^{(\lambda)}| = \frac{1}{4K}\cdot p_{K,j}(U_n) = \frac{1}{4K}\left(I + \sum_{\ell=1}^{4K-1}(e^{-2\pi ij/(4K)})^\ell U_n^\ell\right).$$

If K is odd, then $U_n^{4K} = -I$ and the eigenvalues of U_n are $4K$-th roots of -1. These are $\exp(i\frac{\pi}{4K}+\frac{2\pi ij}{4K})$ for $j \in [0 : 4K-1]$. For notational convenience, let $\gamma(K) := \begin{cases} e^{i\pi/(4K)}, & K \text{ odd} \\ 1, & K \text{ even}\end{cases}$, so we can write for any $K \in \mathbb{N}$,

$$P^{(n,j)} := \sum_{\lambda=e^{2\pi ij/(4K)}\gamma(K)} |\psi^{(\lambda)}\rangle\langle\psi^{(\lambda)}| = \frac{1}{4K}\left(I + \sum_{\ell=1}^{4K-1}(e^{-2\pi ij/(4K)}\overline{\gamma(K)})^\ell U_n^\ell\right). \tag{4.2.11}$$

4.2.2. Powers of U_n. To estimate the matrix elements of (4.2.11), we need some properties on the powers of U_n. Since by Proposition 4.3(b), $U_n^{2K+r} = (-1)^K(U_n^{2K-r})^T$ for $r = 0, \ldots, 2K-1$, to understand all the powers $U_n, U_n^2, \ldots, U_n^{4K-1}$, it is enough to know the powers U_n^m for $m \in [1 : K] \cup [2K : 3K]$. We will only need to know where the entries of U_n^m are nonzero, which follows from matrix multiplication:

Lemma 4.7 (Powers up to K)**.** *Let $n = 2^K$. For $m \le K$, let $\mathcal{A}_m$ be the set of real $n \times n$ matrices A such that*

$$|A_{ij}| = \begin{cases} 1, & j \in \{2^m i, 2^m i - 1, \ldots, 2^m i - (2^m - 1)\} \mod 2^K \\ 0, & else \end{cases}.$$

$\mathcal{A}_m$ consists of matrices whose nonzero entries are ± 1 arranged in 2^m descending "staircases" with steps of length 2^m. Then for $m \le K-1$ and $A \in \mathcal{A}_m$,

$$A \cdot \sqrt{2}U_n \in \mathcal{A}_{m+1}.$$

In particular, since $2^{1/2}U_n \in \mathcal{A}_1$, then for $m \le K$,

$$2^{m/2}U_n^m \in \mathcal{A}_m.$$

Between $2K$ and $3K$, U_n^m has a flipped staircase structure:

Corollary 4.8 (Powers from $2K$ to $3K$)**.** *Let $n = 2^K$. For $m \leq K$, let $\mathcal{B}_m$ be the set of $n \times n$ matrices B such that the matrix A defined by $A_{ij} := B_{(n-i)j}$ is in $\mathcal{A}_m$. Then for $m \leq K-1$ and $B \in \mathcal{B}_m$,*

$$B \cdot \sqrt{2} U_n \in \mathcal{B}_{m+1}.$$

In particular, using that U_n^{2K} is a "flipped diagonal" matrix with nonzero entries ± 1, so that $2^{1/2} U_n^{2K+1} \in B_1$, then for $m \in [1 : K]$,

$$2^{m/2} U_n^{2K+m} \in \mathcal{B}_m.$$

Proof. If $A_{ij} = B_{(n-i)j}$, then $(BU_n)_{(n-i)j} = \sum_{\ell=1}^n A_{i\ell}(U_n)_{\ell j} = (AU_n)_{ij}$, and since $\sqrt{2} \cdot AU_n \in \mathcal{A}_{m+1}$, then $\sqrt{2} \cdot BU_n \in \mathcal{B}_{m+1}$. That U_n^{2K} is a "flipped diagonal" matrix with nonzero entries ± 1 along the flipped diagonal $(i, n-i)$ follows from equation (4.2.3). Then the matrix A defined by

$$A_{ij} := (U_n^{2K} \cdot 2^{1/2} U_n)_{(n-i)j} = 2^{1/2} \sum_{\ell=1}^n \pm \delta_{i,\ell} (U_n)_{\ell j} = (\pm 1) 2^{1/2} (U_n)_{ij}$$

is in $\mathcal{A}_1$ so $2^{1/2} U_n^{2K+1} \in \mathcal{B}_1$. □

4.2.3. Removing potentially bad points. This mirrors Subsection 3.2.3 from the proof of Theorem 2.1, although due to the structure of U_n^ℓ here, we consider $1 \leq \ell \leq 4K$ instead of just $1 \leq \ell \leq \widetilde{K} + 1$. (Figure 4.4.)

Let the set of potentially bad coordinates be

$$B_K := \{x \in [n] : (U_n^m)_{xx} \neq 0 \text{ for some } m \in [1 : r(K)] \cup [2K - r(K) : 2K]\}. \tag{4.2.12}$$

Indexing the 2^K atoms of $[0,1]$ by length K bit strings as in the proof of Proposition 4.3, we see that for $1 \leq \ell \leq K$, the entry $(U_n^\ell)_{xx}$ is nonzero iff x is of the form $(x_1, \ldots, x_\ell, x_1, \ldots, x_\ell, x_1, \ldots)$, that is x corresponds to a periodic orbit of length ℓ. There are 2^ℓ choices for the sequence $(x_1, x_2, \ldots, x_\ell)$, so the diagonal of U_n^ℓ contains exactly 2^ℓ nonzero entries for $\ell \in [1 : r(K)]$. Additionally, by the staircase structure from Corollary 4.8, the

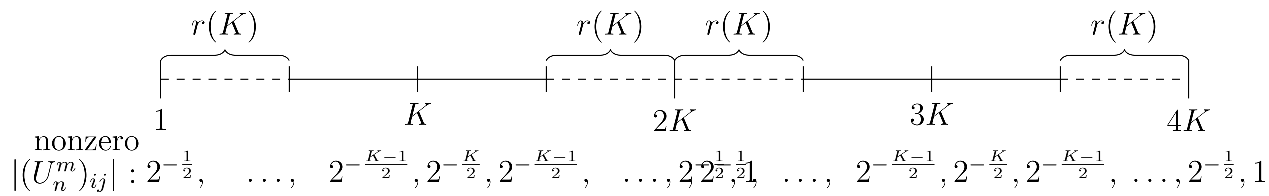

FIGURE 4.4. Eliminating bad coordinates in regions where the nonzero entries of U_n^m are large. By Proposition 4.3(b), we only need to consider powers up to $2K$ in the definition of B_K, since the powers reflect across $2K$.

diagonal of U_n^ℓ has at most $2^{2K-\ell}$ nonzero entries for $\ell \in [2K - r(K) : 2K - 1]$. Thus

$$\#B_K \leq 2 \sum_{\ell=1}^{r(K)} 2^\ell = 4(2^{r(K)} - 1) = o(2^K/K). \tag{4.2.13}$$

Let the set of good coordinates be $G_K := [n] \setminus B_K$. For $x \in G_K$, then $(U_n^\ell)_{xx} = 0$ for $\ell \in [1 : r(K)] \cup [2K - r(K) : 2K]$ (and also for $\ell \in [2K : 2K + r(K)] \cup [4K - r(K) : 4K]$), so that for any $j \in [0 : 4K - 1]$,

$$\begin{aligned} \|P^{(n,j)} e_x\|_2^2 = \frac{1}{4K} \Bigg(1 + \sum_{\ell=r(K)+1}^{2K-r(K)-1} (e^{-2\pi i j/(4K)} \overline{\gamma(K)})^\ell (U_n^\ell)_{xx} + \\ + \sum_{\ell=2K+r(K)+1}^{4K-r(K)-1} (e^{-2\pi i j/(4K)} \overline{\gamma(K)})^\ell (U_n^\ell)_{xx} \Bigg) \end{aligned}$$

$$= \frac{1}{4K} (1 + \mathcal{O}(2^{-r(K)/2})), \tag{4.2.14}$$

since

$$\left| \sum_{\ell=r(K)+1}^{2K-r(K)-1} (e^{-2\pi i j/(4K)} \overline{\gamma(K)})^\ell (U_n^\ell)_{xx} \right| \leq 2 \sum_{\ell=r(K)+1}^{K} 2^{-\ell/2} \leq 10 \cdot 2^{-r(K)/2}, \tag{4.2.15}$$

and similarly for the second sum. This proves (4.2.6).

4.2.4. Removing potentially bad pairs of points. Let the set of potentially bad pairs of coordinates be

$$BP_K := \{(x,y) \in [n]^2, x \neq y : (U_n^\ell)_{xy} \neq 0 \text{ for some } \ell \in [1 : r(K)] \cup [2K - r(K) : 2K + r(K)] \cup [4K - r(K) : 4K - 1]\}. \tag{4.2.16}$$

The matrix U_n^ℓ contains $2^\ell \cdot n$ nonzero entries (n entries in each staircase and 2^ℓ staircases) for $\ell \in [1 : r(K)]$, and $2^{2K-\ell} \cdot n$ nonzero entries for $\ell \in [2K - r(K) : 2K - 1]$ (and the same for flipping ℓ across $2K$). Then

$$\#BP_K \leq 4 \sum_{\ell=1}^{r(K)} 2^\ell \cdot n = 8(2^{r(K)} - 1) \cdot n = o(n^2), \tag{4.2.17}$$

and for good pairs $(x,y) \in GP_K := ([n]^2 \setminus \{(x,y) : x = y\}) \setminus BP_K$,

$$|\langle e_y, P^{(n,j)} e_x \rangle| = \left| \frac{1}{4K} \left(\left(\sum_{m=r(K)+1}^{2K-r(K)-1} + \sum_{m=2K+r(K)+1}^{4K-r(K)-1} \right) (e^{-2\pi i j/(4K)} \overline{\gamma(K)})^m (U_n^\ell)_{xy} \right) \right|$$

$$\leq 10 \cdot \frac{2^{-r(K)/2}}{4K}, \tag{4.2.18}$$

by the same estimates as before. □

CHAPTER 5

Additional remarks

5.1. Coordinates that fail the pointwise Weyl law

We give a specific example where not all coordinates satisfy the pointwise Weyl law (2.3.2). For $n \in 2\mathbb{N}$, let

$$U_n = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 & & & & \\ & & 1 & -1 & & \\ & & & & \ddots & \\ & & & & 1 & -1 \\ 1 & 1 & & & & \\ & & 1 & 1 & & \\ & & & & \ddots & \\ & & & & 1 & 1 \end{pmatrix}.$$

Letting $P_I^{(n)}$ be the spectral projection matrix of U_n onto the arc $I = [-\pi/2, \pi/2]$ on the unit circle, we will show that

$$(P_I^{(n)})_{11} \geq 0.89182655 + o(1) \neq \frac{1}{2}(1 + o(1)). \tag{5.1.1}$$

Thus for $I(n) = [-\pi/2, \pi/2]$, the sequence of coordinates $(x_n)_n$ with just $x_n = 1$ does not satisfy the pointwise Weyl law. Note the coordinate $x = 1$ was one of the "bad" points that was removed during the proof of the pointwise Weyl law, since it always has the very short periodic loop consisting of just itself.

To approximate $(P_I^{(n)})_{11}$ from below, we use the piecewise linear approximation h_Δ on $\mathbb{R}/(2\pi\mathbb{Z})$ in Figure 5.1, defined by

$$h_\Delta(x) = \begin{cases} 1, & -\frac{\pi}{2} + \Delta \leq x \leq \frac{\pi}{2} - \Delta \\ \frac{1}{\Delta}\left(x + \frac{\pi}{2}\right), & -\frac{\pi}{2} \leq x \leq \frac{-\pi}{2} + \Delta \\ -\frac{1}{\Delta}\left(x - \frac{\pi}{2}\right), & \frac{\pi}{2} - \Delta \leq x \leq \frac{\pi}{2} \\ 0, & |x| \geq \frac{\pi}{2} \end{cases}.$$

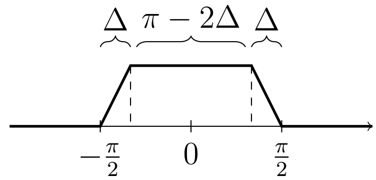

FIGURE 5.1. Plot of h_Δ.

Since this $I(n) = [-\pi/2, \pi/2]$ is not shrinking, we don't need further smoothness of the approximation, and the piecewise linear h_Δ has Fourier coefficients that are easy to work with. We only need continuity and absolutely summable Fourier coefficients, so that the Fourier series converges uniformly to h_Δ. For convenience we also use the same notation h_Δ or $\chi_{[-\pi/2,\pi/2]}$ to denote the corresponding function on the unit circle in $\mathbb{C}$ (via $\mathbb{R}/2\pi\mathbb{Z} \ni t \leftrightarrow e^{it} \in \mathbb{S}^1$). Since pointwise $h_\Delta(t) \le \chi_{[-\pi/2,\pi/2]}(t)$ for any $\Delta \ge 0$, then by the spectral theorem $(h_\Delta(U_n))_{xx} \le \chi_{[-\pi/2,\pi/2]}(U_n)_{xx} = (P_I^{(n)})_{xx}$ for any coordinate $x \in [n]$.

To approximate $h_\Delta(U_n)$, we compute its Fourier coefficients $(\widehat{h}_\Delta)_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} h_\Delta(x) e^{-ijx}\,dx$,

$$(\widehat{h}_\Delta)_j = \frac{2}{\pi j^2 \Delta} \sin\left(\frac{j(\pi - \Delta)}{2}\right) \sin\left(\frac{j\Delta}{2}\right), \quad j \neq 0 \tag{5.1.2}$$

$$(\widehat{h}_\Delta)_0 = \frac{\pi - \Delta}{2\pi}. \tag{5.1.3}$$

Since the Fourier coefficients are absolutely summable, the partial sums $\sum_{j\in\mathbb{Z}} (\widehat{h}_\Delta)_j e^{ijx}$ converge uniformly to h_Δ, with the Kth partial sum $S_K h_\Delta(x) := \sum_{|j|\le K} (\widehat{h}_\Delta)_j e^{ijx}$ having error bound

$$\|S_K h_\Delta - h_\Delta\|_\infty \le \sum_{j=K+1}^{\infty} \frac{4}{\pi j^2 \Delta} \le \frac{4}{\pi K \Delta}. \tag{5.1.4}$$

As long as $\Delta \gg K^{-1}$, this is $o(1)$, and then

$$\begin{aligned}
|\langle y|S_K h_\Delta(U_n) - h_\Delta(U_n)|x\rangle| &= \left| \sum_{j=1}^{n} \left(S_K h_\Delta(e^{i\theta^{(n,j)}}) - h_\Delta(e^{i\theta^{(n,j)}}) \right) \langle y|\psi^{(n,j)}\rangle \langle \psi^{(n,j)}|x\rangle \right| \\
&\le \|S_K(h_\Delta) - h_\Delta\|_\infty \left(\sum_{j=1}^{n} |\psi_y^{(n,j)}|^2 \right)^{1/2} \left(\sum_{j=1}^{n} |\psi_x^{(n,j)}|^2 \right)^{1/2} \\
&= o(1).
\end{aligned}$$

So then

$$(P_I^{(n)})_{11} \geq (h_\Delta(U_n))_{11} = ((S_K h_\Delta)U_n)_{11} + o(1).$$

As usual, we take $K = \lfloor \log_2 n \rfloor$, and consider

$$(5.1.5)\quad (S_K h_\Delta)(U_n)_{11} = \frac{\pi - \Delta}{2\pi} + \sum_{j=1}^{K} \frac{2}{\pi j^2 \Delta} \sin\left(\frac{j(\pi-\Delta)}{2}\right) \sin\left(\frac{j\Delta}{2}\right) \left((U_n)_{11}^j + (U_n)_{11}^{-j}\right).$$

Take $K^{-1} \ll \Delta \ll K^{-1/2}$, for example $\Delta = K^{-3/4}$. We split up the sum over j in (5.1.5) into two regions, first from $j = 1$ to $\sqrt{K}$, and then from $\sqrt{K}+1$ to K. In the first region, $j\Delta \leq \sqrt{K}\Delta \ll 1$, so we can Taylor expand the sine terms and evaluate the sum. In the second region, the exponential decay from $(U_n)_{11}^j = 2^{-j/2}$ (for $j \leq \lfloor \log_2 n \rfloor$, there is only the path $1 \to 1 \to \cdots \to 1$ that starts and ends at 1 and has length j) will make the sum $o(1)$ as $K \to \infty$.

For $j\Delta \ll 1$, we have $\sin(j\Delta/2) = \frac{j\Delta}{2} + \mathcal{O}(j^2\Delta^2)$ and

$$\sin\left(\frac{j(\pi-\Delta)}{2}\right) = \begin{cases} -\frac{j\Delta}{2} + \mathcal{O}(j^3\Delta^3), & j = 0 \mod 4 \\ 1 - \frac{j^2\Delta^2}{4} + \mathcal{O}(j^4\Delta^4), & j = 1 \mod 4 \\ \frac{j\Delta}{2} + \mathcal{O}(j^3\Delta^3), & j = 2 \mod 4 \\ -1 + \frac{j^2\Delta^2}{4} + \mathcal{O}(j^4\Delta^4), & j = 3 \mod 4 \end{cases}.$$

Thus

$$\begin{aligned}
(S_K h_\Delta)(U)_{11} &= \frac{1}{2} - \mathcal{O}(\Delta) + \sum_{j=1}^{\sqrt{K}} \frac{2}{\pi j^2 \Delta} \sin\left(\frac{j(\pi-\Delta)}{2}\right) \sin\left(\frac{j\Delta}{2}\right) 2 \cdot 2^{-j/2} + \sum_{j=\sqrt{K}+1}^{K} \mathcal{O}(2^{-j/2}) \\
&= \frac{1}{2} + \sum_{\substack{j=1 \\ j \text{ odd}}}^{\sqrt{K}} \frac{2}{\pi j} \left((-1)^{((j \bmod 4)-1)/2} + \mathcal{O}(j\Delta)\right) \cdot 2^{-j/2} + \sum_{\substack{j=1 \\ j \text{ even}}}^{\sqrt{K}} \frac{\mathcal{O}(j^2\Delta^2)}{j^2\Delta} 2^{-j/2} + o(1) \\
&= \left(\frac{1}{2} + \frac{2}{\pi} \sum_{\substack{j=1 \\ j=1 \mod 4}}^{\sqrt{K}} \left(\frac{1}{j 2^{j/2}} - \frac{1}{(j+2)2^{(j+2)/2}}\right)\right) + o(1) \\
&= \left(\frac{1}{2} + \frac{1}{\pi} \sum_{\ell=0}^{(\sqrt{K}-1)/4} \frac{5+4\ell}{(1+4\ell)(3+4\ell)2^{(1+4\ell)/2}}\right) + o(1).
\end{aligned}$$

Numerically,

$$\frac{1}{\pi}\sum_{\ell=0}^{\infty}\frac{5+4\ell}{(1+4\ell)(3+4\ell)2^{(1+4\ell)/2}} \approx 0.39182655,$$

so that

$$(P_I^{(n)})_{11} \geq (S_K h_\Delta)(U)_{11} + o(1) \geq 0.89182655 - o(1) \neq \frac{1}{2}(1+o(1)). \tag{5.1.6}$$

Remark 5.1.1. A similar statement can be shown with phases $e^{i\Phi}U_n$ for the interval $[t_0 - \frac{\pi}{2}, t_0 + \frac{\pi}{2}]$ for a t_0 depending on Φ. The only difference in (5.1.5) is that the $\big((U_n)_{11}^j + (U_n)_{11}^{-j}\big)$ term becomes $\big(e^{-it_0 j}(e^{i\Phi}U_n)_{11}^j + e^{it_0 j}(e^{i\Phi}U_n)_{11}^{-j}\big)$. For $j \leq \lfloor \log_2 n \rfloor$, $(e^{i\Phi}U_n)_{11}^j = e^{ij\Phi_1}2^{-j/2}$ and $(e^{i\Phi}U_n)_{11}^{-j} = e^{-ij\Phi_1}2^{-j/2}$, so taking $t_0 = \Phi_1$ reduces this Fourier series back to just (5.1.5).

5.2. Logarithmic factors in ℓ^p norms

5.2.1. From pointwise Weyl law. The pointwise Weyl law gives some logarithmic factors on sums of ℓ^p norm bounds of eigenvectors for $2 < p < \infty$. (Since the pointwise Weyl law doesn't have to hold for all coordinates x, we do not obtain the immediate ℓ^∞ estimate.) Fix $2 < p < \infty$. By Hölder and $|\psi_x^{(n,j)}| \leq 1$, there are the general bounds

$$\frac{1}{n^{p/2-1}} \leq \|\psi^{(n,j)}\|_p^p \leq 1,$$

with the lower bound corresponding to a completely delocalized vector and the upper bound corresponding to a completely localized vector. For $p > 2$, ℓ^p norm comparison gives

$$\Big(\sum_{j\in J}|\psi_x^{(n,j)}|^p\Big)^{1/p} \leq \Big(\sum_{j\in J}|\psi_x^{(n,j)}|^2\Big)^{1/2}.$$

To average $\|\psi^{(n,j)}\|_p^p$ over a window $I(n)$, apply the pointwise Weyl law, choosing the function $r(\widetilde{K}(n))$ in (3.2.2) sufficiently small so that the set $B_n = [n] \setminus G_n$ of "bad coordinates" is small enough that the resulting error term below is lower order,[1]

$$\begin{aligned}\frac{1}{\#\{j : \theta^{(n,j)} \in I(n)\}} \sum_{j:\lambda^{(n,j)}\in I(n)} \|\psi^{(n,j)}\|_p^p &= \frac{2\pi(1+o(1))}{n|I(n)|} \sum_{j:\lambda^{(n,j)}\in I(n)} \sum_{x=1}^{n} |\psi_x^{(n,j)}|^p \\ &\leq \frac{2\pi}{|I(n)|}\left(\frac{|I(n)|}{2\pi}\right)^{p/2}(1+o(1)) + \frac{2\pi(1+o(1))\cdot \#B_n}{n|I(n)|} \\ &= \left(\frac{|I(n)|}{2\pi}\right)^{p/2-1}(1+o(1)).\end{aligned}$$

At best this decays a bit slower than something like $(\log n)^{1-p/2}$ by choosing $|I(n)|$ close to smallest allowed size $|I(n)| \gg \widetilde{K}(n)^{-1}$, and assuming n is chosen so $\widetilde{K}(n) \sim \log n$. This is still very far from what we expect from numerics and the random wave conjecture, even with an average over many eigenvectors.

5.2.2. From small-scale quantum ergodicity. Logarithmic factors in ℓ^p norms can be obtained using the small-scale ergodicity argument from [HR16], this time in a limiting density one set without having to average over eigenvectors in an interval $I(n)$, though possibly with a worse power of $\log n$. The setting in [HR16] is much more complicated on manifolds, but the small-scale quantum ergodicity idea carries over to this simpler discrete case if we have an estimate on decay of correlations or the rate of convergence in the L^2-ergodic theorem for certain functions.[2] For example, for S the doubling map and $A \subseteq [0,1]$ an interval, direct computation shows,

$$\left|\mu(S^{-t}(A)\cap A) - \mu(A)^2\right| = \left|\int_0^1 \chi_A(2^t x)\chi_A(x)\,dx - |A|^2\right| \leq C|A|2^{-t}, \tag{5.2.1}$$

[1] For example, if $r(\widetilde{K}(n)) = \log_{L_0} n$, then $\#B_n \leq C\widetilde{K}(n)$, so $\frac{\#B_n}{n|I(n)|} \leq \frac{C\widetilde{K}(n)^2}{n} \leq \frac{C(\log_{L_0} n)^2}{n} \ll \frac{1}{(\log n)^{p/2-1}} \ll |I(n)|^{p/2-1}$.

[2] For general functions, nothing can be said about the rate of convergence [KP81], but for small-scale ergodicity we only need to consider indicator functions of intervals (or possibly their smooth approximations).

which then implies

$$\int_0^1 \left| \frac{1}{T}\sum_{t=0}^{T-1} \chi_A(2^t x) - |A| \right|^2 dx = \frac{1}{T^2}\sum_{t,r=0}^{\infty} \int_0^1 \left(\chi_A(2^{|t-r|}x)\chi_A(x) - |A|^2\right) dx$$
$$\leq \frac{1}{T^2}\sum_{t,r=0}^{T-1} 2^{-|t-r|} C|A| \leq \frac{C|A|}{T}. \tag{5.2.2}$$

From the above one gets a quantitative decay of the quantum variance in the quantum ergodic theorem, which allows taking n-dependent intervals $A = A_n$. Taking $A_n = [\frac{x_n}{n}, \frac{x_n+\alpha_n}{n}] \subseteq [0,1]$ for some $x_n \in [0 : n-1]$ and $\alpha_n \in [1 : n-1]$, and using [BKS07, Lemma 5], which holds for any L^2 observable, shows for $T = \lfloor \log_2 n \rfloor$,

$$V_n(\chi_{A_n}) := \frac{1}{n}\sum_{j=1}^{n} \left| \langle \psi^{(n,j)}, O_n(\chi_{A_n})\psi^{(n,j)} \rangle - |A_n| \right|^2$$
$$= \frac{1}{n}\sum_{j=1}^{n} \left| \sum_{x \in nA_n} |\psi_x^{(n,j)}|^2 - |A_n| \right|^2$$
$$\leq \int_0^1 \left| \frac{1}{T}\sum_{t=0}^{T-1} \chi_{A_n}(2^t x) - |A_n| \right|^2 dx \leq \frac{C|A_n|}{T}.$$

Then we follow the standard the proof of quantum ergodic theorem, except we have to choose $|A_n|$ and error terms more carefully due to the changing $|A_n|$ and the need for more and more intervals of length $|A_n|$ to cover $[0,1]$, similar to [HR16, §3]. The Chebyshev–Markov bound shows

$$\frac{1}{n} \cdot \#\left\{ j : \left| \sum_{x \in nA_n} |\psi_x^{(n,j)}|^2 - |A_n| \right| > \varepsilon_n |A_n| \right\} \leq \frac{1}{\varepsilon_n^2 |A_n|^2} V_n(\chi_{A_n})$$
$$\leq \frac{C}{\varepsilon_n^2} \cdot \frac{1}{|A_n| \lfloor \log_2 n \rfloor}. \tag{5.2.3}$$

Take $|A_n| = \frac{\alpha_n}{n} \to 0$ but satisfying $\alpha_n \gg \frac{n}{\sqrt{\log n}}$ so that $|A_n| \gg \frac{1}{\sqrt{\log n}}$, and set $\varepsilon_n^2 = |A_n|^{-1} \lfloor \log_2 n \rfloor^{-1/2}$ which tends to zero as $n \to \infty$. Cover $[0,1]$ with $b_n := \left\lceil \frac{1}{|A_n|} \right\rceil$ intervals $A_n^{(1)}, \ldots, A_n^{(b_n)}$, each of the form $[\frac{x_k}{n}, \frac{x_k+\alpha_n}{n}]$ for $x_k \in \mathbb{Z}/n$, and of the same length $|A_n| = \frac{\alpha}{n}$.

Let $\Lambda_n(A_n^{(k)}) := \left\{ j : \left| \sum_{x \in nA_n^{(k)}} |\psi_x^{(n,j)}|^2 - |A_n^{(k)}| \right| \leq \varepsilon_n |A_n^{(k)}| \right\}$ be the set of good indices $j \subseteq [n]$ associated with $A_n^{(k)}$. With the choice of $|A_n|$ and ε_n, (5.2.3) implies

$$\#\Lambda_n(A_n^{(k)})^c \leq \frac{Cn}{\lfloor \log_2 n \rfloor^{1/2}}.$$

Take

$$\widetilde{\Lambda}_n := \Lambda_n(A_n^{(1)}) \cap \Lambda_n(A_n^{(2)}) \cap \cdots \cap \Lambda_n(A_n^{(b_n)}),$$

which by union bound has size

$$\#\widetilde{\Lambda}_n \geq n - \frac{C}{|A_n|} \cdot \frac{n}{\lfloor \log_2 n \rfloor^{1/2}} = n(1 - \mathcal{O}(\varepsilon_n)).$$

For any $j \in \widetilde{\Lambda}_n$ and $k \in [1 : b_n]$,

$$\sum_{x \in nA_n^{(k)}} |\psi_x^{(n,j)}|^2 = |A_n|(1 + \mathcal{O}(\varepsilon_n)),$$

with the error term independent of k. Thus for any sequence $(j_n)_n$ with $j_n \in \widetilde{\Lambda}_n$,

$$\begin{aligned} \|\psi^{(n,j_n)}\|_p^p = \sum_{x=1}^{n} |\psi_x^{(n,j_n)}|^p \leq \sum_{k=1}^{b_n} \sum_{x \in nA_n^{(k)}} |\psi_x^{(n,j_n)}|^p \leq \sum_{k=1}^{b_n} \left(\sum_{x \in nA_n^{(k)}} |\psi_x^{(n,j_n)}|^2 \right)^{p/2} \\ \leq \frac{1}{|A_n|} \left(|A_n|(1 + \mathcal{O}(\varepsilon_n)) \right)^{p/2} = |A_n|^{p/2-1}(1 + o(1)), \end{aligned}$$

which decays a bit slower than something like $(\sqrt{\log n})^{1-p/2}$ by choosing $|A_n| \gg \frac{1}{\sqrt{\log n}}$ close to the smallest allowable by this method. We had to take $|A_n| \gg \frac{1}{\sqrt{\log n}}$ to ensure $\#\widetilde{\Lambda}_n^c \leq b_n \cdot \#\Lambda_n(A_n^{(k)}) \sim \frac{1}{|A_n|} \frac{Cn}{\varepsilon_n^2 |A_n| \log n} = o(n)$.

5.3. Other miscellaneous remarks

(i) The condition $\widetilde{K}(n)|I(n)| \to \infty$ for the pointwise Weyl law is essentially optimal without further restrictions, as the example for the doubling map when $n = 2^K$ shows (Section 4.2). In this case, if $\frac{|I(n)|}{2\pi} < \frac{1}{4K}$, then one can take an interval that avoids the spectrum entirely and thus produces a projection matrix filled with zeros.

(ii) The same method to estimate the diagonal entries of the projection matrix $P_{I(n)}$ can also be used to estimate the off-diagonal entries of $P_{I(n)}$. In this case, the constant term in the Fourier series expansion is zero, and then one can use similar arguments to show that the higher order terms are small. Alternatively, one can also obtain some bounds using that by the Weyl law, $\sum_{x,y\in[n]}(P_{I(n)})^2_{xy} = \operatorname{tr} P^2_{I(n)} = \operatorname{tr} P_{I(n)} = \frac{n|I(n)|}{2\pi}(1+o(1))$.

(iii) For a Markov transition matrix P_n, one usually views the indices of P_n as corresponding to vertices in a directed (pseudo)graph. (Loops are allowed but multi-arcs are not.) However, one can sometimes view the indices of P_n as corresponding to directed *edges* in a digraph instead, similar to a vertex scattering matrix on a quantum graph. In this case, a path ij corresponds to edge i followed by edge j, which can describe scattering from edge i, through the shared vertex, and out through edge j.

However, not every Markov transition matrix P_n corresponds naturally to a Markov chain on the directed edges of a digraph. For a directed pseudograph $G = (V, E)$, the *line digraph* $L(G)$ is the directed pseudograph with vertex set E and an edge between $e, f \in E$ iff e and f are incident (with correct orientations) in G. A directed pseudograph is called a line digraph if it is some directed pseudograph's line digraph. From the following classification of line digraphs, not all digraphs can be realized as the line digraph of a digraph:

Theorem 5.1 (Line digraph classification, see Theorem 11.2.3 in the book [GS18])**.** *Let D be a directed pseudograph with no multiple arcs. Then D is a line digraph iff every pair of rows (or of columns) of its adjacency matrix M are either identical or orthogonal.*

In particular, the digraphs from the doubling map on $[0, 1]$ are line digraphs. However, the digraphs from the four-legs map are not.

5.4. Numerics for other interval maps

Numerics for the doubling map and one of its unitary quantizations were displayed in Figures 2.1 and 2.2. Similar results appear for the "four legs map" F (drawn in Figure 2.3) and the tripling map $T(x) = 3x \pmod 1$,

$$F(x) = \begin{cases} 2x, & 0 \leq x < \frac{1}{4} \\ 4\left(x - \frac{1}{4}\right), & \frac{1}{4} \leq x < \frac{1}{2} \\ 4\left(x - \frac{1}{2}\right), & \frac{1}{2} \leq x < \frac{3}{4} \\ 2x - 2, & \frac{3}{4} \leq x \leq 1 \end{cases}, \qquad T(x) = \begin{cases} 3x, & 0 \leq x < \frac{1}{3} \\ 3\left(x - \frac{1}{3}\right), & \frac{1}{3} \leq x < \frac{2}{3} \\ 3\left(x - \frac{2}{3}\right), & \frac{2}{3} \leq x \leq 1 \end{cases}.$$

For the four legs map F, the corresponding Markov matrix $P_n^{(F)}$ and a particularly simple unitary (orthogonal) quantization $U_n^{(F)}$ given in [PZK01, Appendix B] for $n \in 4\mathbb{N}$ are

$$P_n^{(F)} = \frac{1}{4} \begin{pmatrix} 2\;2 & & & & & & & \\ & 2\;2 & & & & & \mathbf{0} & \\ & & \ddots & & & & & \\ & & & 2\;2 & & & & \\ 1\;1\;1\;1 & & & & & & & \\ & 1\;1\;1\;1 & & & & & & \\ & & \ddots & & & & & \\ & & & & \ddots & & & \\ & & & & & & 1\;1\;1\;1 & \\ 1\;1\;1\;1 & & & & & & & \\ & 1\;1\;1\;1 & & & & & & \\ & & \ddots & & & & & \\ & & & & \ddots & & & \\ & & & & & & 1\;1\;1\;1 & \\ & & & & 2\;2 & & & \\ & \mathbf{0} & & & & \ddots & & \\ & & & & & & & 2\;2 \end{pmatrix}, \tag{5.4.1}$$

$$U_n^{(F)} = \frac{1}{2} \begin{pmatrix} \sqrt{2}\;-\sqrt{2} & & & & & & & \\ & \sqrt{2}\;-\sqrt{2} & & & & & \mathbf{0} & \\ & & \ddots & & & & & \\ & & & \sqrt{2}\;-\sqrt{2} & & & & \\ 1\;\;1\;\;1\;\;1 & & & & & & & \\ & 1\;1\;\;1\;\;1 & & & & & & \\ & & \ddots & & & & & \\ & & & & \ddots & & & \\ & & & & & & 1\;1\;\;1\;\;1 & \\ 1\;\;1\;\;-1\;\;-1 & & & & & & & \\ & 1\;1\;-1\;-1 & & & & & & \\ & & \ddots & & & & & \\ & & & & \ddots & & & \\ & & & & & & 1\;1\;-1\;\;-1 & \\ & & & & \sqrt{2}\;-\sqrt{2} & & & \\ & \mathbf{0} & & & & \ddots & & \\ & & & & & & & \sqrt{2}\;-\sqrt{2} \end{pmatrix}. \tag{5.4.2}$$

For the tripling map, the corresponding Markov matrix $P_n^{(T)}$ and a simple unitary quantization $U_n^{(T)}$ (using a 3×3 DFT matrix) for $n \in 3\mathbb{N}$ are

(5.4.3)

$$P_n^{(T)} = \frac{1}{3}\begin{pmatrix} 1\,1\,1 & & & \\ & 1\,1\,1 & & \\ & & \ddots & \\ & & & 1\,1\,1 \\ 1\,1\,1 & & & \\ & 1\,1\,1 & & \\ & & \ddots & \\ & & & 1\,1\,1 \\ 1\,1\,1 & & & \\ & 1\,1\,1 & & \\ & & \ddots & \\ & & & 1\,1\,1 \end{pmatrix}, \quad U_n^{(T)} = \frac{1}{\sqrt{3}}\begin{pmatrix} 1\ e^{\frac{2\pi i}{3}}\ e^{-\frac{2\pi i}{3}} & & & \\ & 1\ e^{\frac{2\pi i}{3}}\ e^{-\frac{2\pi i}{3}} & & \\ & & \ddots & \\ & & & 1\ e^{\frac{2\pi i}{3}}\ e^{-\frac{2\pi i}{3}} \\ 1\ e^{-\frac{2\pi i}{3}}\ e^{\frac{2\pi i}{3}} & & & \\ & 1\ e^{-\frac{2\pi i}{3}}\ e^{\frac{2\pi i}{3}} & & \\ & & \ddots & \\ & & & 1\ e^{-\frac{2\pi i}{3}}\ e^{\frac{2\pi i}{3}} \\ 1\ \ 1\ \ 1 & & & \\ & 1\ \ 1\ \ 1 & & \\ & & \ddots & \\ & & & 1\ \ 1\ \ 1 \end{pmatrix}.$$

As plotted in Figures 5.2 and 5.3, the unitary matrices $U_n^{(F)}$ and $U_n^{(T)}$ appear to have CUE-like level spacings and eigenvectors with approximately Gaussian coordinate statistics.

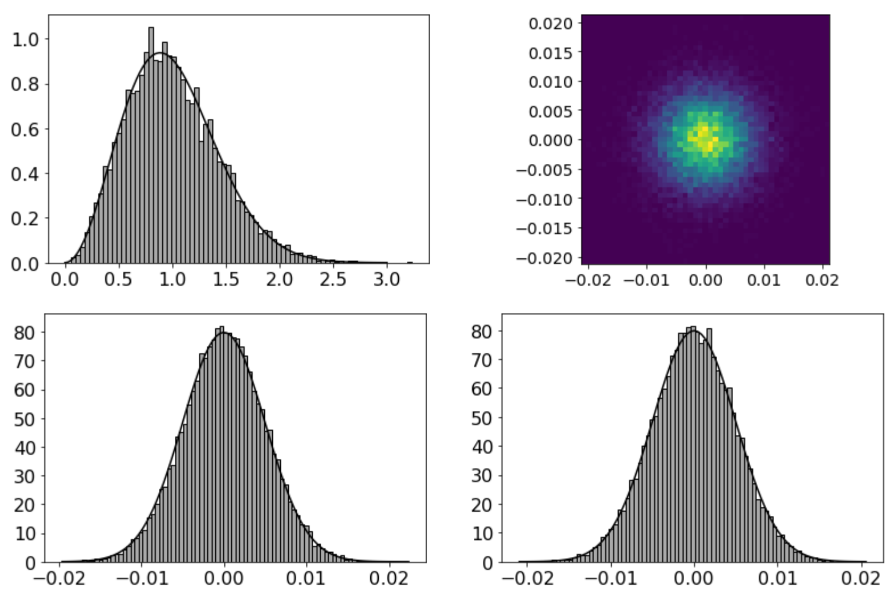

FIGURE 5.2. Four legs map numerics for $U_n^{(F)}$ for $n = 20\,000$. Top left: angle level spacing distribution against Wigner surmise for GUE. Top right: 2D coordinate histogram in $\mathbb{C}$ for a randomly chosen eigenvector ψ (this one with eigenvalue $0.972610 - 0.232443i$). Bottom left: Histogram of the values $(\mathfrak{Re}\,\psi_x)_{x=1}^{20\,000}$ against the pdf of $N(0, \frac{1}{40\,000})$. Bottom right: Histogram of the values $(\mathfrak{Im}\,\psi_x)_{x=1}^{20\,000}$ against the pdf of $N(0, \frac{1}{40\,000})$.

We note that the main focus in [PZK01] was on spectral statistics averaged over phases $\Phi \in [0, 2\pi)^n$ that appear through multiplication by the diagonal matrix $e^{i\Phi} := \mathrm{diag}(e^{i\Phi_1}, \dots, e^{i\Phi_n})$

to form the ensemble of unitary matrices $e^{i\Phi}U_n$. With averaging over these phases, the spectral statistics agree even more closely with those of CUE matrices. It was also observed there that the spectral statistics for just a single unitary matrix corresponding to the four legs map still continued to exhibit CUE-like spectral behavior, which we also see for the level spacings of the matrices here.

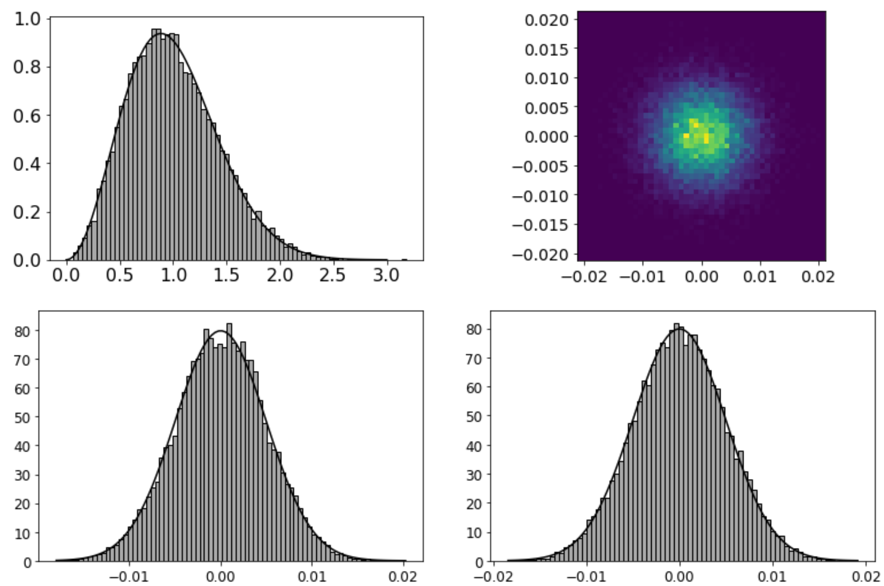

FIGURE 5.3. Tripling map numerics for $U_n^{(T)}$ for $n = 20\,001$, as in Figure 5.2, with randomly chosen eigenvector (this one with eigenvalue $0.546244 + 0.837691i$).

Part 2

Localization-delocalization transition for nonhomogeneous random matrices

CHAPTER 6

Introduction and main result

6.1. Introduction

Understanding the eigenvectors of large random matrices, and particularly whether they are delocalized or localized, is of interest in many areas including mathematical physics, computer science, and combinatorics. A delocalized vector has roughly equal mass spread throughout its coordinates, while a localized vector has much of its mass concentrated on relatively few coordinates. A prime example of generally delocalized vectors is a uniform random vector from the unit sphere. This describes for example eigenvectors of rotationally invariant ensembles like the classical Gaussian orthogonal ensemble (GOE). Properties of the uniform distribution on the sphere are then a benchmark for measuring how delocalized other vectors are in comparison. Much work has been done to show delocalization of eigenvectors of general *Wigner-type matrices*. For a summary and many references, see the book [EY17] or survey [OVW16].

In contrast to delocalized vectors, the most localized vectors are simply the coordinate directions, with all mass concentrated on a single coordinate. These arise as eigenvectors of diagonal matrices, such as a diagonal matrix with iid diagonal Gaussian entries. To interpolate between the two extremes of a diagonal matrix and a GOE matrix, one can consider models of varying degrees of sparseness. One such model of interest in mathematical physics is random band matrices, which (for 1D) are zero outside of a band around the diagonal, and undergo a transition depending on the band width; see [Bou18] for a survey.

In this part of the thesis we will consider Gaussian matrices arising from d-regular graphs, which includes the above mentioned Gaussian models (GOE, diagonal, and band[1]), as well as many other matrices with a fixed structure. As we will see in Section 6.2, the matrix norm

[1]If one considers periodic band matrices, which have nonzero entries in the corners as well, then these matrices have d nonzero entries in each row and column.

of such matrices is known to undergo a phase transition at $d \sim \log N$ [BvH16]. We will show this transition is detected not by the localization or delocalization of the top eigenvectors, but by that of *approximate* top eigenvectors, by which we will mean a unit vector v with $\|Xv\|_2$ close to $\|X\|$.

Localization and delocalization of vectors can be described by various non-equivalent notions, such as the ℓ^∞ norm or other ℓ^p norms [ESY09b, ESY09a], joint distribution of coordinates [TV11], or no-gaps delocalization [RV16]; see also [OVW16] for a survey including results on different notions of delocalization. Here we will use the complementary notion of the (L, η)-localization used in [ESY09b, §7]. A vector delocalized in this sense will be one that has no large "peaks" in a small set of coordinates.

Definition 1 (delocalization)**.** Call a vector $v \in \mathbb{S}^{N-1}$ (L, κ)-*delocalized* if for every set $A \subset [N]$ of size $|A| = L$, we have $\sum_{j\in A} v_j^2 \leq \kappa^2$. The set of (L, κ)-delocalized vectors will be denoted by

$$D_{L,\kappa} := \left\{ v \in \mathbb{S}^{N-1} : \forall A \subset [N], |A| = L, \sum_{j\in A} |v_j|^2 \leq \kappa^2 \right\}.$$

Thus a vector in $D_{L,\kappa}$ is one that has no peaks of mass $> \kappa^2$ supported on size $\leq L$. The condition becomes stricter for smaller κ and larger L. Since for very small L, the condition to be in $D_{L,\kappa}$ becomes very weak, we will typically take L to be proportional to N, say $L = \nu N$ (or more precisely $L = \lfloor \nu N \rfloor$) for some $0 < \nu < 1$. In this case we will always assume $\nu < \kappa^2$ otherwise $D_{\nu N,\kappa} = \emptyset$. Thus colloquially we will refer to a vector, or more precisely a sequence of vectors, as *delocalized* if it is (L, κ)-delocalized for some L proportional to N as $N \to \infty$, and otherwise we will refer to it as *localized*. In this sense, a delocalized vector is one that is not concentrated in a vanishing fraction of the coordinates.

Since a uniform random vector on the sphere is the benchmark example of a delocalized vector, the example in Section 7.2 will show that such a vector is indeed $(\nu N, \kappa)$-delocalized with high probability for a ν chosen depending on κ.

6.2. Structured matrices from d-regular graphs

We will look at eigenvectors and approximate eigenvectors for the largest eigenvalues of structured random matrices arising from d-regular graphs. These matrices can be sparse and can behave very differently from traditional Wigner matrices. Given an undirected graph $G = (V, E)$ with vertices $V = [N]$, let $x \sim y$ mean that nodes x and y are connected by an edge. Here self-loops are allowed (and counted as a single edge) but multi-edges are not allowed. Define the random $N \times N$ symmetric matrix X by $X_{xy} = \delta_{x\sim y} g_{xy}$, where g_{xy} are iid standard normal random variables modulo the symmetry requirement $g_{xy} = g_{yx}$. This assigns an independent standard normal variable to each edge (xy). For example, the complete graph on N vertices corresponds to an $N \times N$ GOE matrix, while the graph of N isolated points with self-loops corresponds to a diagonal matrix with independent $\mathcal{N}(0, 1)$ variables on the diagonal. Note these are not random graphs, but fixed graphs with Gaussian entries on the adjacency matrix.

Concerning the largest eigenvalue magnitude, for such matrices that come from a d-regular graph in the method described above, it was shown in [BvH16] that $\mathbb{E}\|X\| \asymp \sqrt{d} + \sqrt{\log N}$.

Theorem 6.1 ([BvH16])**.** *Let $X = (X_{xy})_{xy}$ be a symmetric $N \times N$ matrix corresponding to a d-regular graph, with $d \to \infty$ and $X_{xy} = \delta_{x\sim y} g_{xy}$, where g_{xy} are iid $\mathcal{N}(0, 1)$ modulo symmetry requirements. Then for any $\varepsilon > 0$,*

$$\max\left\{(2 - o(1))\sqrt{d}, C\sqrt{\log N}\right\} \le \mathbb{E}\|X\| \le (2 + \varepsilon)\sqrt{d} + K_\varepsilon\sqrt{\log N}. \tag{6.2.1}$$

In particular, if $d/\log N \to \infty$, then

$$(2 - o(1))\sqrt{d} \le \mathbb{E}\|X\| \le (2 + o(1))\sqrt{d}. \tag{6.2.2}$$

The upper bound is Theorem 1.1 in [BvH16]. The lower bound up to a constant is Lemma 3.14 in [BvH16]. The specific constant $(2 - o(1))\sqrt{d}$ in the lower bound can be obtained by the moment method or semicircle law, see for example [BvH16, §4.1] or [vH17, §4.2].

Based on this theorem, one can expect a phase transition to occur at $d \sim \log N$, where the dominating term in the norm bound changes. In the extreme case of a diagonal matrix ($d = 1$), the matrix norm is simply the largest (in magnitude) diagonal entry, which is of order $\sqrt{\log N}$, and the eigenvectors are just the coordinate directions. In the other extreme where $d = N$, X is GOE and has matrix norm of order $2\sqrt{N} = 2\sqrt{d}$ with delocalized eigenvectors. Equation (6.2.1) suggests that for $d \ll \log N$, we might obtain the largest eigenvalue by taking an eigenvector that is localized on the large outliers of order $\sqrt{\log N}$. On the other hand, for $d \gg \log N$, we might obtain the largest eigenvalue by instead taking a delocalized vector as in the GOE case.

However, this intuition cannot be entirely correct, as can be seen in the example of a *block Wigner matrix* $X = \bigotimes_{i=1}^{N/d} Y_i^{(d)}$, where each $(Y_i^{(d)})_i$ is a $d \times d$ symmetric Wigner matrix with iid $\mathcal{N}(0, 1)$ entries, and independent of other $Y_i^{(d)}$s. Then X is block diagonal with blocks size $d \times d$, and so its eigenvectors are localized to each block of size d. Then if $d \gg \log N$ but $d \ll \nu N$, the eigenvectors will not be $(\nu N, \kappa)$-delocalized. However, if $d \gg \log N$ we can create a delocalized *approximate* largest eigenvector by taking the top eigenvector of each block and averaging them. In this context, recall we consider an approximate largest eigenvector to be a unit vector v such that $\|Xv\|_2$ is close to achieving the maximum possible value $\|X\|$. Motivated by this block Wigner matrix example, we will look for whether or not we can find near-maximizers of the norm with good delocalization properties, in order to identify the phase transition at $d \sim \log N$.

6.2.1. Main result.

Theorem 6.2 (localization-delocalization transition)**.** *Let X_N be a sequence of symmetric $N \times N$ matrices, each from a $d = d(N)$-regular graph, with $(X_N)_{xy} = \delta_{x\sim y} g_{xy}$, where g_{xy} are iid $\mathcal{N}(0, 1)$ modulo symmetry requirements.*

(i) Localization for $d = o(\log N)$: Fix $0 < \varepsilon < 1$ and $0 < \kappa < 1 - \varepsilon$. If

$$\mathbb{E} \sup_{v \in D_{\nu N,\kappa}} \|X_N v\| \geq (1 - \varepsilon)\mathbb{E}\|X_N\|, \tag{6.2.3}$$

then $\nu \lesssim_{\varepsilon,\kappa} \frac{d}{\log N}$*. In particular, if* $d = o(\log N)$*, then* $\nu \to 0$ *as* $N \to \infty$*.*

(ii) Delocalization for $d \gg \log N$*: Let* $0 < \kappa < 1$ *and* $0 < \nu < \frac{c\kappa^2}{\log \frac{e}{\kappa}}$*, where* $c > 0$ *is an absolute constant. If* $d \gg \log N$*, then with probability* $1 - o(1)$*,*

$$\sup_{v \in D_{\nu N,\kappa}} \|X_N v\|_2 \geq (1 - o(1))\mathbb{E}\|X_N\|. \tag{6.2.4}$$

Also

$$\mathbb{E} \sup_{v \in D_{\nu N,\kappa}} \|X_N v\|_2 \geq (1 - o(1))\mathbb{E}\|X_N\|. \tag{6.2.5}$$

Remark 6.2.1. (a) Morally, part (i) says that if $d = o(\log N)$, then vectors v such that $\|X_N v\|_2$ is even within just a constant fraction of the top eigenvalue must be localized. Thus not only are the top eigenvectors localized in this regime, but all approximate top eigenvectors must be localized as well. In contrast, part (ii) says that if $d \gg \log N$, then we can find a good approximation to an eigenvector for $\|X_N\| \approx 2\sqrt{d}$ by searching in the set $D_{\nu N,\kappa}$ of $(\nu N, \kappa)$-delocalized vectors, even if the actual eigenvectors are localized.

(b) The condition $0 < \nu < \frac{c\kappa^2}{\log \frac{e}{\kappa}}$ in (ii) is equivalent, up to constant factors, to that which is required for high probability delocalization of the uniform distribution of the sphere; see Remarks 7.2.1 and 9.4.1. Thus in the sense of delocalization in Definition 1, the delocalized approximate largest eigenvector guaranteed by (ii) is roughly just as delocalized as a typical vector chosen uniformly from the sphere.

(c) For equation (6.2.4), we will prove a lower bound of $(2 - o(1))\sqrt{d}$, which by Theorem 6.1 is $(1 - o(1))\mathbb{E}\|X_N\|$. We first note that obtaining a lower bound of just $(2 - \varepsilon)\sqrt{d}$ for $\varepsilon > 0$ and some $\nu = \nu(\varepsilon) > 0$ is much simpler, because one can take the top αN eigenvectors of X_N, and just use that given any αN orthonormal vectors, their span must contain a $(\nu N, \kappa)$-delocalized vector for sufficiently small ν. This works even if the αN vectors are all coordinate vectors, but no longer works if we consider only $o(N)$ vectors. In order to get the $(2 - o(1))\sqrt{d}$ bound, we will consider a number $o(N)$ of the top eigenvectors, and use projection matrix estimates to show that their span contains a delocalized vector. We mention that by [BGP14, Theorem 2.9], the top eigenvectors of X_N cannot be *too* localized in this regime, and so this makes the possibility of an $o(1)$ term seem intuitively plausible, as long as the top eigenvectors do not cluster together too strongly in one area.

(d) Besides block Wigner matrices, another example where $d \gg \log N$ but all eigenvectors can be localized are the (1D) band matrices [FM91, Sch09]. In this case, the proof of Theorem 6.2(ii) implies the top eigenvectors cannot all be concentrated in too few of the same coordinates, since their span must contain a delocalized vector.

The proof for (i) is to split up a vector v in $D_{\nu N,\kappa}$ into its large entries, which are controlled by delocalization, and its smaller entries, which can be controlled by subgaussian estimates.

For (ii), the proof idea is to consider the eigenvectors of $X_N/\sqrt{d}$ with eigenvalues at least $2 - o(1)$ for some $o(1)$ term, and show that there is a delocalized vector in their span. Then this delocalized vector automatically satisfies $\|X_N v\|_2 \geq (2 - o(1))\sqrt{d}$. Finding such a delocalized vector is done by approximating the diagonal entries of the associated projection matrix using resolvents and a (weak) local semicircle law. The estimates on the diagonal entries will provide Gaussian moment bounds, and these moments restrict how much mass can accumulate in just νN coordinates.

6.2.2. Outline. In Chapter 7, we review several facts about subgaussian random variables, discuss the semicircle law, and give the example of delocalization for uniform random vectors on the sphere. In Chapter 8, we prove part (i) of Theorem 6.2, the localization for $d = o(\log N)$. In Chapter 9, we prove part (ii), starting with proving a (weak) local semicircle law for the Green's function in Section 9.1. Then in Section 9.2 we approximate the spectral projection matrix onto $[2 - \varepsilon_N, b]$ in terms of the Green's function. In Section 9.3, we take the spectral projection matrix approximation and replace the Green's function with the Stieltjes transform of the semicircle law, using the local semicircle law and Gaussian concentration. This gives the asymptotics for the matrix entries of the spectral projection matrix, which are then used in Section 9.4 to show we can pick $\nu > 0$ that ensures a $(\nu N, \kappa)$-delocalized vector v exists with $\|X_N v\|_2 \geq (2 - o(1))\sqrt{d}$.

CHAPTER 7

Background

7.1. Gaussian concentration and suprema

We gather several useful properties concerning subgaussian random variables. For further background, see for example the books and notes [BLM13, vH16, Ver18, Wai19]. Recall a real random variable X is σ^2-subgaussian if $\mathbb{E}[e^{t(X-\mathbb{E}X)}] \le e^{\sigma^2 t^2/2}$ for all $t \in \mathbb{R}$. Tail bounds follow from the Chernoff bound, for any $\lambda, t > 0$,

$$\mathbb{P}[X - \mathbb{E}X \ge t] = \mathbb{P}[e^{\lambda(X-\mathbb{E}X)} \ge e^{\lambda t}] \le \frac{\mathbb{E}[e^{\lambda(X-\mathbb{E}X)}]}{e^{\lambda t}} \le e^{\frac{\sigma^2\lambda^2}{2} - \lambda t},$$

by taking $\lambda = t/\sigma^2$ to minimize the bound, yielding for $t \ge 0$,

$$\mathbb{P}[X - \mathbb{E}X \ge t] \le e^{-\frac{t^2}{2\sigma^2}}. \tag{7.1.1}$$

By symmetry, $-X$ is subgaussian if and only if X is subgaussian, so one also obtains for $t \ge 0$,

$$\mathbb{P}[\mathbb{E}X - X \ge t] \le e^{-\frac{t^2}{2\sigma^2}}. \tag{7.1.2}$$

The tail bounds give moment bounds, for any $p > 0$,

$$\begin{aligned} \mathbb{E}|X - \mathbb{E}X|^p &= \int_0^\infty p t^{p-1} \mathbb{P}[|X - \mathbb{E}X| > t]\, dt \\ &\le 2\int_0^\infty p t^{p-1} e^{\frac{-t^2}{2\sigma^2}}\, dt = p 2^{p/2} \sigma^p \Gamma(p/2). \end{aligned} \tag{7.1.3}$$

One can use this and the series expansion of e^{tx} to show that if X satisfies (7.1.1) and (7.1.2), then X is subgaussian, though possibly with a possibly larger variance proxy $\widetilde{\sigma}^2$. Thus up to constants, subgaussian variables can be characterized using the moment generating function bound, tail bounds, or moment bounds. They can also be characterized by the ψ_2-condition, that there exists $s > 0$ so that $\mathbb{E}e^{sX^2} \leq 2$, with the subgaussian norm $\|X\|_{\psi_2} := \inf\{t > 0 : \mathbb{E}\exp(X^2/t^2) \leq 2\}$, for details see [Ver18, §2].

One can get better constants for $p = 1, 2$ than in (7.1.3) by using the series expansion of e^{tx}. Letting $Y = X - \mathbb{E}X$, then by Fubini or dominated convergence, since $\mathbb{E}[e^{|t||Y|}] \leq \mathbb{E}[e^{|t|Y}] + \mathbb{E}[e^{-|t|Y}]$,

$$\mathbb{E}[e^{tY}] = \sum_{n=0}^{\infty} \frac{t^n \mathbb{E}Y^n}{n!} \leq e^{\sigma^2 t^2/2} = 1 + \frac{\sigma^2 t^2}{2} + \mathcal{O}(t^3).$$

Since $\mathbb{E}Y = 0$, this yields

$$\frac{\mathbb{E}Y^2}{2} + \mathcal{O}(t) \leq \frac{\sigma^2}{2} + \mathcal{O}(t),$$

so taking $t \to 0$ yields $\operatorname{Var} X = \mathbb{E}Y^2 \leq \sigma^2$, and by Cauchy–Schwarz, $\mathbb{E}|X - \mathbb{E}X| \leq \sigma$.

An important property of Gaussian random variables is the following classical concentration of measure statement.

Theorem 7.1 (Gaussian concentration)**.** *Let $X_1, \ldots, X_n$ be iid $\mathcal{N}(0,1)$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz with respect to the Euclidean norm. Then $f(X) - \mathbb{E}[f(X)]$ is subgaussian with parameter $\sigma^2 \leq L^2$, and so*

$$\mathbb{P}\big[\,|f(X) - \mathbb{E}[f(X)]| \geq t\big] \leq 2e^{-\frac{t^2}{2L^2}}, \qquad \text{for all } t \geq 0. \tag{7.1.4}$$

There are several different proofs of Gaussian concentration. A standard one is via the entropy method and log-Sobolev inequalities, see [BLM13, §5.4] or [vH16, §3]. A very direct proof is given in [AT07, §2.1] on the way to proving the Borell-TIS inequality, an important result that suprema of Gaussian processes concentrate. Here though we only deal with finite sets of Gaussians and will not really need concentration of the maximum anyway.

Note, one cannot replace the Gaussian $X_1, \ldots, X_n$ with general independent subgaussian random variables, in particular [LT91, §1.1/p.25] shows (7.1.4) can fail for Rademacher/Bernoulli random variables. However, concentration with bounded random variables holds with additional convexity assumptions on f, see [BLM13, §6.6] or [vH16, §4.3].

Gaussian concentration implies concentration of the norm $\|X\|_2$ for $X \sim \mathcal{N}(0, I_n)$, as it is 1-Lipschitz, as well as concentration on the sphere.

Corollary 7.2 (norm concentration and concentration on the sphere)**.**

- *Let $X \sim \mathcal{N}(0, I_n)$. Then for any $t \geq 0$,*

$$\mathbb{P}\left[\left|\|X\|_2 - \sqrt{n}\right| \geq t\right] \leq 2e^{-ct^2}, \tag{7.1.5}$$

  where $c > 0$ is an absolute constant. Note that $\mathbb{E}\|X\|_2 = \sqrt{2}\Gamma(\frac{n+1}{2})/\Gamma(\frac{n}{2}) \approx \sqrt{n}$.

- *Let $\theta \sim \mathrm{Unif}(\mathbb{S}^{n-1})$, so $\theta \sim X/\|X\|$. Then for any $f : \mathbb{R}^n \to \mathbb{R}$ L-Lipschitz and $t \geq 0$,*

$$\mathbb{P}\left[|f(\theta) - \mathbb{E}f(\theta)| \geq t\right] \leq Ce^{-\frac{ct^2 n}{L^2}}, \tag{7.1.6}$$

  where $c, C > 0$ are absolute constants.

Example 7.1 (quantities that concentrate)**.** Useful functions of a random Gaussian matrix that concentrate include the norm, the largest eigenvalue, and the Green's function. Let H be an $N \times N$ real symmetric matrix with some pattern of independent $\mathcal{N}(0,1)$ entries, and view a function $f(H)$ as a function of the matrix entries, $\mathbb{R}^{\triangledown} \to \mathbb{R}$, where $\mathbb{R}^{\triangledown}$ corresponds to the nonzero entries in the upper triangular part of H. (For the matrices coming from d-regular graphs, these entries are the (i,j) so that $i \sim j$ and $j \geq i$.)

- For the operator norm $f(H) = \|H\| = \sup_{\|v\|=1}\langle v, Hv\rangle$, note that for H and H' with the same sparsity pattern corresponding to entries from $\mathbb{R}^{\triangledown}$,

$$\begin{aligned}\left|\sup_{v\in V}\langle v, Hv\rangle - \sup_{w\in V}\langle w, H'w\rangle\right| &\leq \sup_{v\in V}|\langle v, Hv\rangle - \langle v, H'v\rangle| \\ &\leq \sup_{v\in V}\|H - H'\|_F\left(\sum_{i,j=1}^N v_i^2 v_j^2\right)^{1/2} \\ &= \|H - H'\|_F \leq \sqrt{2}\|H - H'\|_{\mathbb{R}^{\triangledown}},\end{aligned}$$

where $\|\cdot\|_F$ is the Frobenius or Hilbert–Schmidt norm, which is the Euclidean norm on the matrix entries. Thus the norm is $\sqrt{2}$-Lipschitz, and concentrates according to Theorem 7.1.

- More generally, for an L-Lipschitz function $F : \mathbb{R}^N \to \mathbb{R}$, the corresponding function $f(H) := F(\lambda_1(H), \ldots, \lambda_N(H))$ is L-Lipschitz with respect to the Frobenius/Hilbert–Schmidt norm. This follows from the Hoffman–Wielandt inequality $\sum_{i=1}^N (\lambda_i(H) - \lambda_i(H'))^2 \le \operatorname{tr}[(H-H')^2] = \|H-H'\|_F^2$ (for ordered eigenvalues of real symmetric or hermitian matrices), and covers quantities like the norm, largest eigenvalue, and trace of the resolvent. For details see [AGZ09, Ch.2].
- The Green's function $G_H(z;x,y)$ is the kernel of the resolvent $R_H(z) := (H-z)^{-1}$, defined as $G_H(z;x,y) := \langle x|(H-z)^{-1}|y\rangle$. It is complex-valued, but Theorem 7.1 applies to its real and imaginary parts. For H self-adjoint, since $\|(H-z)^{-1}\| = \operatorname{dist}(z, \sigma(H)) \le |\mathfrak{Im}\, z|^{-1}$, then using the second resolvent identity shows,

$$\begin{aligned} |G_H(z;x,y) - G_{H'}(z;x,y)| &= \left|\langle x|(H-z)^{-1}(H-H')(H'-z)^{-1}|y\rangle\right| \\ &\le \|H-H'\|_F \left(\sum_{\ell,m} |\langle x|R_H(z)|\ell\rangle|^2 \cdot |\langle m|R_{H'}(z)|y\rangle|^2\right)^{1/2} \\ &= \|H-H'\|_F \left(|\langle x|R_H(z)R_H(\overline{z})|x\rangle| \cdot |\langle y|R_{H'}(z)R_{H'}(\overline{z})|y\rangle|\right)^{1/2} \\ &\le \|H-H'\|_F \frac{1}{|\mathfrak{Im}\, z|^2} \le \|H-H'\|_{\mathbb{R}\triangledown} \frac{\sqrt{2}}{|\mathfrak{Im}\, z|^2}, \end{aligned}$$

so the Green's function has Lipschitz constant $L \le \frac{\sqrt{2}}{|\mathfrak{Im}\, z|^2}$, and

$$\begin{aligned} \mathbb{P}[|G_H(z;x,y) - \mathbb{E}G_H(z;x,y)| \ge t] &\le \mathbb{P}[\mathfrak{Re}\, G_H(z;x,y) - \mathbb{E}\,\mathfrak{Re}\, G_H(z;x,y)| \ge t/\sqrt{2}] \\ &\quad + \mathbb{P}[|\,\mathfrak{Im}\, G_H(z;x,y) - \mathbb{E}\,\mathfrak{Im}\, G_H(z;x,y)| \ge t/\sqrt{2}] \\ &\le 4\exp\left(-\frac{1}{8}t^2|\,\mathfrak{Im}\, z|^4\right). \end{aligned}$$

- For considering delocalization for the uniform distribution on the sphere, a useful (non-matrix) quantity that concentrates is $\sup_{\substack{A\subset[N] \\ \#A=\nu N}} \left(\sum_{x\in A} v_x^2\right)^{1/2}$. For a vector $w \in \mathbb{R}^N$, let the jth largest coordinate in absolute value be $w_{(j)} := j\text{-}\max(|w_1|, \ldots, |w_N|)$, and define the function $f(w) = \left(\sum_{j=1}^{\nu N} w_{(j)}^2\right)^{1/2}$. Then f is 1-Lipschitz, using the reverse triangle inequality and that if $\widetilde{w}$ and $\widetilde{y}$ are the sorted vectors $(w_{(1)}, \dots, w_{(N)})$ and $(y_{(1)}, \dots, y_{(N)})$ respectively, then $\|\widetilde{w} - \widetilde{y}\|_2 \leq \|w - y\|_2$. (If two entries are out of order, then swapping them decreases the norm.) Concentration on the sphere then shows that for $w \sim \mathrm{Unif}(\mathbb{S}^{N-1})$,

$$\mathbb{P}\left[\left|\left(\sum_{j=1}^{\nu N} w_{(j)}^2\right)^{1/2} - \mathbb{E}\left(\sum_{j=1}^{\nu N} w_{(j)}^2\right)^{1/2}\right| \geq t\right] \leq Ce^{-ct^2 N}. \tag{7.1.7}$$

Next, an upper bound on the maximum of a finite number of subgaussian random variables will be useful in Section 8.

Theorem 7.3 (maximum of a finite number of subgaussians)**.** *Suppose $X_1, \dots, X_n$ are all mean zero and σ^2-subgaussian. Then*

$$\mathbb{E}\left[\max_{i=1,\dots,n} X_i\right] \leq \sqrt{2\sigma^2 \log n}.$$

Proof. For the first part, by Jensen,

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in T} X_t\right] &\leq \frac{1}{\lambda}\log \mathbb{E}\left[e^{\lambda \sup_{t\in T} X_t}\right] \leq \frac{1}{\lambda}\log\sum_{t\in T}\mathbb{E}\left[e^{\lambda X_t}\right] \\ &\leq \frac{1}{\lambda}\left(\log|T| + \frac{\sigma^2\lambda^2}{2}\right) \leq \sqrt{2\sigma^2\log|T|},\end{aligned}$$

by optimizing over $\lambda > 0$ to take $\lambda = \sigma^{-1}\sqrt{2\log|T|}$. Note this does not require independence.

□

Finally, for looking at the uniform distribution on the sphere in Section 7.2, it will be useful to have bounds on the expected value of the kth order statistics as well.

Theorem 7.4 (Gaussian order statistics)**.** *Let $Z \sim \mathcal{N}(0, I_n)$, and let $Z_{(k)}$ denote the absolute values of the entries of Z sorted in decreasing order, $Z_{(1)} \geq Z_{(2)} \geq \cdots \geq Z_{(N)} \geq 0$. Then for any $1 \leq k \leq \frac{n}{2}$,*

$$c\sqrt{\log\frac{n}{k}} \leq \mathbb{E}Z_{(k)} \leq C\sqrt{\log\frac{n}{k}}, \tag{7.1.8}$$

where $c, C > 0$ are absolute constants.

See also [GLSW06, Example 10], which includes bounds for $\frac{n}{2} \le k \le n$. Here, for $1 \le k \le \frac{n}{2}$, if one does not care about the constants, the proof can be done using union bound and various estimates on the normal cdf.

7.2. Unit sphere delocalization

For a GOE matrix, the eigenvectors are uniformly distributed on the unit sphere by rotational symmetry. As this is our benchmark example of delocalized vectors, we demonstrate that with high probability such vectors satisfy the $(\nu N, \kappa)$-delocalization definition. There are several ways to show this for uniform random vectors on the sphere, e.g. by Gaussian concentration, estimating integrals, or using order statistics. For this example we start with Gaussian concentration. Much more precise information on the distributions can be found in [OVW16, §2].

Lemma 7.5 (unit sphere delocalization)**.** *Let v be distributed uniformly on the unit sphere $\mathbb{S}^{N-1} \subset \mathbb{R}^N$. Fix $0 < \kappa < 1$. Then there are $\nu = \nu(\kappa) > 0$ and $\beta = \beta(\nu) = \beta(\kappa) > 0$ such that*

$$\mathbb{P}_{\mathbb{S}^{N-1}}[x \in D_{\nu N, \kappa}] \ge 1 - 2e^{-\beta(\nu) N}. \tag{7.2.1}$$

Proof. Let $g \sim \mathcal{N}(0, I_N)$, so that $\frac{g}{\|g\|_2} \sim \mathrm{Unif}(\mathbb{S}^{N-1})$. Estimating by union bound and using coordinate symmetry of g,

$$\begin{aligned}
\mathbb{P}\left[\frac{g}{\|g\|_2} \notin D_{\nu N,\kappa}\right] &= \mathbb{P}\left[\exists A \subset [N],\ \#A = \nu N : \sum_{j \in A} \frac{g_j^2}{\|g\|_2^2} \ge \kappa^2\right] \\
&\le \binom{N}{\nu N} \mathbb{P}\left[\sum_{j=1}^{\nu N} g_j^2 \ge \kappa^2 \|g\|_2^2\right] \\
&\le \left(\frac{e}{\nu}\right)^{\nu N} \left(\mathbb{P}\left[\sum_{j=1}^{\nu N} g_j^2 \ge \kappa^2 \frac{N}{2}\right] + \mathbb{P}\left[\|g\|_2^2 \le \frac{N}{2}\right]\right)
\end{aligned}$$

Since the norm concentrates, $\mathbb{P}\left[\|g\|_2 \le \sqrt{N/2}\right] \le e^{-cN}$, for some $c > 0$, and similarly (for $2\nu < \kappa^2$),

$$\mathbb{P}\left[\sum_{j=1}^{\nu N} g_j^2 \ge \frac{\kappa^2}{2\nu} \cdot \nu N\right] \le \exp\left(-c\Big(\frac{\kappa}{\sqrt{2\nu}} - 1\Big)^2 \nu N\right).$$

For $\nu < \frac{\kappa^2}{8}$, $\left(\frac{\kappa}{\sqrt{2\nu}} - 1\right)\sqrt{\nu} \ge \frac{\kappa}{2\sqrt{2}}$. Then by taking ν sufficiently small, we can ensure the term $(e/v)^{\nu N}$ term from the union bound is small enough so that $\left(\frac{e}{\nu}\right)^{\nu N}(2e^{-\frac{c}{8}\kappa^2 N}) = 2e^{-\left[\frac{c}{8}\kappa^2 - \nu \log\frac{e}{\nu}\right]N}$ with $\beta(\nu) := \frac{c}{8}\kappa^2 - \nu\log\frac{e}{\nu} > 0$. □

Remark 7.2.1. The condition $\nu\log\frac{e}{\nu} < c\kappa^2$ required above is around the best we can hope to get, since from Gaussian order statistics or [OVW16, Theorem 2.7], we expect that $\sup_{\substack{A\subset[N]\\ \#A=\nu N}} \sum_{x\in A} v_j^2 \approx \Theta(\nu\log\frac{1}{\nu})$ for ν small.

One can see this using the order statistics bounds in Theorem 7.4 as follows. As described in (7.1.7), the function $f : \mathbb{S}^{N-1} \to \mathbb{R}$ defined by $f(w) = \left(\sum_{j=1}^{\nu N} w_{(j)}^2\right)^{1/2}$ concentrates around its mean. Estimates on the mean will follow from (7.1.8), using norm concentration and that each $h_j(z) := j\text{-max}(|z_1|, \dots, |z_n|)$, which gives the jth largest entry, is also 1-Lipschitz. The general estimate is that for $Z \sim \mathcal{N}(0, I_n)$,

$$\frac{1}{N}\sum_{j=1}^{\nu N}(\mathbb{E}Z_{(j)})^2 \asymp \frac{C'}{N}\log\prod_{j=1}^{\nu N}\frac{N}{j} = \frac{C'}{N}\log\left(\frac{N^{\nu N}}{(\nu N)!}\right),$$

and by Stirling's formula, assuming $\nu N \ge 1$,

$$\nu\log\frac{e}{\nu} - o(1) \le \frac{1}{N}\log\left(\frac{N^{\nu N}}{(\nu N)!}\right) \le \nu\log\frac{e}{\nu},$$

with the $o(1)$ term as $N \to \infty$, so that

$$\frac{1}{N}\sum_{j=1}^{\nu N}(\mathbb{E}Z_{(j)})^2 \asymp C'\nu\log\frac{e}{\nu} + o(1).$$

Using concentration to replace $\|Z\|^2$ with N, and $\mathbb{E}Z_{(j)}^2$ with $(\mathbb{E}Z_{(j)})^2$, up to some errors that may just affect the constants then yields for $w \sim \mathrm{Unif}(\mathbb{S}^{N-1})$,

$$c\sqrt{\nu\log\frac{e}{\nu}} - o(1) \le \mathbb{E}\Big(\sum_{j=1}^{\nu N} w_{(j)}^2\Big)^{1/2} \le C\sqrt{\nu\log\frac{e}{\nu}} + o(1).$$

Thus for a typical $w \sim \mathrm{Unif}(\mathbb{S}^{n-1})$, we expect that $f(w)^2 = \sup_{\substack{A\subset[N]\\ \#A=\nu N}} \sum_{x\in A} w_x^2$ is around $C\nu \log \frac{e}{\nu}$.

7.3. Semicircle law

If X_N corresponds to a d-regular graph with $d \to \infty$, it is known to follow the semicircle law, meaning that when scaled as $X_N/\sqrt{d}$, its empirical eigenvalue distribution μ_N converges weakly in probability to the (non-random) semicircle distribution $\rho_{\mathrm{sc}}(x) = \frac{1}{2\pi}\sqrt{(4-x^2)_+}$. As usual, this can be done using the resolvent method (see for example §2.4.1 in the book [AGZ09]), with self-consistent equations (9.1.4), or by the moment method. The Stieltjes transform $S_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z}\, d\mu(x)$ of the semicircle law ρ_{sc} will be denoted

$$S_{\rho_{\mathrm{sc}}}(z) \equiv m_{\mathrm{sc}}(z) = -\frac{z}{2} + \frac{\sqrt{z^2-4}}{2},$$

which branch cut taken in $(-2,2)$ and $\sqrt{z^2-4} \approx z$ for large z. The semicircle law is equivalent to the statement that the Stieltjes transform $S_{\mu_N}(z) = \frac{1}{N}\operatorname{tr}\left(X_N/\sqrt{d} - z\right)^{-1}$ of the empirical eigenvalue distribution μ_N converges in probability to $m_{\mathrm{sc}}(z)$ for any fixed $z \in \mathbb{H}$ the upper half-plane. Note that the semicircle law does not predict the largest eigenvalue or norm $\|X_N\|$, since one can have a vanishing fraction of eigenvalues escape the semicircle. However, estimating $\|X_N\|$ can be done instead by comparing moments as in Theorem 6.1.

Going beyond the above global semicircle law, a *local* semicircle law states that $S_{\mu_N}(z) - m_{\mathrm{sc}}(z)$ is uniformly small for all $z \in \mathbb{H}$ with $\mathfrak{Im}\, z \geq \delta_N$ for some $\delta_N \to 0$. It is also often useful to know that the Green's function $G_N(z; x, x) := \left(X_N/\sqrt{d} - z\right)^{-1}_{xx}$ is close to $m_{\mathrm{sc}}(z)$ for all coordinates $x = 1, \ldots, N$, and we will need that here as well. The parameter δ_N represents the "spectral resolution", or roughly how large a window $\mathfrak{Im}\, S_{\mu_N}(z)$ sees. Since $\delta_N \to 0$, a local semicircle law then states that the empirical eigenvalue distribution still looks like $\rho_{\mathrm{sc}}(x)$ at a shrinking window of scale δ_N. Since eigenvalues typically have separation distance N^{-1}, the optimal δ_N is any $\delta_N \gg N^{-1}$. For general Wigner matrices, local semicircle laws have been proved on the essentially optimal scale $\mathfrak{Im}\, z \gg N^{-1}$, first for Wigner matrices in

[ESY09a], and later improved and generalized to other cases in many papers, see the notes [BGK16] for background and references in this direction.

In our model which allows for many different matrix structures and can be very sparse, one cannot hope to understand the spectrum at such a small scale. However, we will only actually need a weak local semicircle law that allows for $\mathfrak{Im}\, z \to 0$ at *some* (possibly highly non-optimal) rate.

CHAPTER 8

Proof of localization for $d \ll \log N$

The proof for localization will only require Gaussian concentration and suprema bounds. We first prove the following lemma.

Lemma 8.1. *Let X be a symmetric $N \times N$ matrix with $X_{xy} = \delta_{x\sim y} g_{xy}$, where g_{xy} are iid $\mathcal{N}(0,1)$ modulo symmetry requirements. Let d be the maximum degree of the associated graph, i.e. the maximum number of nonzero entries in a row. Then*

$$\mathbb{E} \sup_{v \in D_{\nu N,\kappa}} \|Xv\| \leq 3\frac{\sqrt{d}}{\sqrt{\nu}} + \kappa \mathbb{E}\|X\|. \tag{8.0.1}$$

Proof. Let $L = \nu N$ and split up the vector v via

$$\mathbb{E} \sup_{v \in D_{L,\kappa}} \|Xv\| \leq \mathbb{E} \sup_{v \in D_{L,\kappa}} \|X(v\mathbb{1}_{|v_j| \leq \frac{1}{\sqrt{L}}})\| + \mathbb{E} \sup_{v \in D_{L,\kappa}} \|X(v\mathbb{1}_{|v_j| > \frac{1}{\sqrt{L}}})\|. \tag{8.0.2}$$

The term with $|v_i| \leq \frac{1}{\sqrt{L}}$ will be bounded by the supremum bound for subgaussian random variables, while the term with $|v_i| > \frac{1}{\sqrt{L}}$ will be bounded using the definition of $D_{L,\kappa}$.

We start with $\mathbb{E}\sup_{v\in D_{L,\kappa}} \|X(v\mathbb{1}_{|v_i|>\frac{1}{\sqrt{L}}})\|$. For a unit vector v, the set $A := \{i : |v_i| > \frac{1}{\sqrt{L}}\}$ has cardinality $\leq L$. If $v \in D_{L,\kappa}$, then

$$\|(v\mathbb{1}_{|v_i|>\frac{1}{\sqrt{L}}})\|^2 = \sum_{i:|v_i|>\frac{1}{\sqrt{L}}} v_i^2 \leq \kappa^2,$$

so that

$$\sup_{v \in D_{L,\kappa}} \|X(v\mathbb{1}_{|v_i|>\frac{1}{\sqrt{L}}})\| \leq \sup_{v \in D_{L,\kappa}} \|X\|\|(v\mathbb{1}_{|v_i|>\frac{1}{\sqrt{L}}})\| \leq \kappa\|X\|. \tag{8.0.3}$$

For the other term, write

$$\sup_{v \in D_{L,\kappa}} \|X(v\mathbb{1}_{|v_i|\leq\frac{1}{\sqrt{L}}})\| \leq \sup_{w\in\mathbb{S}^{N-1}} \sup_{\|v\|_\infty \leq \frac{1}{\sqrt{L}}} \langle w, Xv\rangle = \frac{1}{\sqrt{L}} \sup_{w\in\mathbb{S}^{N-1}} \max_{v:v_i=\pm 1} \langle Xw, v\rangle,$$

since the supremum of $\langle Xw, v\rangle$ over $\|v\|_\infty \le \frac{1}{\sqrt{L}}$ occurs when $v_i = \frac{1}{\sqrt{L}}\operatorname{sgn}((Xw)_i)$. Thus

$$\sup_{v\in D_{L,\kappa}} \|X(v\mathbb{1}_{|v_i|\le\frac{1}{\sqrt{L}}})\| \le \frac{1}{\sqrt{L}}\max_{v:v_i=\pm1}\|Xv\|_2.$$

For $\|v\|_\infty \le 1$, $f(X) := \|Xv\|_2$ is $2d$-subgaussian on $(\mathbb{R}^\triangledown, \|\cdot\|_2)$, since for symmetric matrices $M = (x_{ij})$, $M' = (x'_{ij})$,

$$\begin{aligned}|f(M)-f(M')|^2 \le \|(M-M')v\|^2 &= \sum_{i=1}^N \left(\sum_{j:j\sim i}(x_{ij}-x'_{ij})v_j\right)^2 \le \sum_{i=1}^N\sum_{j:j\sim i}(x_{ij}-x'_{ij})^2\cdot d \\ &\le 2d\sum_{i=1}^N\sum_{\substack{j:j\sim i\\ j\ge i}}(x_{ij}-x'_{ij})^2 = 2d\|M-M'\|^2_{\mathbb{R}^\triangledown},\end{aligned}$$

where $\|\cdot\|_{\mathbb{R}^\triangledown}$ denotes the Euclidean norm on the nonzero entries in the upper triangular part of the structured matrix (these are the (i,j) so that $i\sim j$ and $j\ge i$). Since there are 2^N possible v's to maximize over, the supremum bound for subgaussian variables yields

$$\begin{aligned}\frac{1}{\sqrt{\nu N}}\mathbb{E}\max_{v_i=\pm1}\|Xv\|_2 &\le \frac{1}{\sqrt{\nu N}}\mathbb{E}\max_{v_i=\pm1}\left(|\|Xv\|-\mathbb{E}\|Xv\||+\mathbb{E}\|Xv\|\right) \\ &\le \frac{1}{\sqrt{\nu N}}\left(\sqrt{4d\log 2^N}+\sqrt{Nd}\right) \le 3\frac{\sqrt{d}}{\sqrt{\nu}},\end{aligned}$$

where we also used that

$$\max_{v_i=\pm1}\mathbb{E}\|Xv\| = \max_{v_i=\pm1}\mathbb{E}\sqrt{\sum_{j=1}^N\left|\sum_{k:k\sim j}X_{jk}v_k\right|^2} \le \max_{v_i=\pm1}\sqrt{\sum_{j=1}^N\mathbb{E}|\mathcal{N}(0,d)|^2} = \sqrt{Nd}.$$

Then with (8.0.3) we obtain (8.0.1). □

Proof of Theorem 6.2(i). If (6.2.3) holds, then

$$(1-\varepsilon)\mathbb{E}\|X\| \le \mathbb{E}\sup_{v\in D_{L,\kappa}}\|Xv\| \le 3\frac{\sqrt{d}}{\sqrt{\nu}}+\kappa\mathbb{E}\|X\|. \tag{8.0.4}$$

Isolating $\mathbb{E}\|X\|$ and using $\mathbb{E}\|X\| \ge C\sqrt{\log N}$ yields

$$\sqrt{\nu} \le \frac{3}{1-\varepsilon-\kappa}\frac{\sqrt{d}}{\mathbb{E}\|X\|} \le \frac{3}{1-\varepsilon-\kappa}\cdot\frac{\sqrt{d}}{C\sqrt{\log N}}. \tag{8.0.5}$$

□

CHAPTER 9

Proof of delocalization for $d \gg \log N$

9.1. Somewhat local semicircle law

We start to prove Theorem 6.2(ii). Recall $G_N(z;x,y) = \left(X_N/\sqrt{d} - z\right)^{-1}_{xy}$ is the Green's function. In this section, we state a (weak) local semicircle law, which will state that $\mathbb{E}G_N(z;x,x)$ is close to $m_{\mathrm{sc}}(z)$, on a somewhat local scale. This will allow us to replace $\mathbb{E}G_N(z;x,x)$ in approximations by the explicit and x-independent $m_{\mathrm{sc}}(z)$, as long as z does not approach the real line too quickly. This local semicircle law follows from a general result in [BBvH21] using free probability, or by a bootstrap method using the self-consistent equations like that described in [EY17, §7].

9.1.1. As a consequence of free probability. Consider an $N \times N$ random matrix of the form $X = \sum_{i=1}^n g_i A_i$, for g_i iid $\mathcal{N}(0,1)$ and A_i self-adjoint $N \times N$ matrices over $\mathbb{C}$. Define the free analogue $X_{\mathrm{free}} := \sum_{i=1}^n A_i \otimes s_i$, for $s_1, \ldots, s_n$ a free semicircular family in a C^*-probability space $(\mathcal{B}, \tau)$. The following was shown in [BBvH21].

Theorem 9.1 ([BBvH21], special case of Theorem 2.8)**.** *For the above model, for $z \in \mathbb{H}$, define the Stieltjes transforms $G(z), G_{\mathrm{free}}(z)$ as*

$$G(z) = \mathbb{E}(X - z)^{-1}, \quad G_{\mathrm{free}}(z) = (\mathrm{id} \otimes \tau)[(X_{\mathrm{free}} - z \otimes \mathbb{1}_{\mathcal{B}})^{-1}] \in M_{N\times N}(\mathbb{C}).$$

Define $\widetilde{v}(X)^2 := \|\operatorname{Cov}(X)\| \cdot \left\| \sum_{i=1}^n A_i^2 \right\|^{1/2}$. Then

$$\|G(z) - G_{\mathrm{free}}(z)\| \le \widetilde{v}(X)^4 |\mathfrak{Im}\, z|^{-5}. \tag{9.1.1}$$

In our case where X_N is associated with a d-regular graph and $X = \frac{X_N}{\sqrt{d}}$, we have $\frac{X_N}{\sqrt{d}} = \sum_{\substack{(i,j):i\sim j\\ i\geq j}} g_{ij} A_{(i,j)}$, where g_{ij} are iid $\mathcal{N}(0,1)$, and $A_{(i,j)} := \frac{1}{\sqrt{d}} \left(|i\rangle\langle j| + \delta_{i\neq j}|j\rangle\langle i|\right)$. Then

$$\Big\| \sum_{\substack{(i,j):i\sim j\\ i\geq j}} A^2_{(i,j)} \Big\| = \frac{1}{d} \Big\| \sum_{\substack{(i,j):i\sim j\\ i\geq j}} |i\rangle\langle i| + \delta_{i\neq j}|j\rangle\langle j| \Big\|$$
$$= \frac{1}{2d} \Big\| \sum_{(i,j):i\sim j} |i\rangle\langle i| + |j\rangle\langle j| \Big\| = \frac{1}{2d} \|2d \cdot \mathrm{Id}\| = 1,$$

and $\|\operatorname{Cov}(X_N/\sqrt{d})\| = \frac{1}{d}$, so that $\widetilde{v}(X)^4 = \frac{1}{d}$. Then it remains to find $G_{\text{free}}(z)$. The Stieltjes transform $G_{\text{free}}(z)$ satisfies the matrix equation (cf. [HT05, equation (1.5)]),

$$\sum_{\substack{(i,j):i\sim j\\ i\geq j}} A_{(i,j)} G_{\text{free}}(z) A_{(i,j)} G_{\text{free}}(z) + z G_{\text{free}}(z) + \mathrm{Id}_N = 0,$$

which can be re-written using $\mathbb{E}[g_{ij} g_{k\ell}] = \delta_{(i,j),(k,\ell)}$ as

$$-zG_{\text{free}}(z) - \frac{1}{d}\mathbb{E}\left[X_N G_{\text{free}}(z) X_N\right] G_{\text{free}}(z) = \mathrm{Id}_N\,. \tag{9.1.2}$$

This is a matrix equation of the form $VW + \eta(W)W = I$, considered in e.g. [HFS07], with $W = -iG_{\text{free}}(z)$, $V = -iz\,\mathrm{Id}$, and $\eta(W) = \frac{1}{d}\mathbb{E}\left[X_N W X_N\right]$. Note that $W_{\text{sc}}(z) := -im_{\text{sc}}(z)\,\mathrm{Id}_N$ satisfies equation (9.1.2) for $W = -iG_{\text{free}}(z)$. By [HFS07, Theorem 2.1], this is then the unique solution with $\mathfrak{Re}\, W > 0$. Note for $z \in \mathbb{H}$, that $\mathfrak{Im}\, G_{\text{free}}(z)$ is positive since X_{free} is self-adjoint and τ is completely positive. Then $\mathfrak{Re}(-iG_{\text{free}}(z)) = \mathfrak{Im}\, G_{\text{free}}(z) > 0$, and so by uniqueness we must have $G_{\text{free}}(z) = m_{\text{sc}}(z)\,\mathrm{Id}_N$.

Thus for any $z \in \mathbb{H}$, (9.1.1) reads

$$\left\|(X_N/\sqrt{d} - z)^{-1} - m_{\text{sc}}(z)\mathrm{Id}_N\right\| \leq \frac{1}{d|\,\mathfrak{Im}\, z|^5}. \tag{9.1.3}$$

9.1.2. Bootstrap method. This kind of method has been used often to prove local semicircle laws in random matrix theory, see for example the book [EY17, §7], or [EKYY13a, §3], [EKYY13b], among others. Our case is simpler and can be done with expectation values. The general idea is to obtain self-consistent equations for the Green's function with an error term, and then prove a stability result that implies the Green's function must be close to the error-term-free solution $m_{\text{sc}}(z)$, as the error term tends to zero. For our case, this can be done in the following steps.

(1) Prove stability ($|\mathbb{E}G_N(z;x,x) - m_{\text{sc}}(z)|$ is very small) for large $\mathfrak{Im}\, z$ using the self-consistent equations.
(2) Use the self-consistent equations again to prove a dichotomy result, that the quantity $|\mathbb{E}G_N(z;x,x) - m_{\text{sc}}(z)|$ is either somewhat large or very small.
(3) Bootstrap down from larger $\mathfrak{Im}\, z$ using the dichotomy result: Start with the large $\mathfrak{Im}\, z$ where $|\mathbb{E}G_N(z;x,x) - m_{\text{sc}}(z)|$ is very small, and use Lipschitz continuity to conclude the same must hold for sufficiently close z by the dichotomy. Continue extending by bootstrapping small distances away from these new z.

Lemma 9.2 (self-consistent equations). *For $z \in \mathbb{C} \setminus \mathbb{R}$ and $x, y \in [N]$, let $G_N(z;x,y) := \left(\frac{X_N}{\sqrt{d}} - z\right)^{-1}_{xy}$. Then*

$$\text{(9.1.4)} \quad z\mathbb{E}G_N(z;x,x) = -1 - \mathbb{E}G_N(z;x,x) \cdot \frac{1}{d} \sum_{\ell:\ell\sim x} \mathbb{E}G_N(z;\ell,\ell) + \Phi_{X_N,d,z,x,x}, \quad \forall x \in [N],$$

with $|\Phi_{X_N,d,z,x,x}| \le d^{-1}(|\mathfrak{Im}\, z|^{-2} + 2|\mathfrak{Im}\, z|^{-3})$. For the off-diagonal terms $x \neq y$,

$$\text{(9.1.5)} \quad z\mathbb{E}G_N(z;x,y) = -\mathbb{E}G_N(z;x,y) \cdot \frac{1}{d} \sum_{\ell:\ell\sim x} \mathbb{E}G_N(z;\ell,\ell) + \Phi_{X_N,d,z,x,y}, \quad \forall x, y \in [N],$$

with $|\Phi_{X_N,d,z,x,y}| \le d^{-1}(|\mathfrak{Im}\, z|^{-2} + 2|\mathfrak{Im}\, z|^{-3})$.

The proof is the standard one for Gaussian matrices, using integration by parts and concentration, just keeping track of the individual entries instead of taking the trace. The details are written in Appendix A.1.

As long as $d \to \infty$, (9.1.4) implies the global semicircle law since the matrix S, defined as $\frac{1}{d}$ times the adjacency matrix of the graph, is bistochastic. For a local semicircle law, we need to estimate how much each diagonal element $\mathbb{E}G_N(z;x,x)$ deviates from $m_{\text{sc}}(z)$, the unique solution in $\mathbb{H}$ (cf. [AEK19, Theorem 2.1]) to (9.1.4) when $\Phi_{X_N,d,z,x,x} = 0$.

Notation: Let $z = E + i\eta$, and let $g_{N,x}(z) := \mathbb{E}G_N(z;x,x)$ and $\widetilde{g}_{N,x}(z) := \mathbb{E}G_N(z;x,x) - m_{\mathrm{sc}}(z)$. Let g_N and $\widetilde{g}_N$ be the vectors $(g_{N,x})_{x=1}^N$ and $(\widetilde{g}_{N,x})_{x=1}^N$, respectively. If z is fixed or its value is clear we may omit it in $g_{N,x}$, g_N, $\widetilde{g}_{N,x}$, $\widetilde{g}_N$, or m_{sc}.

Lemma 9.3 (Large η bound)**.** *Suppose* (9.1.4) *holds for a z with $\eta \geq 2$. Then for this z,*

$$\max_{x=1,\ldots,N} |g_{N,x}(z) - m_{\mathrm{sc}}(z)| \leq \frac{C_z}{d}, \tag{9.1.6}$$

where $C_z < \infty$ is a z-dependent constant independent of N.

Proof. Same as in random matrix papers. The self-consistent equation (9.1.4) can be rearranged to read

$$g_{N,x} = \frac{-1 + \Phi_{X_N,d,z,x,x}}{z + (Sg_N)_x}.$$

Then using that $m_{\mathrm{sc}}(z) = -\frac{1}{z+m_{\mathrm{sc}}(z)}$,

$$\widetilde{g}_{N,x} = \frac{-1 + \Phi_{X_N,d,z,x,x}}{z + m_{\mathrm{sc}}(z) + (S\widetilde{g}_N)_x} + \frac{1}{z + m_{\mathrm{sc}}(z)} = \frac{(z + m_{\mathrm{sc}}(z))\Phi_{X_N,d,z,x,x} + (S\widetilde{g}_N)_x}{(z + m_{\mathrm{sc}}(z) + (S\widetilde{g}_N)_x)(z + m_{\mathrm{sc}}(z))}.$$

The denominator can be bounded from below, using $|z + m_{\mathrm{sc}}(z)| = |m_{\mathrm{sc}}(z)^{-1}| \geq |\operatorname{\mathfrak{Im}} z|$ and $|(S\widetilde{g}_N)_x| \leq \frac{2}{|\operatorname{\mathfrak{Im}} z|}$, to obtain for $|\operatorname{\mathfrak{Im}} z| \geq 2$,

$$|\widetilde{g}_{N,x}(z)| \leq \frac{(|z|+1)\cdot \Phi_{X_N,d,z,x,x} + |(S\widetilde{g}_N(z))_x|}{(|\operatorname{\mathfrak{Im}} z| - \frac{2}{|\operatorname{\mathfrak{Im}} z|})|\operatorname{\mathfrak{Im}} z|} \leq \frac{1}{2}(|z|+1)\frac{|\operatorname{\mathfrak{Im}} z| + 2}{|\operatorname{\mathfrak{Im}} z|^3 d} + \frac{1}{2}\max_k |\widetilde{g}_{N,k}|.$$

Taking the maximum over x yields (9.1.6) with $C_z = (|z|+1)(|\operatorname{\mathfrak{Im}} z| + 2)|\operatorname{\mathfrak{Im}} z|^{-3}$. ☐

The next lemma says that either $\max_x \widetilde{g}_{N,x}$ is somewhat large or very small. In particular, if we can show $\max_x |\widetilde{g}_{N,x}|$ is not very large then it must be very small.

Lemma 9.4 (Dichotomy)**.** *Let Ω be the rectangle $\Omega = \{E + i\eta : b_1 \leq E \leq b_2, 0 < \eta \leq b_3\}$, and by Lemma A.1 let $c > 0$ be chosen so that $1 - |m_{\mathrm{sc}}(z)|^2 \geq c \operatorname{\mathfrak{Im}} z$ in Ω. For a $z \in \Omega$, if both* (9.1.4) *and* $\max_x |\widetilde{g}_{N,x}(z)| \leq \frac{c}{2}|\operatorname{\mathfrak{Im}} z|$ *hold, then*

$$\max_x |\widetilde{g}_{N,x}(z)| \leq \frac{\frac{2}{c}(b_3 + 2)}{d|\operatorname{\mathfrak{Im}} z|^4}. \tag{9.1.7}$$

Proof. Using $zm_{\mathrm{sc}}(z) = -1 - m_{\mathrm{sc}}(z)^2$ along with (9.1.4) yields the self-consistent equations for $\widetilde{g}_N = g_N - m_{\mathrm{sc}}$,

$$z\widetilde{g}_{N,x} = -\left(\widetilde{g}_{N,x}(S\widetilde{g}_N)_x + m_{\mathrm{sc}}(S\widetilde{g}_N)_x + m_{\mathrm{sc}}\widetilde{g}_{N,x}\right) + \Phi_{X_N,d,z,x,x}, \tag{9.1.8}$$

with the same error term $|\Phi_{X_N,d,z,x,x}| \le \frac{C_\Omega}{d|\mathfrak{Im}\, z|^3}$, where $C_\Omega := b_3 + 2$. Then taking absolute values and maximums yields,

$$\begin{aligned} |z + m_{\mathrm{sc}}||\widetilde{g}_{N,x}| &\le |\widetilde{g}_{N,x}||(S\widetilde{g}_N)_x| + |m_{\mathrm{sc}}||(S\widetilde{g}_N)_x| + \frac{C_\Omega}{d|\mathfrak{Im}\, z|^3} \\ &\le \max_k |\widetilde{g}_{N,k}|^2 + |m_{\mathrm{sc}}| \max_k |\widetilde{g}_{N,k}| + \frac{C_\Omega}{d|\mathfrak{Im}\, z|^3}. \end{aligned}$$

Taking the maximum over x and multiplying through by $|m_{\mathrm{sc}}| \le 1$,

$$\max_x |\widetilde{g}_{N,x}|(1 - |m_{\mathrm{sc}}|^2) \le \max_x |\widetilde{g}_{N,x}^2| + \frac{C_\Omega}{d|\mathfrak{Im}\, z|^3},$$

and then using $1 - |m_{\mathrm{sc}}(z)|^2 \ge c\,\mathfrak{Im}\, z$ and $\max_x |\widetilde{g}_{N,x}| \le \frac{c}{2}\mathfrak{Im}\, z$ yields

$$\begin{aligned} \max_j |\widetilde{g}_{N,x}|(c\,\mathfrak{Im}\, z - \max_x |\widetilde{g}_{N,x}|) &\le \frac{C_\Omega}{d|\mathfrak{Im}\, z|^3} \\ \max_x |\widetilde{g}_{N,x}| &\le \frac{2C_\Omega}{cd|\mathfrak{Im}\, z|^4}. \end{aligned}$$

□

Stability for z with $\eta < 2$ will then follow by bootstrapping the bound in Lemma 9.3 to smaller η using Lemma 9.4.

Proposition 9.5 (Diagonal stability). *Suppose* (9.1.4) *holds in the rectangle* $\Omega = \{E + i\eta : b_1 \le E \le b_2, 0 < \eta \le b_3\}$ *where* $b_3 \ge 2$*, and let* $c > 0$ *be chosen so that* $1 - |m_{\mathrm{sc}}(z)|^2 \ge c\,\mathfrak{Im}\, z$ *in* Ω*. Then there is a constant* $C(\Omega)$ *so that for any* $z \in \Omega$ *and* $d \ge \frac{4C(\Omega)}{c|\mathfrak{Im}\, z|^5}$,

$$\max_{x=1,\dots,N} |g_{N,x}(z) - m_{\mathrm{sc}}(z)| \le \frac{C(\Omega)}{d|\mathfrak{Im}\, z|^4}. \tag{9.1.9}$$

Proof. Take $C(\Omega) := \max\left(\frac{2}{c}(b_3+2), (1 + \sqrt{\max(b_1^2, b_2^2) + b_3^2})(b_3+2)b_3\right)$. Then Lemma 9.3 covers the case when $|\mathfrak{Im}\, z| \ge 2$, since $\frac{C(\Omega)}{|\mathfrak{Im}\, z|^4} \ge \sup_{z\in\Omega, |\mathfrak{Im}\, z| \ge 2} C_z$. Fix $z = E + i\eta$ with $\eta < 2$, and let $\Omega_\eta := \Omega \cap \{w \in \mathbb{C} : \mathfrak{Im}\, w \geq \eta\}$. Since

$$\sup_{w\in\Omega_\eta}\left|\frac{\partial \mathbb{E}G_N(w;x,y)}{\partial w}\right| \leq \frac{1}{\eta^2}, \qquad \sup_{w\in\Omega_\eta}\left|\frac{\partial m_{\mathrm{sc}}(w)}{\partial w}\right| \leq \frac{1}{2}\left(1+\frac{\sup_{w\in\Omega_\eta}|w|}{\sqrt{2\eta}}\right), \tag{9.1.10}$$

each $|\widetilde{g}_{N,x}|$ is Lipschitz on Ω_η with some Lipschitz constant $L(\eta,\Omega)$, and thus so is the function $\max_{x\in[N]}|\widetilde{g}_{N,x}|$. Let $z_0 = E + 2i$ and set $\alpha := \frac{c\eta}{4L(\eta,\Omega)}$ the step size. Set $z_k = z_{k-1} - \alpha i$ for $k = 1, \ldots, k_f$, stopping at $z_{k_f} := z = E + i\eta$ (the last step size may be smaller than α), so that $k_f \leq \lceil \frac{2-\eta}{\alpha} \rceil < \infty$. We show by induction that

$$\max_{x\in[N]}|\widetilde{g}_{N,x}(z_k)| \leq \frac{C(\Omega)}{d\eta^4}, \tag{9.1.11}$$

for each k. The base case $k = 0$ is covered by Lemma 9.3. If (9.1.11) holds for $k-1$, then since $\max_x |\widetilde{g}_{N,x}|$ is $L(\eta,\Omega)$-Lipschitz in z,

$$\begin{aligned}\max_{x\in[N]}|\widetilde{g}_{N,x}(z_k)| &\leq |z_k - z_{k-1}|L(\eta,\Omega) + \max_{x\in[N]}|\widetilde{g}_{N,x}(z_{k-1})|\\ &\leq \frac{c\eta}{4} + \frac{C(\Omega)}{d\eta^4} \leq \frac{c\eta}{2},\end{aligned}$$

so Lemma 9.4 implies $\max_x |\widetilde{g}_{N,x}(z_k)| \leq \frac{C(\Omega)}{d\eta^4}$. □

Proposition 9.6 (Off-diagonal stability)**.** *Suppose* (9.1.5) *holds for all* z *in the rectangle* $\Omega = \{E + i\eta : b_1 \leq E \leq b_2, 0 < \eta \leq b_3\}$ *where* $b_3 \geq 2$*, and let* c *and* $C(\Omega)$ *be defined as in Proposition 9.5. Then there is* $C'(\Omega)$ *so that for any* $z \in \Omega$ *and* $d \geq \max\left(\frac{4C(\Omega)}{c|\mathfrak{Im}\,z|^5}, \frac{2C(\Omega)}{|\mathfrak{Im}\,z|^4}\right)$,

$$|\mathbb{E}G_N(z;x,y)| \leq \frac{C'(\Omega)}{d|\mathfrak{Im}\,z|^3}, \quad x \neq y. \tag{9.1.12}$$

Proof. Recall the self-consistent equations for the off-diagonal terms is (9.1.5). Using Proposition 9.5, (9.1.5) becomes

$$z\mathbb{E}G_N(z;x,y) = -\mathbb{E}G_N(z;x,y)\cdot(m_{\mathrm{sc}}(z) + \Xi_{X_N,d,x}(z)) + \Phi_{X_N,d,z,x,y}, \tag{9.1.13}$$

with

$$|\Xi_{X_N,d,x}(z)| \leq \frac{C(\Omega)}{d|\mathfrak{Im}\,z|^4} \leq \frac{1}{2}.$$

Thus

$$|\mathbb{E}G_N(z;x,y)| = \frac{|\Phi_{X_N,d,z,x,y}|}{|z+m_{\mathrm{sc}}(z)| - |\Xi_{X_N,d,x}(z)|} \leq \frac{4+2|\,\mathfrak{Im}\,z|}{d|\,\mathfrak{Im}\,z|^3},$$

since $|z+m_{\mathrm{sc}}(z)| > 1$. $\square$

9.2. Projection matrix approximation

Let $P_{[2-\varepsilon_N,b]}$ be the projection matrix of $X_N/\sqrt{d}$ onto the subspace spanned by the eigenvectors $\{\psi^{(N,j)} : \lambda^{(N,j)} \in [2-\varepsilon_N, b]\}$, where $b > 2$ is a fixed number like $b = 10$. We will estimate the matrix elements of $P_{[2-\varepsilon_N,b]}$ using expressions involving resolvents.

Let $a_N = 2 - \varepsilon_N$, and approximate the indicator function of $[a_N, b]$ using arctangents, which will then involve the Green's function which will be close to $m_{\mathrm{sc}}(z)$. One motivation for the approximation is Stone's formula, or alternatively just properties of the Poisson kernel on $\mathbb{H}$.

Proposition 9.7 (Stone's formula)**.** *If H is self-adjoint, as $\delta \to 0$*

$$\frac{1}{\pi}\int_a^b \mathfrak{Im}\,\frac{1}{H-(\lambda+i\delta)}\,d\lambda \xrightarrow{s} \frac{1}{2}(P_{(a,b)} + P_{[a,b]}), \tag{9.2.1}$$

where convergence is in the strong operator topology.

For the diagonal matrix elements, if we have enough control on the convergence of $\mathbb{E}G_N(z;j_N,j_N) \to m_{\mathrm{sc}}(z)$, then we can replace $\mathbb{E}G_N(z;j_N,j_N)$ by $m_{\mathrm{sc}}(z)$ and compute the expected projection matrix elements as $N \to \infty$ using $m_{\mathrm{sc}}(z)$. More specifically, let

$$a_N = 2 - \varepsilon_N$$

$$a_N^- = 2 - \varepsilon_N - \gamma_N$$

$$a_N^+ = 2 - \varepsilon_N + \gamma_N,$$

where we will take $\delta_N \ll \gamma_N \ll \varepsilon_N$, and set

$$A^-_{N,\delta} = \frac{1}{\pi}\int_{a_N^-}^{b} \mathfrak{Im}\left(\frac{X}{\sqrt{d}} - z\right)^{-1} d\lambda \tag{9.2.2}$$

$$A^+_{N,\delta} = \frac{1}{\pi}\int_{a_N^+}^{b} \mathfrak{Im}\left(\frac{X}{\sqrt{d}} - z\right)^{-1} d\lambda. \tag{9.2.3}$$

$A^-_{N,\delta}$ approximates the projection matrix diagonal elements roughly mostly from above, while $A^+_{N,\delta}$ approximates the projection matrix diagonal elements roughly mostly from below. We will take $\delta_N \to 0$ so that $(A^-_{N,\delta_N})_{xx}$ and $(A^+_{N,\delta_N})_{xx}$ are close to $(P_{[a_N,b]})_{xx}$, but will take $\delta_N \to 0$ sufficiently slowly that we can guarantee $G_N(\lambda + i\delta_N; x, x)$ is close to $m_{\rm sc}(\lambda + i\delta_N)$. In the following estimates we use the asymptotic $\tan^{-1} x = \frac{\pi}{2} - \frac{1}{x} + \mathcal{O}(x^{-3})$ as $x \to \infty$, and relation $0 \le \tan^{-1}\left(\frac{b-\lambda^{(j)}}{\delta}\right) - \tan^{-1}\left(\frac{a-\lambda^{(j)}}{\delta}\right) \le \pi$ for $a \le b$.

Assume the spectrum of $X_N/\sqrt{d}$ is bounded away below b. For example, take $b = 10$, so the largest eigenvalue is $< b$ with very high probability, by Theorem 6.1 and matrix norm concentration. By the spectral theorem, we can compute $(A^{\pm}_{N,\delta})_{xy}$ as follows, using

$$\begin{aligned}\int_a^b \mathfrak{Im}\,\frac{1}{\lambda^{(j)} - E - i\delta_N}\,dE &= \int_a^b \frac{\delta_N}{(\lambda^{(j)} - E)^2 + \delta_N^2}\,dE \\ &= \tan^{-1}\left(\frac{b-\lambda^{(j)}}{\delta_N}\right) - \tan^{-1}\left(\frac{a-\lambda^{(j)}}{\delta_N}\right).\end{aligned}$$

9.2.1. Diagonal entry estimates. For the upper approximation, neglecting the sum over $\lambda^{(j)} < a_N$ which is nonnegative, and then using the asymptotic for arctangent, yields,

(9.2.4)

$$\begin{aligned}(A^-_{N,\delta_N})_{xx} &= \left(\sum_{j:\lambda^{(j)}\ge a_N} + \sum_{j:\lambda^{(j)}<a_N}\right)\frac{1}{\pi}\left(\tan^{-1}\left(\frac{b-\lambda^{(j)}}{\delta_N}\right) - \tan^{-1}\left(\frac{a_N-\gamma_N-\lambda^{(j)}}{\delta_N}\right)\right)|\psi_x^{(j)}|^2 \\ &\ge \sum_{j:\lambda^{(j)}\ge a_N}\frac{1}{\pi}\left(\frac{\pi}{2} - \mathcal{O}(\delta_N) - \left(-\frac{\pi}{2} + \mathcal{O}(\delta_N/\gamma_N)\right)\right)|\psi_x^{(j)}|^2 \\ &\ge \sum_{j:\lambda^{(j)}\ge a_N}(1 - \mathcal{O}(\delta_N/\gamma_N))\,|\psi_x^{(j)}|^2 \\ &= (P_{[a_N,b]})_{xx}(1 - \mathcal{O}(\delta_N/\gamma_N))\end{aligned}$$

since if $\lambda^{(j)} \geq a_N$ then $\frac{a_N - \gamma_N - \lambda^{(j)}}{\delta_N} \leq -\frac{\gamma_N}{\delta_N} \to -\infty$ if $\delta_N \ll \gamma_N$.

For the lower approximation,

(9.2.5)

$$
\begin{aligned}
(A^+_{N,\delta_N})_{xx} &= \left(\sum_{j:\lambda^{(j)} < a_N} + \sum_{j:\lambda^{(j)} \geq a_N} \right) \frac{1}{\pi} \left(\tan^{-1}\left(\frac{b - \lambda^{(j)}}{\delta_N} \right) - \tan^{-1}\left(\frac{a_N + \gamma_N - \lambda^{(j)}}{\delta_N} \right) \right) |\psi_x^{(j)}|^2 \\
&\leq \sum_{j:\lambda^{(j)} < a_N} \frac{1}{\pi} \left(\tan^{-1}\left(\frac{b - \lambda^{(j)}}{\delta_N} \right) - \tan^{-1}\left(\frac{a_N + \gamma_N - \lambda^{(j)}}{\delta_N} \right) \right) |\psi_x^{(j)}|^2 + (P_{[a_N,b]})_{xx} \\
&= \sum_{j:\lambda^{(j)} < a_N} \frac{1}{\pi} \left[\frac{\pi}{2} - \mathcal{O}(\delta_N) - \left(\frac{\pi}{2} - \mathcal{O}(\delta_N/\gamma_N) \right) \right] |\psi_x^{(j)}|^2 + (P_{[a_N,b]})_{xx} \\
&= \mathcal{O}(\delta_N/\gamma_N) + (P_{[a_N,b]})_{xx}.
\end{aligned}
$$

since if $\lambda^{(j)} < a_N$ then $\frac{a_N + \gamma_N - \lambda^{(j)}}{\delta_N} > \frac{\gamma_N}{\delta_N} \to +\infty$.

Thus if $\max \lambda^{(j)} \leq \beta < b$ and $\delta_N \ll \gamma_N$, then

$$
(A^+_{N,\delta})_{xx} - \mathcal{O}\Big(\frac{\delta_N}{\gamma_N}\Big) \leq (P_{[a_N,b]})_{xx} \leq (A^-_{N,\delta})_{xx} \left(1 + \mathcal{O}\Big(\frac{\delta_N}{\gamma_N}\Big) \right). \tag{9.2.6}
$$

We will see later that $(P_{[a_N,b]})_{xx} \approx \frac{2}{3\pi} \varepsilon_N^{3/2}$, so in order for the error terms above to be $o(\varepsilon_N^{3/2})$, we will need $\delta_N \gamma_N^{-1} \ll \varepsilon_N^{3/2}$.

9.3. Replacement with semicircle law

Next, we want to replace the Green's function in the integral expression of $A^{\pm}_{N,\delta_N}$ with the Stieltjes transform $m_{\mathrm{sc}}(z)$ of the semicircle law. The estimates for this are the weak local semicircle stability results from Section 9.1, and Gaussian concentration for the Green's function. These estimates control the error terms from replacing $G_N(z;x,x)$ with $\mathbb{E}G_N(z;x,x)$ and then with $m_{\mathrm{sc}}(z)$. Then evaluating the integral in $A^{\pm}_{N,\delta_N}$ with $m_{\mathrm{sc}}(z)$ instead yields the desired result.

Theorem 9.8 (Projection matrix elements). *Suppose $d/\log N \to \infty$ and suppose that for each fixed $z \in \mathbb{H}$, that*

$$\max_{x\in[N]} |\mathbb{E}G_N(z;x,x) - m_{\mathrm{sc}}(z)| \to 0, \qquad \text{as } N \to \infty. \tag{9.3.1}$$

Let $P_{[a,b]}$ be the spectral projection matrix of $X_N/\sqrt{d}$ onto the interval $[a,b]$. Then there is a sequence $\varepsilon_N \to 0$ so that with probability at least $1-o(1)$, the matrix $P_{[2-\varepsilon_N,b]}$ has diagonal elements

$$(P_{[2-\varepsilon_N,b]})_{xx} = \frac{2}{3\pi}\varepsilon_N^{3/2}(1+o(1)), \tag{9.3.2}$$

with the error term uniform in $x \in [N]$.

9.3.1. Proof of Theorem 9.8. A uniform continuity argument (Lemma A.3) shows that (9.3.1) implies the existence of a sequence $\widetilde{\delta}_N > 0$, $\widetilde{\delta}_N \to 0$, so that as $N \to \infty$,

$$\mathcal{E}(N,\widetilde{\delta}_N) := \sup_{\widetilde{\delta}_N \le \gamma \le 1} \sup_{\lambda\in[0,b]} \max_{x\in[N]} |\mathbb{E}G_N(\lambda+i\gamma;x,x) - m_{\mathrm{sc}}(\lambda+i\gamma)| \to 0. \tag{9.3.3}$$

Any sequence $\alpha_N \to 0$ with $\alpha_N \ge \widetilde{\delta}_N$ will also satisfy $\mathcal{E}(N,\alpha_N) \to 0$. Then take

$$\delta_N := \max\left(\left(\frac{16\log N}{d}\right)^{5/22}, \widetilde{\delta}_N\right) \tag{9.3.4}$$

$$\varepsilon_N \gg \max(\mathcal{E}(N,\delta_N)^2, \delta_N^{2/5}), \quad \varepsilon_N \to 0 \tag{9.3.5}$$

$$t_N := \delta_N^{1/5} \tag{9.3.6}$$

$$\gamma_N := \varepsilon_N^{1/2}\delta_N^{1/5}, \tag{9.3.7}$$

so that $t_N \ll \varepsilon_N^{1/2}$, and $\delta_N \ll \delta_N^{2/5} \ll \gamma_N \ll \varepsilon_N$, and $\frac{\delta_N}{\gamma_N} \ll \varepsilon_N^{3/2}$. Since $\delta_N \to 0$, this will allow us to approximate $P_{[2-\varepsilon_N,b]}$ well using $A^{\pm}_{N,\delta_N}$ which has the Green's function. Then we replace the Green's function with its expectation by concentration of measure, and then since $\mathcal{E}(N,\delta_N) \le \mathcal{E}(N,\widetilde{\delta}_N) \to 0$, this will then let us replace the Green's function with $m_{\mathrm{sc}}(z)$.

Remark 9.3.1. Equations (9.1.3) or (9.1.9) from the (somewhat) local semicircle law give a quantitative bound on $\mathcal{E}(N, \delta_N)$, which allows us to take any sequence $\varepsilon_N > 0$ satisfying

$$\varepsilon_N \to 0, \quad \varepsilon_N \gg \delta_N^{2/5} \sim \max\left[\left(\frac{\log N}{d}\right)^{1/11}, \frac{1}{d^{2/25}}\right], \tag{9.3.8}$$

by taking for example

$$\delta_N := \max\left(\left(\frac{16\log N}{d}\right)^{5/22}, \left(\frac{4C(\Omega)}{cd}\right)^{1/5}\right) \tag{9.3.9}$$

$$t_N := \delta_N^{1/5}, \quad \gamma_N := \varepsilon_N^{1/2}\delta_N^{1/5}, \tag{9.3.10}$$

where the constants $C(\Omega)$ and c are the same as in Propositions 9.5.

Returning to the proof of Theorem 9.8, we start with the estimate for replacing $G_N(z; x, y)$ by its expectation.

Lemma 9.9 (Green's function concentration)**.** *For any $t \geq 0$, $z \in \mathbb{H}$, and $x, y \in [N]$,*

$$\mathbb{P}\left[|G_N(z; x, y) - \mathbb{E}G_N(z; x, y)| \geq t\right] \leq 4\exp\left(-\frac{1}{8}t^2 d|\mathfrak{Im}\, z|^4\right). \tag{9.3.11}$$

Proof. The Lipschitz constant for $f(X) = \langle x|(X/\sqrt{d} - z)^{-1}|y\rangle$ is bounded above by $\frac{\sqrt{2}}{\sqrt{d}|\mathfrak{Im}\, z|^2}$ by Example 7.1. $\square$

By Lemma 9.9 and a union bound,

$$\begin{aligned}
\mathbb{P}\left[\max_{x\in[N]}\left|G_N(\lambda + i\delta_N; x, x) - \mathbb{E}G_N(\lambda + i\delta_N; x, x)\right| \geq t_N\right]& \\
&\leq 4N\exp\left(-\frac{1}{8}t_N^2 d\delta_N^4\right) \\
&= 4\exp\left(-\log N\left[\frac{1}{8}t_N^2\delta_N^4\frac{d}{\log N} - 1\right]\right) \\
&\leq 4\exp\left(-\log N\right) = \frac{4}{N},
\end{aligned}$$

where we used (9.3.6) and the first term in the max() in the definition (9.3.4) of δ_N. Thus with probability at least $1 - \frac{4}{N}$, all $G_N(\lambda + i\delta_N; x, x)$ are within t_N of their expected values.

From (9.3.3), the expected value $\mathbb{E}G_N(\lambda + i\delta_N; x, x)$ will be close to $m_{\mathrm{sc}}(\lambda + i\delta_N)$. For $\star \in \{+, -\}$, write

$$(9.3.12)\quad (A^{\star}_{N,\delta_N})_{xx} = \frac{1}{\pi}\mathfrak{Im}\int_{a_N^\star}^{b} m_{\mathrm{sc}}(\lambda + i\delta_N)\, d\lambda + \\ + \frac{1}{\pi}\mathfrak{Im}\int_{a_N^\star}^{b} \mathbb{E}G_N(\lambda + i\delta_N; x, x) - m_{\mathrm{sc}}(\lambda + i\delta_N)\, d\lambda + \\ + \frac{1}{\pi}\mathfrak{Im}\int_{a_N^\star}^{b} G_N(\lambda + i\delta_N; x, x) - \mathbb{E}G_N(\lambda + i\delta_N; x, x)\, d\lambda.$$

The two error terms are bounded as follows. For N sufficiently large that $\delta_N \leq 1$, then by (9.3.3) and (9.3.5), the first error term is

$$(9.3.13)\quad \left|\frac{1}{\pi}\mathfrak{Im}\int_{a_N^\star}^{b} \mathbb{E}G_N(\lambda + i\delta_N; x, x) - m_{\mathrm{sc}}(\lambda + i\delta_N)\, d\lambda\right| \leq \frac{\varepsilon_N + \gamma_N}{\pi}\mathcal{E}(N, \delta_N) = o(\varepsilon_N^{3/2}).$$

For the second error term, with probability at least $1 - \frac{4}{N}$,

(9.3.14)

$$\max_{x=1,\dots,N}\left|\frac{1}{\pi}\mathfrak{Im}\int_{a_N^\star}^{b} G_N(\lambda + i\delta_N; x, x) - \mathbb{E}G_N(\lambda + i\delta_N; x, x)\, d\lambda\right| \leq \frac{\varepsilon_N + \gamma_N}{\pi}\cdot t_N = o(\varepsilon_N^{3/2}).$$

For $\delta_N \ll \varepsilon_N^2$, the integration of m_{sc} is (computation in Lemma A.2)

$$\frac{1}{\pi}\mathfrak{Im}\int_{a_N^\star}^{b} m_{\mathrm{sc}}(\lambda + i\delta_N)\, d\lambda = \frac{2}{3\pi}\varepsilon_N^{3/2}(1 + o(1)).$$

Thus with probability at least $1 - \frac{4}{N}$, for $\star \in \{+, -\}$ and any $x \in [N]$,

$$(9.3.15)\quad (A^{\star}_{N,\delta_N})_{xx} = \frac{2}{3\pi}\varepsilon_N^{3/2}(1 + o(1)),$$

with the error term uniform in x. Combining with (9.2.6), noting that $\frac{\delta_N}{\gamma_N} = o(\varepsilon_N^{3/2})$, then with probability at least $1 - o(1)$,

$$(9.3.16)\quad (P_{[2-\varepsilon_N, b]})_{xx} = \frac{2}{3\pi}\varepsilon_N^{3/2}(1 + o(1)),$$

with the error term uniform in x, which is (9.3.2).

9.4. Delocalized random vector

Since $P_{[2-\varepsilon_N,b]}$ is the projection matrix onto the space spanned by the eigenvectors with eigenvalues in $[2-\varepsilon_N, b]$, any vector in this subspace will satisfy $\|X_N v\|_2 \geq (2-\varepsilon_N)\sqrt{d}$. We just need to find a vector in this subspace that is delocalized. This will follow from Theorem 9.8, which describes the covariance matrix of a random Gaussian from this subspace, and the following consequences.

Lemma 9.10 (Gaussian moments)**.** *For $N \in \mathbb{N}$, let $P^{(N)}$ be orthogonal projection onto a subspace $V^{(N)} \subset \mathbb{R}^N$ of dimension m_N with $m_N \to \infty$ as $N \to \infty$, and suppose*

$$(P^{(N)})_{xx} \equiv \sum_{j=1}^{m_N} |\psi_x^{(N,j)}|^2 = \frac{m_N}{N}(1+o(1)), \tag{9.4.1}$$

with error term uniform in $x \in [N]$. Choose $q > 2$. If $w \in V^{(N)}$ is a unit vector chosen randomly from Haar measure on the unit sphere of $V^{(N)}$, then as $N \to \infty$,

$$\mathbb{E}\left[\sum_{x=1}^{N} |w_x|^q\right] = \frac{C_q}{N^{q/2-1}}(1+o(1)), \tag{9.4.2}$$

where $C_q = \mathbb{E}|g|^q = \frac{2^{q/2}\Gamma\left(\frac{q+1}{2}\right)}{\sqrt{\pi}}$ with $g \sim \mathcal{N}(0,1)$.

Proof. Let $(\psi^{(N,j)})$ be an orthonormal basis for $V^{(N)}$, and write $w = \sum_{j=1}^{m_N} \alpha_j \psi^{(N,j)}$, where $\alpha \sim \mathrm{Unif}(\mathbb{S}^{m_N-1})$. Since w is essentially multivariate Gaussian $\frac{1}{m_N^{1/2}}\mathcal{N}(0,P^{(N)})$, (9.4.1) follows from computing multivariate Gaussian moments.

More carefully, $\mathcal{N}(0,P^{(N)}) \sim rw$, where $r^2 \sim \chi^2(m_N)$ is independent of w.[1] Then $\mathbb{E}|w_x|^q = \frac{1}{\mathbb{E}r^q}\mathbb{E}|Z_x|^q$, for $(Z_1,\ldots,Z_n) \sim \mathcal{N}(0,P^{(N)})$. The chi-square distribution $\chi^2(m)$ has probability density function

$$f_m(t) = \begin{cases} \frac{t^{m/2-1}e^{-t/2}}{2^{m/2}\Gamma\left(\frac{m}{2}\right)}, & t \geq 0 \\ 0, & t < 0 \end{cases},$$

[1] For $A^{(N)}$ the $N \times m_N$ matrix whose columns are the vectors $\psi^{(N,j)}$, then $\mathcal{N}(0,P^{(N)}) \sim A^{(N)}\mathcal{N}(0,I_{m_N}) \sim rA^{(N)}\alpha \sim rw$, where $\alpha \in \mathrm{Unif}(\mathbb{S}^{m_N-1})$, and r and α are independent by writing the density of $\mathcal{N}(0,I_{m_N})$ in polar coordinates which factors.

so that

$$\mathbb{E}r^q = \int_0^\infty t^{q/2} \frac{t^{m_N/2-1}e^{-t/2}}{2^{m_N/2}\Gamma\left(\frac{m_N}{2}\right)}\,dt = 2^{q/2}\frac{\Gamma\left(\frac{m_N+q}{2}\right)}{\Gamma\left(\frac{m_N}{2}\right)} \sim m_N^{q/2}, \text{ as } m_N \to \infty.$$

Since $(Z_1, \ldots, Z_n) \sim \mathcal{N}(0, P^{(N)})$, each Z_x by itself is centered normal with variance $P^{(N)}_{xx}$, so

$$\mathbb{E}|Z_x|^q = (P^{(N)}_{xx})^{q/2}\mathbb{E}|g|^q = (P^{(N)}_{xx})^{q/2}C_q,$$

where $g \sim \mathcal{N}(0,1)$. Thus

$$\mathbb{E}|w_x|^q = C_q(P^{(N)}_{xx})^{q/2}\frac{\Gamma\left(\frac{m_N}{2}\right)}{2^{q/2}\Gamma\left(\frac{m_N+q}{2}\right)} = \frac{C_q}{N^{q/2}}(1+o(1)).$$

□

Since the qth moment of some w is bounded (after suitable scaling), this lets us find a $\nu > 0$ for delocalization. Essentially, if we had to take $\nu \to 0$ to ensure $\sup_{\substack{A\subset[N]\\ \#A=\nu N}} \sum_{x\in A}|w_x|^2 \le \kappa^2$, then all qth moments of $\sqrt{N}w$ divided by N must diverge as $N \to \infty$ for $q > 2$.

Corollary 9.11 (delocalization and moments)**.** *Fix $0 < \kappa < 1$, and let $P^{(N)}$ and $V^{(N)}$ be as in Lemma 9.10, including* (9.4.1)*. There for any $0 < \nu < \frac{c\kappa^2}{\log\frac{1}{\kappa}}$, where $0 < c < 1$ is an absolute constant, and for sufficiently large N, there is a $(\nu N, \kappa)$-delocalized unit vector $w \in V^{(N)}$.*

Proof. Let $q = 4\log\frac{e}{\kappa} \ge 4$. By Lemma 9.10, there is a unit vector $w \in V^{(N)}$ so that $N^{q/2-1}\sum_{x=1}^N |w_x|^q \le C_q + o(1)$ as $N \to \infty$. By Hölder, then for any set $A \subset [N]$ with $\#A = \nu N$,

$$\sum_{x\in A}|w_x|^2 \le (\nu N)^{1-2/q}\left(\sum_{x\in A}|w_x|^q\right)^{2/q} \le \nu^{1-2/q}C_q^{2/q} + o(1). \tag{9.4.3}$$

By Stirling's formula inequality,

$$C_q^{2/q} = \frac{2\left(\Gamma\left(\frac{q+1}{2}\right)\right)^{2/q}}{\pi^{1/q}} \le \frac{2(q-1)}{(2e)^{1-1/q}}e^{\frac{1}{3q(q-1)}} \le \frac{8\log\frac{e}{\kappa}}{(2e)^{3/4}}e^{1/36} = C\log\frac{e}{\kappa}.$$

So for $\nu \leq \frac{c\kappa^2}{\log \frac{e}{\kappa}}$,

$$\nu^{1-2/q} C_q^{2/q} \leq \frac{c^{1-2/q}(\kappa^2)^{1-2/q}}{(\log \frac{e}{\kappa})^{1-2/q}} \cdot C \log \frac{e}{\kappa} \leq c^{1/2} C \kappa^2 \cdot \frac{(\log \frac{e}{\kappa})^{2/q}}{(\kappa^2)^{2/q}}.$$

Since $q = 4 \log \frac{e}{\kappa}$, then $(\log \frac{e}{\kappa})^{2/q} = \exp\left(\frac{1}{2\log \frac{e}{\kappa}} \log\log \frac{e}{\kappa}\right) \leq \widetilde{C}$, since the expression is continuous in κ and tends to 1 as $\kappa \to 0$ or $\kappa \to 1$. Similarly,

$$(\kappa^2)^{2/q} = \exp\left(\frac{-\log \frac{1}{\kappa}}{1 + \log \frac{1}{\kappa}}\right) \geq \frac{1}{e},$$

since $0 < \log \frac{1}{\kappa} < \infty$. Thus $\nu^{1-2/q} C_q^{2/q} \leq c^{1/2} C \widetilde{C} e \kappa^2$, so choosing c a sufficiently small absolute constant ensures that (9.4.3) becomes,

$$\sup_{\substack{A \subset [N] \\ \#A = \nu N}} \sum_{x \in A} |w_x|^2 < (1-\delta)\kappa^2 + o(1), \text{ for some } \delta > 0,$$

which will eventually be $\leq \kappa^2$. Thus for sufficiently large N (depending on κ and ν), a w chosen in this way is $(\nu N, \kappa)$-delocalized. □

Remark 9.4.1. (a) We compare the condition $\nu < \frac{c\kappa^2}{\log \frac{e}{\kappa}}$ here to the condition $C\nu \log \frac{e}{\nu} < \kappa^2$ from Lemma 7.5 for delocalization on the sphere. If $\nu = \frac{c\kappa^2}{\log \frac{e}{\kappa}}$, then for small ν and κ, the two conditions are essentially equivalent up to constants, since

$$\nu \log \frac{e}{\nu} = \frac{c\kappa^2}{\log \frac{e}{\kappa}} \log\left(\frac{e \log \frac{e}{\kappa}}{c\kappa^2}\right) = 2c\kappa^2 \left(1 + \mathcal{O}\Big(\frac{C + \log\log \frac{e}{\kappa}}{\log \frac{e}{\kappa}}\Big)\right),$$

and conversely, if $C\nu \log \frac{e}{\nu} = \kappa^2$, then

$$\frac{\kappa^2}{\log \frac{e}{\kappa}} = \frac{2C\nu \log \frac{e}{\nu}}{\log\left(\frac{e^2}{C\nu \log \frac{e}{\nu}}\right)} = 2C\nu \left(1 + \mathcal{O}\Big(\frac{C + \log\log \frac{e}{\nu}}{\log \frac{e}{\nu}}\Big)\right).$$

Thus up to constants, the vector w chosen here is essentially just as delocalized as a typical random unit vector from the sphere.

(b) The choice $q = 4 \log \frac{e}{\kappa}$ was taken by (approximately) optimizing the condition $\nu^{1-2/q} C_q^{2/q} < \kappa^2$ over q. The approximation was applying Stirling in C_q before optimizing, and then assuming q large, and choosing convenient factors so that $q \geq 4$.

Proof of Theorem 6.2(ii). The condition for Theorem 9.8 is met by the weak local semicircle law in Section 9.1. Then choosing $P_{[2-\varepsilon_N,b]}$ as in Theorem 9.8, with probability $1-o(1)$, there are the projection matrix estimates (9.3.2), and the rank of $P_{[2-\varepsilon_N,b]}$ is $m_N = \frac{2}{3\pi}\varepsilon_N^{3/2}N(1+o(1))$. Then Corollary 9.11 implies for sufficiently large N, there is a $(\nu N, \kappa)$-delocalized vector v_N in the span of $\{\psi^{(N,j)} : \lambda^{(N,j)} \in [2-\varepsilon_N, b]\}$. By construction, $\|X_N v_N\|_2 \geq (2-\varepsilon_N)\sqrt{d}$. Since $\mathbb{E}\|X_N\| \leq (2+r_N)\sqrt{d}$ for some $r_N = o(1)$ by Theorem 6.1, then taking $\delta_N := 1 - \frac{2-\varepsilon_N}{2+r_N} = o(1)$ shows $(2-\varepsilon_N)\sqrt{d} = (1-\delta_N)(2+r_N)\sqrt{d} \geq (1-o(1))\mathbb{E}\|X_N\|$. $\square$

APPENDIX A

A.1. Proof of Lemma 9.2

This is the standard proof with Gaussian integration by parts and concentration (e.g. see [AGZ09]), with keeping track of the error terms. As in the usual semicircle law proof for Gaussian matrices, start with

$$\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1} = -z^{-1} + z^{-1}\frac{X_N}{\sqrt{d}}\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1},$$

then take expectations using Gaussian integration by parts $\mathbb{E}[X_{ij}f(X)] = \mathbb{E}\left[\partial_{X_{ij}}f(X)\right]$ and the relation $\partial_x Y^{-1} = -Y^{-1}(\partial_x Y)Y^{-1}$. Then

$$\text{(A.1.1)}\quad z\mathbb{E}\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1}_{xy} = \\ -\delta_{xy} - \frac{1}{d}\mathbb{E}\left[\sum_{\ell:\ell\sim x}\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1}_{\ell\ell}\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1}_{xy} + \Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1}_{\ell x}\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1}_{\ell y}\mathbb{1}_{\ell\neq x}\right].$$

The last term in (A.1.1) is bounded as

$$\begin{aligned}
&\frac{1}{d}\sum_{\ell:\ell\sim x}\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1}_{\ell x}\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1}_{\ell y} \\
&\quad\le \frac{1}{d}\left(\sum_{\ell:\ell\sim x}\left|\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1}_{\ell x}\right|^2\right)^{1/2}\left(\sum_{\ell:\ell\sim x}\left|\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1}_{\ell y}\right|^2\right)^{1/2} \\
&\quad\le \frac{1}{d}\left(\sum_{\ell=1}^{N}\left|\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1}_{\ell x}\right|^2\right)^{1/2}\left(\sum_{\ell=1}^{N}\left|\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1}_{\ell y}\right|^2\right)^{1/2} \\
&\quad= \frac{1}{d}\left(\Big|\Big\langle x\Big|\Big(\frac{X_N}{\sqrt{d}} - \overline{z}\Big)^{-1}\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1}\Big|x\Big\rangle\Big|\Big|\Big\langle y\Big|\Big(\frac{X_N}{\sqrt{d}} - \overline{z}\Big)^{-1}\Big(\frac{X_N}{\sqrt{d}} - z\Big)^{-1}\Big|y\Big\rangle\Big|\right)^{1/2} \\
&\quad\le \frac{1}{d|\,\mathfrak{Im}\,z|^2},
\end{aligned}$$

since $\|(X_N/\sqrt{d}-z)^{-1}\| \le |\mathfrak{Im}\, z|^{-1}$.

It remains to distribute the expectation over the product in the term $\mathbb{E}\left[\left(\frac{X_N}{\sqrt{d}}-z\right)^{-1}_{xy}\sum_{\ell:\ell\sim i}\left(\frac{X_N}{\sqrt{d}}-z\right)^{-1}_{\ell\ell}\right]$, which can be done via Gaussian concentration applied to $\sum_{\ell:\ell\sim x}\left(\frac{X_N}{\sqrt{d}}-z\right)^{-1}_{\ell\ell}$. Using the second resolvent identity and several instances of Cauchy-Schwarz, for symmetric matrices A and B, (or one can compute the gradient)

$$\begin{aligned}
&\left|\sum_{\ell:\ell\sim x}\langle\ell|(A-z)^{-1}|\ell\rangle-\langle\ell|(B-z)^{-1}|\ell\rangle\right| = \left|\sum_{\ell:\ell\sim x}\langle\ell|(A-z)^{-1}(B-A)R_B(z)|\ell\rangle\right| \\
&\quad\le \|A-B\|_F\left(\sum_{j,k=1}^{N}\left|\sum_{\ell:\ell\sim x}\langle\ell|R_A(z)|j\rangle\langle k|R_B(z)|\ell\rangle\right|^2\right)^{1/2} \\
&\quad= \|A-B\|_F\left(\sum_{j,k=1}^{N}\sum_{\ell:\ell\sim x}\langle\ell|R_A(z)|j\rangle\langle k|R_B(z)|\ell\rangle\sum_{m:m\sim x}\langle j|R_A(\overline{z})|m\rangle\langle m|R_B(\overline{z})|k\rangle\right)^{1/2} \\
&\quad= \|A-B\|_F\left(\sum_{\ell:\ell\sim x}\sum_{m:m\sim x}\langle\ell|R_A(z)R_A(\overline{z})|m\rangle\langle m|R_B(\overline{z})R_B(z)|\ell\rangle\right)^{1/2} \\
&\quad\le \|A-B\|_F\left(\sum_{\ell:\ell\sim x}\left(\sum_{m:m\sim x}|\langle\ell|R_A(z)R_A(\overline{z})|m\rangle|^2\right)^{1/2}\left(\sum_{m:m\sim x}|\langle\ell|R_B(\overline{z})R_B(z)|m\rangle|^2\right)^{1/2}\right)^{1/2} \\
&\quad\le \|A-B\|_F\left(\sum_{\ell:\ell\sim x}\langle\ell|(R_A(z)R_A(\overline{z}))^2|\ell\rangle^{1/2}\langle\ell|(R_B(z)R_B(\overline{z}))^2|\ell\rangle^{1/2}\right)^{1/2} \\
&\quad\le \|A-B\|_F\left(\sum_{\ell:\ell\sim x}\frac{1}{|\mathfrak{Im}\, z|^4}\right)^{1/2} \le \sqrt{2}\|A-B\|_{\mathbb{R}^{\triangledown}}\frac{\sqrt{d}}{|\mathfrak{Im}\, z|^2},
\end{aligned}$$

where $\|\cdot\|_{\mathbb{R}^{\triangledown}}$ is the Euclidean norm on the upper triangular elements that are nonzero in the sparsity pattern of the matrix. (These are the (i,j) such that $i\sim j$ and $i\ge j$.) Thus $\sum_{\ell:\ell\sim x}\left(\frac{X_N}{\sqrt{d}}-z\right)^{-1}_{\ell\ell}$ is $\sqrt{2}/|\operatorname{Im} z|^2$-Lipschitz, and its standard deviation is bounded by

$2/|\mathfrak{Im}\, z|^2$, so that using Cauchy–Schwarz,

$$\left|\mathbb{E}\left[\Big(\frac{X_N}{\sqrt{d}}-z\Big)^{-1}_{xy}\sum_{\ell:\ell\sim x}\Big(\frac{X_N}{\sqrt{d}}-z\Big)^{-1}_{\ell\ell}\right]-\mathbb{E}\Big(\frac{X_N}{\sqrt{d}}-z\Big)^{-1}_{xy}\mathbb{E}\sum_{\ell:\ell\sim x}\Big(\frac{X_N}{\sqrt{d}}-z\Big)^{-1}_{\ell\ell}\right|$$
$$\leq \frac{2}{|\mathfrak{Im}\, z|^2}\cdot\left[\mathbb{E}\left|\Big(\frac{X_N}{\sqrt{d}}-z\Big)^{-1}_{xy}\right|^2\right]^{1/2}\leq\frac{2}{|\mathfrak{Im}\, z|^3}.$$

Thus (A.1.1) becomes

$$z\mathbb{E}\Big(\frac{X_N}{\sqrt{d}}-z\Big)^{-1}_{xy}=-\delta_{xy}-\mathbb{E}\Big(\frac{X_N}{\sqrt{d}}-z\Big)^{-1}_{xy}\cdot\frac{1}{d}\sum_{\ell:\ell\sim i}\mathbb{E}\Big(\frac{X_N}{\sqrt{d}}-z\Big)^{-1}_{\ell\ell}+\Phi_{X_N,d,z,x,y} \tag{A.1.2}$$

with

$$|\Phi_{X_N,d,z,x,y}|\leq\frac{1}{d|\mathfrak{Im}\, z|^2}+\frac{2}{d|\mathfrak{Im}\, z|^3}. \tag{A.1.3}$$

A.2. Semicircle law computations

Let $m_{\mathrm{sc}}(z)=-\frac{z}{2}+\frac{\sqrt{z^2-4}}{2}$ the Stieltjes transform of the semicircle law, with branch cut taken in $(-2,2)$ and $\sqrt{z^2-4}\approx z$ for large z. With the standard square root branch, this is $m_{\mathrm{sc}}(z)=-\frac{z}{2}+\frac{\sqrt{z-2}\sqrt{z+2}}{2}$ at least for $z\notin\mathbb{R}_-$.

Lemma A.1 (Norm bounds)**.** *For $z\in\mathbb{H}$,*

$$|m_{\mathrm{sc}}(z)|<1. \tag{A.2.1}$$

For a bounded set $R\subset\mathbb{H}$, there is a constant $c>0$ (depending on R) so that for all $z\in R$,

$$1-|m_{\mathrm{sc}}(z)|^2\geq c\,\mathfrak{Im}\, z. \tag{A.2.2}$$

The above estimates are well-known and simply an application of the formula for $m_{\mathrm{sc}}(z)$. We provide proofs here for completeness since they are generally not written. There is no effort to obtain a good bound on the constant.

Proof. From the self-consistent equation, $|m_{\mathrm{sc}}(z)||z+m_{\mathrm{sc}}(z)|=1$. For $|z|$ large, $|z+m_{\mathrm{sc}}(z)|\approx|z|>1$ and then $|m_{\mathrm{sc}}(z)|<1$. Since $|m_{\mathrm{sc}}(z)|$ is continuous, to show that

$|m_{\mathrm{sc}}(z)| < 1$ for all z, it is enough to show that for any $z \in \mathbb{H}$, $|m_{\mathrm{sc}}(z)| \neq |z + m_{\mathrm{sc}}(z)|$. Let $w := \frac{1}{2}\sqrt{z-2}\sqrt{z+2}$, so $|m_{\mathrm{sc}}(z)| = \frac{1}{2}|z-w|$ and $|z+m_{\mathrm{sc}}(z)| = \frac{1}{2}|z+w|$. Then $|z-w| = |z+w|$ iff $\mathfrak{Re}\,\overline{z}w = 0$. However, $\mathfrak{Re}\,\overline{z}w > 0$ for any $z \in \mathbb{H}$:

For $z = E + i\eta$,

$$\tag{A.2.3} \begin{aligned} 2\,\mathfrak{Re}(\overline{z}\sqrt{z-2}\sqrt{z+2}) = \\ & E\sqrt{((E-2)^2+\eta^2)^{1/2}+(E-2)}\sqrt{((E+2)^2+\eta^2)^{1/2}+(E+2)} - \\ & - E\sqrt{((E-2)^2+\eta^2)^{1/2}-(E-2)}\sqrt{((E+2)^2+\eta^2)^{1/2}-(E+2)} + \\ & + \eta\sqrt{((E-2)^2+\eta^2)^{1/2}+(E-2)}\sqrt{((E+2)^2+\eta^2)^{1/2}-(E+2)} + \\ & \qquad + \eta\sqrt{((E+2)^2+\eta^2)^{1/2}+(E+2)}\sqrt{((E-2)^2+\eta^2)^{1/2}-(E-2)} \end{aligned}$$

This is unchanged under $E \mapsto -E$ so it suffices to consider $E \geq 0$.

The last two terms (starting with η) are always positive, so we just check the first term is larger than the second. This is clear for $E \geq 2$. For $0 \leq E \leq 2$, then $E - 2 < 0$. Letting $\xi = E - 2$ and $\omega = E + 2$, the first two terms then are

$$\tag{A.2.4} E\sqrt{(\xi^2+\eta^2)^{1/2}-|\xi|}\sqrt{(\omega^2+\eta^2)^{1/2}+\omega} - E\sqrt{(\xi^2+\eta^2)^{1/2}+|\xi|}\sqrt{(\omega^2+\eta^2)^{1/2}-\omega}.$$

Note $(\xi^2+\eta^2)^{1/2}-|\xi| > (\omega^2+\eta^2)^{1/2}-\omega$ since $(x^2+\eta^2)^{1/2}-x$ is (strictly, for $\eta > 0$) decreasing in x. Similarly, $(\omega^2+\eta^2)^{1/2}+\omega > (\xi^2+\eta^2)^{1/2}+|\xi|$ since $(x^2+\eta^2)^{1/2}+x$ is (strictly, for $\eta > 0$) increasing in x. (e.g. check by derivatives). Thus, $\mathfrak{Re}(\overline{z}w) > 0$ so $|m_{\mathrm{sc}}(z)| \neq |z + m_{\mathrm{sc}}(z)|$ and $|m_{\mathrm{sc}}(z)| < 1$ for all $z \in \mathbb{H}$.

For (A.2.2), since $|z+w|^2 - |z-w|^2 = 4\,\mathfrak{Re}\,\overline{z}w$, then using $|z-w||z+w| = 4$,

$$\tag{A.2.5} 1 - |m_{\mathrm{sc}}(z)|^2 = 1 - \frac{1}{4}|z-w|^2 = \frac{\mathfrak{Re}\,\overline{z}w}{|z+w|+|z-w|}|z-w|.$$

Since $|z|$ is bounded, $1 < |z+w| \leq C$ and thus $\frac{1}{C} \leq |z-w| < 1$. Then

$$\tag{A.2.6} 1 - |m_{\mathrm{sc}}(z)|^2 = 1 - \frac{1}{4}|z-w|^2 \geq \frac{1}{C(C+1)}\,\mathfrak{Re}(\overline{z}w).$$

By casework, there will be $c > 0$ so that

$$\mathfrak{Re}(\overline{z}\sqrt{z-2}\sqrt{z+2}) \geq c\,\mathfrak{Im}\,z. \tag{A.2.7}$$

- $E \geq 2$: Since for $0 \leq \alpha, \beta \leq 1$,

$$\sqrt{1+\alpha}\sqrt{1+\beta} - \sqrt{1-\alpha}\sqrt{1-\beta} = \frac{2(\alpha+\beta)}{\sqrt{1+\alpha}\sqrt{1+\beta}+\sqrt{1-\alpha}\sqrt{1-\beta}} \geq C(\alpha+\beta), \tag{A.2.8}$$

  then the first two terms of (A.2.3) are bounded below by

$$CE(\omega^2+\eta^2)^{1/4}(\xi^2+\eta^2)^{1/4}\left[\frac{\xi}{(\xi^2+\eta^2)^{1/2}} + \frac{\omega}{(\omega^2+\eta^2)^{1/2}}\right] \geq CE(E-2) + C'E\eta^{1/2},$$

  where C' depends on the maximum possible η in the region R.

- $1 \leq E \leq 2$: Again with just the first two terms of (A.2.3). Note for $0 \leq x \leq b$ and $C_b := \frac{\sqrt{b+1}-1}{b}$,

$$1 + C_b x \leq \sqrt{1+x} \leq 1 + \frac{x}{2}. \tag{A.2.9}$$

  Then

$$\begin{aligned} &E\sqrt{(\xi^2+\eta^2)^{1/2}-|\xi|}\sqrt{(\omega^2+\eta^2)^{1/2}+\omega} - E\sqrt{(\xi^2+\eta^2)^{1/2}+|\xi|}\sqrt{(\omega^2+\eta^2)^{1/2}-\omega} \\ &\qquad \geq E\sqrt{(\xi^2+\eta^2)^{1/2}-|\xi|}\sqrt{2\omega} - E\sqrt{\eta+2|\xi|}\frac{\eta}{\sqrt{2\omega}}. \end{aligned} \tag{A.2.10}$$

  Since $\mathfrak{Re}(\overline{z}w) > 0$ in all of $\mathbb{H}$, it suffices to consider small η, since $\mathfrak{Re}(\overline{z}w) \geq c > 0$ on any bounded region with η bounded away from zero (by compactness). If $\eta \leq 2|\xi|$, then by (A.2.9), along with $3 \leq \omega \leq 4$ and $0 \leq |\xi| \leq 1$,

$$\sqrt{(\xi^2+\eta^2)^{1/2}-|\xi|} \geq \frac{\eta}{|\xi|^{1/2}}\left(\frac{\sqrt{5}-1}{4}\right)^{1/2},$$

  so for sufficiently small η,

$$2\,\mathfrak{Re}(\overline{z}\sqrt{z-2}\sqrt{z+2}) \geq \frac{E\eta}{\sqrt{\omega|\xi|}}\left[3\sqrt{2}\sqrt{C_4} - |\xi| - \frac{1}{\sqrt{2}}\sqrt{\eta|\xi|}\right] \geq c\eta. \tag{A.2.11}$$

If $\eta \geq 2|\xi|$, then

$$\sqrt{(\xi^2+\eta^2)^{1/2}-|\xi|} \geq \sqrt{\eta - \frac{\eta}{2}} = \sqrt{\frac{\eta}{2}}.$$

- $0 \leq E \leq 1$: The fourth term of (A.2.3) is at least $2\sqrt{2}\cdot\eta$. (Usually we can't really use the 3rd and 4th terms since they can be at least order $\eta^{3/2} \ll \eta$ for small η, but for E away from ± 2 the 4th term is order η.)

□

Lemma A.2 (Semicircle integration)**.** *Let $a_N^\star$ be $2-\varepsilon_N+\gamma_N$ or $2-\varepsilon_N-\gamma_N$ where $\gamma_N = o(\varepsilon_N)$, and suppose $\delta_N \ll \varepsilon_N^2$. Then*

$$\frac{1}{\pi}\,\mathfrak{Im}\int_{a_N^\star}^{b} m_{\mathrm{sc}}(\lambda+i\delta_N)\,d\lambda = \frac{2}{3\pi}\varepsilon_N^{3/2}(1+o(1)). \tag{A.2.12}$$

Proof. Compute

$$\begin{aligned}\mathfrak{Im}\int_{a_N^\star}^{b} -\frac{\lambda+i\delta_N}{2} + \frac{\sqrt{\lambda+i\delta_N-2}\sqrt{\lambda+i\delta_N+2}}{2}\,d\lambda \\ = \mathcal{O}(\varepsilon_N\delta_N) + \frac{1}{2}\,\mathfrak{Im}\int_{a_N^\star}^{b}\sqrt{\lambda+i\delta_N-2}\sqrt{\lambda+i\delta_N+2}.\end{aligned} \tag{A.2.13}$$

Computing antiderivatives,

$$\begin{aligned}&\mathfrak{Im}\int_{a_N^\star}^{b}\sqrt{\lambda+i\delta_N-2}\sqrt{\lambda+i\delta_N+2}\,d\lambda \\ &\quad= \mathfrak{Im}\Bigg[\frac{1}{2}(\lambda+i\delta_N)\sqrt{(\lambda+i\delta_N)-2}\sqrt{(\lambda+i\delta_N)+2} - \\ &\qquad - 2\log\left(\lambda+i\delta_N+\sqrt{(\lambda+i\delta_N)-2}\sqrt{(\lambda+i\delta_N)+2}\right)\Bigg]_{a_N^\star}^{b}\end{aligned} \tag{A.2.14}$$

For the terms with b, Taylor expansion in δ_N yields,

$$\begin{aligned}\mathfrak{Im}\left[\frac{1}{2}b\sqrt{b^2-4} - 2\log[b+\sqrt{b^2-4}] - i\sqrt{b^2-4}\cdot\delta_N - \frac{b}{2\sqrt{b^2-4}}\delta_N^2 + \mathcal{O}(\delta_N^3)\right] \\ = \mathcal{O}(\delta_N) = o(\varepsilon_N^2).\end{aligned}$$

For the terms with $a_N^\star$, first a useful computation. For notational convenience, let ε be $\varepsilon_N + \gamma_N$ or $\varepsilon_N - \gamma_N$. Then

$$\begin{aligned}
&\sqrt{(2-\varepsilon+i\delta_N)-2}\sqrt{(2-\varepsilon+i\delta_N)+2} \\
&\qquad = \sqrt{\varepsilon}\sqrt{4-\varepsilon}\sqrt{-1+i\frac{\delta_N}{\varepsilon}}\sqrt{1+i\frac{\delta_N}{4-\varepsilon}} \\
&\qquad = 2\sqrt{\varepsilon}\left(i+\frac{\delta_N}{\varepsilon}+\mathcal{O}(\delta_N^2\varepsilon^{-2})\right)\left(1-\frac{1}{8}\varepsilon+\mathcal{O}(\varepsilon^2)\right)(1+i\mathcal{O}(\delta_N)) \\
&\qquad = 2i\sqrt{\varepsilon}-\frac{i}{4}\varepsilon^{3/2}+\mathcal{O}(\delta_N\varepsilon^{-1/2}+\varepsilon^{5/2})
\end{aligned}$$

since for $\alpha > 0$, $\sqrt{-1+\alpha i} = i + \frac{\alpha}{2} + \mathcal{O}(\alpha^2)$, e.g. by Taylor expanding several times $\sqrt{-1+\alpha i} = \sqrt{1+\alpha^2}\exp(i(\pi - \tan^{-1}\alpha))$. Using that $\mathcal{O}(\delta_N\varepsilon^{-1/2}) = \mathcal{O}(\delta_N\varepsilon_N^{-1/2}) = o(\varepsilon_N^{3/2})$, the non-logarithm term involving $a_N^\star$ in (A.2.14) is then

$$\begin{aligned}
-\mathfrak{Im}\,\frac{1}{2}(\lambda+i\delta_N)\sqrt{(\lambda+i\delta_N)-2}\sqrt{(\lambda+i\delta_N)+2}\bigg|_{\lambda=a_N^\star} &= -\frac{1}{2}(2-\varepsilon)\left[2\varepsilon^{1/2}-\frac{1}{4}\varepsilon^{3/2}+o(\varepsilon_N^{3/2})\right] \\
&= -2\varepsilon^{1/2}+\frac{5}{4}\varepsilon^{3/2}+o(\varepsilon_N^{3/2}).
\end{aligned}$$

For the logarithm term,

$$\begin{aligned}
&\log(2-\varepsilon+i\delta+\sqrt{-\varepsilon+i\delta}\sqrt{4-\varepsilon+i\delta}) \\
&\qquad = \log((2-\varepsilon+i\delta)+\log\left(1+\frac{1}{2-\varepsilon+i\delta}(2i\sqrt{\varepsilon}-\frac{i}{4}\varepsilon^{3/2}+o(\varepsilon_N^{3/2}))\right) \\
&\qquad = \log 2-\frac{\varepsilon}{2}+i\sqrt{\varepsilon}+\frac{i}{24}\varepsilon^{3/2}+\mathcal{O}(\delta)+o(\varepsilon_N^{3/2}).
\end{aligned}$$

The end logarithm term is then

$$\begin{aligned}
&\mathfrak{Im}\,2\log\left(\lambda+i\delta+\sqrt{(\lambda+i\delta_N)-2}\sqrt{(\lambda+i\delta)+2}\right) \\
&\qquad = 2\left(\sqrt{\varepsilon}+\frac{1}{24}\varepsilon^{3/2}\right)+o(\varepsilon^{3/2}) = 2\varepsilon^{1/2}+\frac{1}{12}\varepsilon^{3/2}+o(\varepsilon^{3/2}).
\end{aligned}$$

Thus in total, (A.2.13) is

$$\frac{1}{\pi} \cdot \frac{1}{2} \left(\frac{5}{4} + \frac{1}{12} \right) \varepsilon^{3/2} + o(\varepsilon_N^{3/2}) = \frac{2}{3\pi} \varepsilon^{3/2} + o(\varepsilon_N^{3/2}).$$

Since $\varepsilon = \varepsilon_N(1 \pm o(1))$, this is $\frac{2}{3\pi}\varepsilon_N^{3/2}(1 + o(1))$ as desired. □

This agrees to leading order with what we expect from the semicircle law,

$$\frac{1}{2\pi} \int_{2-\varepsilon}^{2} \sqrt{4 - x^2}\, dx \approx \frac{2}{3\pi} \varepsilon^{3/2},$$

e.g. by using that $m_{\mathrm{sc}}(z)$ is the Stieltjes transform of the semicircle density and applying Fubini and Taylor expanding.

A.3. Stability without quantitative stability

In Section 9.1, we obtained a quantitative bound on the rate of convergence of $\mathbb{E}G_N(z; x, y)$ to $\delta_{xy} m_{\mathrm{sc}}(z)$. This makes it easy to choose a sequence $\delta_N \to 0$ such that

$$\max_{\lambda \in [0,b]} \max_{x \in [N]} |\mathbb{E}G_N(\lambda + i\delta_N; x, x) - m_{\mathrm{sc}}(\lambda + i\delta_N)| \to 0,$$

since the quantitative bound doesn't depend on λ or x. (Remark 9.3.1.) If we did not have such a quantitative bound, but only knew that for any fixed $z = \lambda + i\delta$,

$$\max_{x \in [N]} |\mathbb{E}G_N(\lambda + i\delta; x, x) - m_{\mathrm{sc}}(\lambda + i\delta)| \to 0$$

as $N \to \infty$, with no knowledge of the convergence rate or dependence on δ, we can still find a sequence $\delta_N \to 0$ with the desired property. We will use the fact that for functions on a compact metric space, pointwise convergence plus equicontinuity implies uniform convergence.

Lemma A.3. *Suppose* $\max_{x \in [N]} |\mathbb{E}G_N(\lambda + i\delta; x, x) - m_{\mathrm{sc}}(\lambda + i\delta)| \to 0$ *as* $N \to \infty$, *for any fixed* $z = \lambda + i\delta$ *with* $\delta > 0$. *Then there is a positive sequence* $\delta_N \to 0$ *so that*

(A.3.1) $$\mathcal{E}(N, \delta_N) := \sup_{\delta_N \le \gamma \le 1} \sup_{\lambda \in [0,b]} \max_{x \in [N]} |\mathbb{E}G_N(\lambda + i\gamma; x, x) - m_{\mathrm{sc}}(\lambda + i\gamma)| \to 0,$$

as $N \to \infty$. Additionally, any sequence (α_N) satisfying $\alpha_N \geq \delta_N$ will also satisfy $\mathcal{E}(N, \alpha_N) \to 0$.

Proof. The "additionally" statement is immediate from the $\sup_{\delta_N \leq \gamma \leq 1}$ and $\mathcal{E}(N, \delta_N) \to 0$. For (A.3.1), define $\widetilde{g}_{N,x}(z) = \mathbb{E} G_N(z; x, x) - m_{\mathrm{sc}}(z)$ and let R_δ be the compact region $[0, b] \times i[\delta, 1]$ in $\mathbb{C}$. Then $|\widetilde{g}_{N,x}(z)|$ is Lipschitz continuous for $z \in R_\delta$ with some Lipschitz constant $L(\delta)$, since using the first resolvent identity,

$$\begin{aligned}
\big| |\mathbb{E} G_N(z; x, x) - m_{\mathrm{sc}}(z)| - |\mathbb{E} G_N(w; x, x) - m_{\mathrm{sc}}(w)| \big| & \\
&\leq |\mathbb{E} G_N(z; x, x) - \mathbb{E} G_N(w; x, x)| + |m_{\mathrm{sc}}(z) - m_{\mathrm{sc}}(w)| \\
&\leq |z - w| \mathbb{E}(R_N(z) R_N(w))_{xx} + |z - w| \sup_{\delta \leq \mathfrak{Im}\, z \leq 1} \left| \frac{d}{dz} m_{\mathrm{sc}}(z) \right| \\
&\leq |z - w| \frac{1}{\mathfrak{Im}\, z \, \mathfrak{Im}\, w} + |z - w| \frac{1}{2} \left(1 + \frac{1 + |b|}{\sqrt{2\delta}} \right).
\end{aligned}$$

Since taking a maximum doesn't change the Lipschitz constant, $f_N(z) := \max_{x \in [N]} |\widetilde{g}_{N,x}(z)|$ also has Lipschitz constant $L(\delta)$ in the region R_δ. By assumption $f_N(z) \to 0$ pointwise in z as $N \to \infty$. Because there is the uniform Lipschitz constant $L(\delta)$, $\{f_N\}_N$ is equicontinuous and so the pointwise convergence turns into uniform convergence,

$$\lim_{N \to \infty} \|f_N\|_{C(R_\delta)} = \lim_{N \to \infty} \sup_{z \in R_\delta} \max_{x \in [N]} |\mathbb{E} G_N(z; x, x) - m_{\mathrm{sc}}(z)| = 0.$$

Now for any $\delta > 0$, define

$$D(N, \delta) := \sup_{n \geq N} \mathcal{E}(N, \delta),$$

which is decreasing in N (and finite, $< \frac{2}{\delta}$, by resolvent bound). By the above, then $\lim_{N \to \infty} D(N, \delta) = \limsup_{N \to \infty} \|f_N\|_{C(R_\delta)} = 0$. Choose δ_N as follows: First take $\delta_1 = 1$ and a parameter (which tracks the convergence rate) $m = 2$. Take $\delta_N = \delta_{N-1}$ until N is large enough so that $D(N, \frac{1}{m}) < \frac{1}{m}$. In that case take $\delta_N = \frac{1}{m}$ and increment m, then repeat. Since $D(N, \delta)$ is decreasing in N, this method guarantees once $D(N, \delta_N) < \frac{1}{m}$, that this inequality holds for all subsequent N as well.

Then $\delta_N \to 0$ since they eventually will be smaller than any $\frac{1}{m}$. Also, $D(N, \delta_N) \to 0$ since it will eventually be smaller than any $\frac{1}{m}$. $\square$

Bibliography

[AEK19] O. Ajanki, L. Erdős, and T. Krüger, *Quadratic Vector Equations On Complex Upper Half-Plane*, vol. 261, Memoirs of the American Mathematical Society, 2019.

[AGZ09] G. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, Cambridge University Press, 2009.

[AL15] N. Anantharaman and E. Le Masson, *Quantum ergodicity on large regular graphs*, Duke Math. J. 164 (2015), 723–765.

[Ana17] N. Anantharaman, *Quantum ergodicity on regular graphs*, Comm. Math. Phys. 353 (2017), 633–690.

[Ana18] N. Anantharaman, *Delocalization of Schrödinger eigenfunctions*, in Proceedings ICM 2018, vol. 1, World Sci. Publ., 2018, 341–375.

[AS19] N. Anantharaman and M. Sabri, *Quantum ergodicity on graphs: From spectral to spatial delocalization*, Ann. Math. 189 (2019), 753–835.

[AT07] R. J. Adler and J. E. Taylor, *Random Fields and Geometry*, Springer Monographs in Mathematics, Springer, 2007.

[BBvH21] A. S. Bandeira, M. T. Boedihardjo, and R. van Handel, *Matrix concentration inequalities and free probability*. Preprint arXiv:2108.06312 (2021).

[BD96] A. Bouzouina and S. De Biévre, *Equipartition of the eigenfunctions of quantized ergodic maps on the torus*, Comm. Math. Phys. 178 (1996), 83–105.

[Ber77] M. V. Berry, *Regular and irregular semiclassical wavefunctions*, J. Phys. A 10 (1977), 2083–2091.

[BGGS97] E. B. Bogomolny, B. Georgeot, M. J. Giannoni, and C. Schmit, *Arithmetical chaos*, Phys. Rep. 291 (1997), 219–324.

[BGK16] F. Benaych-Georges and A. Knowles, *Lectures on the local semicircle law for Wigner matrices*. Available at arXiv:1601.04055 (2016).

[BGP14] F. Benaych-Georges and S. Péché, *Largest eigenvalues and eigenvectors of band or sparse random matrices*, Electron. Commun. Probab. 19 (2014), 9 pp.

[BGS84] O. Bohigas, M. J. Giannoni, and C. Schmidt, *Characterization of chaotic quantum spectra and universality of level fluctuation laws*, Phys. Rev. Lett. 52 (1984), 1–4.

[BK10] G. Berkolaiko and P. Kuchment, *Introduction to Quantum Graphs*, Mathematical Surveys and Monographs, American Mathematical Society, 2010.

[BKS07] G. Berkolaiko, J. P. Keating, and U. Smilansky, *Quantum ergodicity for graphs related to interval maps*, Comm. Math. Phys. 273 (2007), 137–159.

[BLM13] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, 2013.

[Bou18] P. Bourgade, *Random band matrices*, in Proceedings ICM 2018, vol. 4, 2018.

[BSW02] G. Berkolaiko, H. Schanz, and R. Whitney, *Leading off-diagonal correction to the form factor of large graphs*, Phys. Rev. Lett. 82 (2002), 104101.

[BSW03] G. Berkolaiko, H. Schanz, and R. Whitney, *Form factor for a family of quantum graphs: An expansion to third order*, J. Phys. A 36 (2003), 8373–8392.

[BvH16] A. Bandeira and R. van Handel, *Sharp nonasymptotic bounds on the norm of random matrices with independent entries*, Ann. Probab. 44 (2016), 2479–2506.

[CG18] S. Chatterjee and J. Galkowski, *Arbitrarily small perturbations of Dirichlet Laplacians are quantum unique ergodic*, J. Spectr. Theory 8 (2018), 909–947.

[CG20] Y. Canzani and J. Galkowski, *Weyl remainders: an application of geodesic beams*. Preprint arXiv:2010.03969 (2020).

[CM06] N. Chernov and R. Markarian, *Chaotic Billiards*, vol. 127 of Mathematical Surveys and Monographs, American Mathematical Society, 2006.

[CM08] S. Chatterjee and E. Meckes, *Multivariate normal approximation using exchangeable pairs*, ALEA 4 (2008), 257–283.

[DF84] P. Diaconis and D. Freedman, *Asymptotics of graphical projection pursuit*, Ann. Stat. 12 (1984), 793–815.

[DG75] J. J. Duistermaat and V. W. Guillemin, *The spectrum of positive elliptic operators and periodic bicharacteristics*, Invent. Math. 29 (1975), 39–79.

[dV85] Y. C. de Verdière, *Ergodicité et fonctions propres du laplacien*, Comm. Math. Phys. 102 (1985), 497–502.

[EKYY13a] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Spectral statistics of Erdős–Rényi graphs I: local semicircle law*, Ann. Probab. 41 (2013), 2279–2375.

[EKYY13b] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *The local semicircle law for a general class of random matrices*, Electron. J. Probab. 18 (2013), 58 pp.

[ESY09a] L. Erdős, B. Schlein, and H.-T. Yau, *Local semicircle law and complete delocalization for Wigner random matrices*, Comm. Math. Phys. 287 (2009), 641–655.

[ESY09b] L. Erdős, B. Schlein, and H.-T. Yau, *Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices*, Ann. Probab. 37 (2009), 815–852.

[EY17] L. Erdős and H.-T. Yau, *A Dynamical Approach to Random Matrix Theory*, vol. 28 of Courant Lecture Notes, AMS and Courant Institute, 2017.

[FM91] Y. V. Fyodorov and A. D. Mirlin, *Scaling properties of localization in random band matrices: a sigma-model approach*, Phys. Rev. Lett. 67 (1991).

[GA04] S. Gnutzmann and A. Altland, *Universal spectral statistics in quantum graphs*, Phys. Rev. Lett. 93 (2004), 194101.

[GA05] S. Gnutzmann and A. Altland, *Spectral correlations of individual quantum graphs*, Phys. Rev. E 72 (2005), 056215.

[GKP08] S. Gnutzmann, J. P. Keating, and F. Piotet, *Quantum ergodicity on graphs*, Phys. Rev. Lett. 101 (2008), 264102.

[GKP10] S. Gnutzmann, J. Keating, and F. Piotet, *Eigenfunction statistics on quantum graphs*, Ann. Phys. 325 (2010), 2595–2640.

[GLSW06] Y. Gordon, A. E. Litvak, S. Schütt, and E. Werner, *On the minimum of several random variables*, Proc. Am. Math. Soc. 134 (2006), 3665–3675.

[GO13] B. Gutkin and V. A. Osipov, *Clustering of periodic orbits in chaotic systems*, Nonlinearity 26 (2013), 177–200.

[GS18] Y. Guo and M. Surmacs, *Miscellaneous digraph classes*, in Classes of Directed Graphs (J. Bang-Jensen and G. Gutin, eds.), Springer Monographs in Mathematics, Springer, 2018, ch. 11, 517–574.

[HFS07] J. W. Helton, R. R. Far, and R. Speicher, *Operator-valued semicircular elements: solving a quadratic matrix equation with positivity constraints*, Int. Math. Res. Not. 2007 (2007), 15 pp.

[HGK18] F. Haake, S. Gnutzmann, and M. Kuś, *Quantum Signatures of Chaos*, Springer Series in Synergetics, Springer, 4 ed., 2018.

[HH20] J. Harrison and T. Hudgins, *Complete dynamical evaluation of the characteristic polynomial of binary quantum graphs*. Preprint arXiv:2011.05213 (2020).

[HR16] H. Hezari and G. Riviére, l^p *norms, nodal sets, and quantum ergodicity*, Adv. Math. 290 (2016), 938–966.

[HT05] U. Haagerup and S. Thorbjørnsen, *A new application of random matrices:* $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ *is not a group*, Ann. Math. 162 (2005), 711–775.

[Hub85] P. Huber, *Projection pursuit*, Ann. Stat. 13 (1985), 435–525.

[Ivr80] V. Ivriĭ, *The second term of the spectral asymptotics for a laplace beltrami operator on manifolds with boundary*, Funktsional. Anal. i Prilozhen. 14 (1980), 25–34.

[KP81] S. Kakutani and K. Petersen, *The speed of convergence in the ergodic theorem*, Mon. Hefte. Math. 91 (1981), 11–18.

[KR00] P. Kurlberg and Z. Rudnick, *Hecke theory and distribution for the quantization of linear maps of the torus*, Duke Math. J. 103 (2000), 47–77.

[KR01] P. Kurlberg and Z. Rudnick, *On quantum ergodicity for linear maps of the torus*, Comm. Math. Phys. 222 (2001), 201–227.

[KS97] T. Kottos and U. Smilansky, *Quantum chaos on graphs*, Phys. Rev. Lett. 79 (1997), 4794–4797.

[KS99] T. Kottos and U. Smilansky, *Periodic orbit theory and spectral statistics for quantum graphs*, Ann. Phys. 274 (1999), 76–124.

[Ler05] P. Leroux, *Coassociative grammar, periodic orbits, and quantum random walk over z*, Int. J. Math. Math. Sci 24 (2005), 3979–3996.

[LS94] W. Luo and P. Sarnak, *Number variance for arithmetic hyperbolic surfaces*, Comm. Math. Phys. 161 (1994), 419–432.

[LT91] M. Ledoux and M. Talagrand, *Probability in Banach Spaces*, vol. 3 of A Series of Modern Surveys in Mathematics, Springer-Verlag, 2002 reprint ed., 1991.

[Mañ87] R. Mañé, *Ergodic Theory and Differentiable Dynamics*, vol. 8 of Ergebnisse der Mathematik und ihrer Grenzgebiete (3), Springer-Verlag, Berlin, 1987. Translated from the Portuguese by Silvio Levy.

[Map13] K. Maples, *Quantum unique ergodicity for random bases of spectral projections*, Math. Res. Lett. 20 (2013), 1115–1124.

[Mec09] E. Meckes, *Quantitative asymptotics of graphical projection pursuit*, Electron. Commun. Probab. 14 (2009), 176–185.

[Meh04] M. L. Mehta, *Random Matrices*, Elsevier Science & Technology, 2004.

[MO05] J. Marklof and S. O'Keefe, *Weyl's law and quantum ergodicity for maps with divided phase space*, Nonlinearity 18 (2005), 277–304.

[Mon94] H. L. Montgomery, *Ten Lectures on the Interface between Analytic Number Theory and Harmonic Analysis*, vol. 84 of CBMS Regional Conference Series in Mathematics, American Mathematical Society, 1994.

[OVW16] S. O'Rourke, V. Vu, and K. Wang, *Eigenvectors of random matrices: A survey*, J. Comb. Theory Ser. A. 144 (2016), 361–442.

[PZK01] P. Pakoński, K. Życzkowski, and M. Kuś, *Classical 1D maps, quantum graphs and ensembles of unitary matrices*, J. Phys. A 34 (2001), 9303–9317.

[RS94] Z. Rudnick and P. Sarnak, *The behaviour of eigenstates of arithmetic hyperbolic manifolds*, Comm. Math. Phys. 161 (1994), 195–213.

[RV13] M. Rudelson and R. Vershynin, *Hanson–Wright inequality and sub-Gaussian concentration*, Electron. Commun. Probab. 18 (2013), 9 pp.

[RV16] M. Rudelson and R. Vershynin, *No-gaps delocalization for general random matrices*, Geom. Funct. Anal. 26 (2016), 1716–1776.

[S21] L. Shou, *Pointwise Weyl law for graphs from quantized interval maps*. Preprint arXiv:2110.15301 (2021).

[Saf88] Y. G. Safarov, *Asymptotics of the spectral function of a positive elliptic operator without a nontrapping condition*, (Russian), Funktsional. Anal. i Prilozhen 22 (1988), 53–65.

[Sch09] J. Schenker, *Eigenvector localization for random band matrices with power law band width*, Comm. Math. Phys. 290 (2009), 1065–1097.

[Sel91] A. Selberg, *Alte Selberg Collected Papers Volume II*, Springer-Verlag, 1991.

[Shn74] A. Shnirelman, *Ergodic properties of eigenfunctions*, Uspehi Mat. Nauk 29 (1974), 181–182.

[SZ02] C. D. Sogge and S. Zelditch, *Riemannian manifolds with maximal eigenfunction growth*, Duke Math. J. 114 (2002), 387–437.

[Tan00] G. Tanner, *Spectral statistics for unitary transfer matrices of binary graphs*, J. Phys. A 33 (2000), 3567–3585.

[Tan01] G. Tanner, *Unitary-stochastic matrix ensembles and spectral statistics*, J. Phys. A 34 (2001), 8485–8500.

[TV11] T. Tao and V. Vu, *Random matrices: Universal properties of eigenvectors*, Random Matrices: Theory Appl. 1 (2011), 1150001.

[Vaa85] J. D. Vaaler, *Some extremal functions in Fourier analysis*, Bull. Amer. Math. Soc. 12 (1985), 183–216.

[Van97] J. M. VanderKam, l^∞ *norms and quantum ergodicity on the sphere*, Int. Math. Res 7 (1997), 329–347.

[Ver18] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge University Press, 2018.

[vH16] R. van Handel, *Probability in High Dimension, APC 550 Lecture Notes*. `https://web.math.princeton.edu/~rvan/APC550.pdf`, 2016.

[vH17] R. van Handel, *Structured random matrices*, in Convexity and Concentration (E. Carlen et al., eds.), vol. 161 of IMA, Springer, 2017, 107–165.

[Wai19] M. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press, 2019.

[Zel87] S. Zelditch, *Uniform distribution of eigenfunctions on compact hyperbolic surfaces*, Duke Math. J. 55 (1987), 919–941.

[Zel97] S. Zelditch, *Index and dynamics of quantized contact transformations*, Ann. Inst. Fourier (Grenoble) 47 (1997), 305–363.

[Zel14] S. Zelditch, *Quantum ergodicity of random orthonormal bases of spaces of high dimension*, Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 372 (2014), 20120511.

[ZSKS03] K. Życzkowski, W. Słomczński, M. Kuś, and H. J. Sommers, *Random unistochastic matrices*, J. Phys. A 36 (2003), 3425–3450.

[Zwo12] M. Zworski, *Semiclassical Analysis*, vol. 138 of Graduate Studies in Mathematics, American Mathematical Society, 2012.