ON EIGENVECTORS OF RANDOM BAND MATRICES WITH LARGE BAND

STEFAN STEINERBERGER

Abstract. We study random, symmetric $N \times N$ band matrices with a band of size $W$ and Bernoulli random variables as entries. This interpolates between nearest neighbour interaction $W = 1$ and Wigner matrices $W = N$. Eigenvectors are known to be localized for $W \ll N^{1/8}$, delocalized for $W \gg N^{4/5}$ and it is conjectured that the transition for the bulk occurs at $W \sim N^{1/2}$. Eigenvalues in the spectral edge change their behavior at $W \sim N^{5/6}$ but nothing is known about the associated eigenvectors. We show that up to $W \ll N^{5/7}$ any random matrix has with large probability some eigenvectors in the spectral edge, which either exhibit mass concentration or interact strongly on a small scale.

1. Introduction

1.1. Introduction. Random band matrices represent quantum systems on a large graph with random quantum transition amplitudes effective up to distances of order $W$. The typical length scale of an eigenvector is denoted by $\ell$, where $\ell \sim 1$ corresponds to highly localized and $\ell \sim N$ completely delocalized eigenvectors. Based on nonrigorous supersymmetric calculations [4], it is expected that $\ell \sim W^2$. This implies that complete delocalization happens at scale $W \sim N^{1/2}$ for the bulk of the eigenvectors. It is of great interest to understand the nature of this transition. The currently best result from above is due to Erdős, Knowles, Yau & Yin [2], who show that delocalization in the bulk does occur if $W \gg N^{4/5}$ (improving an earlier result by Erdős & Knowles [1]). At the natural endpoint $W = N$, Erdős, Schlein & Yau [3] prove that with high probability no eigenvector of a Wigner matrix is localized, which is to be expected because there is no distinguished interaction anymore. The same result for random matrices with independent entries was very recently proven by Rudelson & Vershynin [5]. Lower bounds are less well understood: the only available result is due to Schenker [6] who proves that localization occurs for $W \ll N^{1/8}$ in a particular model. As for the spectral edge, Sodin [7] has identified the threshold $W \sim N^{5/6}$ at which the probability distribution of eigenvalues in the edge undergoes a transition (his results are much more precise than that). However, nothing except localization below $W \ll N^{1/8}$ and delocalization at $W = N$ is known about the structure of eigenvectors with eigenvalues in the spectral edge – and in both these cases, the behavior of eigenvectors from the spectral edge is the very same as for all other eigenvalues.

1.2. Our model. We work with a particular band matrix model introduced by Sodin [7]: consider random, symmetric $N \times N$ matrices. The rows and columns are labelled by elements of $\mathbb{Z}_N := \mathbb{Z}/N\mathbb{Z}$ and the matrix has independent Bernoulli random variables $\{\pm 1\}$ as entries above the main diagonal and satisfies

\[ H_{uv} = 0 \quad \text{if} \quad \min(|u - v|, N - |u - v|) > W \text{ or } u = v. \]

At least for Wigner matrices, one would expect an eigenvector $v$ to not deviate significantly from the following behavior: pick each entry from a nice probability distribution and normalize its $\ell^2$-norm in the end. In particular, we would expect

• equidistribution of mass: for any fixed interval $I \subset \mathbb{Z}_N$

\[ |v|_{\ell^2(I)} \sim \sqrt{\frac{|I|}{N}} \|v\|_{\ell^2}. \]
• small inner products over intervals: any two distinct eigenvectors \(u, v\) are orthogonal \((u, v) = 0\). If we restrict their interaction to any fixed interval \(I \subset \mathbb{Z}_N\), we would expect them to behave like any random vectors with entries \(X_{ij}/\sqrt{n}\), where \(X_{ij}\) with \(1 \leq i, j \leq N\) is a random variable with \(\mathbb{E}X_{ij}^2 = 1\). In this case, we would have

\[
\left| \sum_{i \in I} u_i v_i \right| \sim \frac{\sqrt{T}}{N}.
\]

Both statements are somehow suggested by the fact that a Wigner matrix has interactions between all nodes and there is no reason why locally structured behavior should arise from that. Note that, since we expect random behavior at a local scale, the central limit theorem suggests both statements to be quite accurate for \(|I|\) large (and even stronger statements could be conjectured).

1.3. Delocalization. These two conditions are the basis of our notion of delocalization: we call a set of vectors delocalized if these two properties hold. This notion of delocalization is way too strong: we are not aware of it having been established even for Wigner matrices; of course we do believe that eigenvectors for Wigner matrices (even those with eigenvalues in the spectral edge) are indeed delocalized in this sense. We derive a weaker notion of delocalization by allowing the notion to fail in a quantitative sense.

**Definition.** Fix positive parameters \(\gamma, \delta > 0\). We say that a set \(\Lambda\) of vectors from \(\mathbb{R}^N\) is \((\gamma, \delta)\)–delocalized if

\[
\text{for every } v \in \Lambda, \text{ every interval } I \text{ of length } |I| \sim N^\delta \text{ contains at most } 1\% \text{ of the total } \ell^2\text{–mass of } v, \text{ i.e.}
\]

\[
\|v\|_{\ell^2(I)} \leq \frac{1}{100}\|v\|_{\ell^2}
\]

and for all intervals \(I \subset \mathbb{Z}_N\) with \(|I| \sim N^\delta\)

\[
\forall u, v \in \Lambda \quad \left| \sum_{i \in I} u_i v_i \right| \lesssim N^{\gamma} \frac{\sqrt{T}}{N} \|u\|_{\ell^2} \|v\|_{\ell^2}.
\]

This definition should be understood starting from random vectors. If we pick \(N\) random vectors by taking each entry to be independently and normally distributed, these vectors will be \((0, \delta)\)–delocalized for any \(\delta > 0\) with probability tending rapidly to 1 as \(N \to \infty\). The first condition becomes more restrictive with growing \(\delta\) while the second condition, for fixed \(\delta\), becomes easier to satisfy with growing \(\gamma\). From Hölder’s inequality, we see that the first condition implies the second condition for

\[
\gamma + \frac{\delta}{2} \geq 1.
\]

The first condition also has some connection to the idea of measuring deviation from purely random behavior by looking at \(\ell^p\)–norms of \(\ell^2\)–normalized eigenvectors with \(p \neq 2\). This idea goes back to T. Spencer (see, for example, the result of Erdős, Schlein & Yau [3] for results on eigenvectors of Wigner matrices). We will exclude \((\gamma, \delta)\)–delocalization for certain \(\gamma, \delta\). Note that excluding any form of \((\gamma, \delta)\)–delocalization means that there is either a disproportionate amount of mass of one eigenvector or strong interaction between two eigenvectors at the scale \(N^{\delta}\).

1.4. The statement. Our statement says that there is some structure for eigenvalues in the spectral edge at least up to \(W \ll N^{5/7}\). This is, of course, expected up to \(W \ll N^{1/2}\) for all eigenvalues. We expect localization in the spectral edge to go beyond the transition, however, it is less clear what precise form of localization to expect in the range \(W \gg N^{1/2}\). Our result is only phrased in the case of periodic band matrices, however, the argument is very general: it applies to any band matrix, where one has sufficiently good control on the number of eigenvalues in the spectral edge. Indeed, if one were to obtain even more refined asymptotics on the distribution of eigenvalues in the case of periodic band matrices, it would automatically improve our result. Our argument will also work for symmetric, band matrices with entries decaying sufficiently quickly away from the diagonal.
Theorem. Let $W \sim N^\alpha$ for $0 < \alpha < 5/7$. Consider the set of eigenvectors

$$\Lambda = \left\{ v_i : \frac{|\lambda_i|}{\max_{1 \leq j \leq N} |\lambda_j|} \geq 1 - \frac{1}{N^{4/5}} \right\}$$

of size (determined by Sodin)

$$|\Lambda| \sim N^{1-\frac{4\alpha}{5}}.$$

For every $\varepsilon > 0$ and with probability arbitrarily close to 1

$$\Lambda \text{ is NOT } (\frac{\alpha}{2} - \varepsilon, \frac{7\alpha}{5}) - \text{delocalized.}$$

1.5. Remarks. We consider this statement a modest first step towards establishing structure theorems for eigenvectors in the spectral edge. It states that either mass concentration or strong local correlation holds true – we believe that $(\frac{\alpha}{2} - \varepsilon, \frac{7\alpha}{5}) - \text{delocalization}$ fails because there is mass concentration at scales smaller than $N^\delta$ with $\delta = \frac{7\alpha}{5}$. We consider the extent in which local correlation does or does not occur to be highly interesting in itself – what can be proven?

There is one implicit relation we did not make clear in the statement of the Theorem for reasons of brevity: it states that we can exclude $(\frac{\alpha}{2} - \varepsilon, \frac{7\alpha}{5}) - \text{delocalization}$ with positive probability arbitrarily close to 1. If we wish to exclude it with probability $p = 1 - \varepsilon$, then we need to consider intervals of size $c(\varepsilon)N^\delta$, where the constant $c(\varepsilon)$ tends (very slowly) to infinity as $\varepsilon$ goes to 0.

It is maybe instructive to consider an example. We pick $W \sim N^{4/7}$ and consider the set $\Lambda$, which in this case satisfies

$$|\Lambda| \sim N^{\frac{4}{7}}.$$

Then for any $\varepsilon > 0$ this set either contains a vector $v \in \Lambda$ that has 1% of its mass on an interval $I$ of size $|I| \sim N^{\frac{4}{7}}$ or there are two vectors $v, w$ that correlate on $I$ a polynomial factor stronger than random variables would, i.e.

$$\left| \sum_{i \in I} u_i v_i \right| \gtrsim N^{\frac{4}{7} - \varepsilon} \frac{\sqrt{|I|}}{N}.$$

2. Proof of the Theorem

Main Idea. The proof is a simple combination of several ideas: given a matrix $H$ and a vector $w$, the vector $H^k w$ is for large $k$ essentially a linear superposition of the eigenvectors of $H$ not orthogonal to $w$ whose eigenvalues are large. There is a statement due to Sodin [7] about the distribution of large eigenvalues in our random matrix model. By taking the vector to be supported on a single entry $w = \delta_x$ for some $x \in \mathbb{Z}_N$, we know that $H^k w$ has supported on an interval of length $kW$ (this part of the argument prefers $k$ to be small). However, since we are dealing with a diffusion process, we would expect that while $H^k w$ is typically supported on an interval of length $kW$, most of the $l^2$–mass lives on a much smaller interval of size $\sqrt{kW}$. Using the moment method, this is easily seen to be true most of the time. In particular, for the right $k$, the vectors $H^k w$ will be ‘almost’ compactly supported on a small scale while essentially being a linear combination of relatively few eigenvectors (with eigenvalues from the spectral edge).

2.1. Elementary Probability. The following statement is almost certainly known. However, it seems to be easier to prove than to find in the literature.

Lemma 1. Let $A, B$ be real random variables with $0 < A \leq B \leq 1$. Then

$$E \frac{A}{B} \geq \left( 1 - \sqrt{1 - \frac{E A}{E B}} \right)^2.$$
Proof. If $EA = EB$, the statement is trivially true. Otherwise, consider $\varepsilon > 0$ given by
\[
\varepsilon = 1 - \frac{EA}{EB}.
\]
Let $c > 0$ be a variable to be fixed later. Abbreviate also
\[
x = P(A < (1 - c\varepsilon)B).
\]
Then
\[
1 - \varepsilon = \frac{EA}{EB} \leq (1 - c\varepsilon)x + (1 - x) = 1 - c\varepsilon x.
\]
As a consequence $x \leq 1/c$. However, we easily see that
\[
\frac{EA}{EB} \geq (1 - x)(1 - c\varepsilon) \geq \left(1 - \frac{1}{c}\right)(1 - c\varepsilon).
\]
Setting $c = 1/\sqrt{\varepsilon}$ yields the desired statement.

2.2. Counting paths. Consider the lattice $\mathbb{Z}$. We are interested in the behavior of a random walk starting in $0$ and making $k$ jumps randomly chosen from
\[
\{-W, -W + 1, \ldots, 0, \ldots W - 1, W\}.
\]
It follows from the central limit theorem that the probability distribution of such a jump process has mean zero and an average deviation of $\sim \sqrt{kW}$ with an exponentially decaying tail and approximates in shape a Gaussian as $k$ becomes large. We need a quantitative version and invoke the Berry-Essen theorem. The first three moments of a single jump $X$ are
\[
EX = 0, \quad EX^2 = \sum_{n=1}^{W} \frac{i^2}{W} \sim \frac{W^2}{3} \quad \text{and} \quad EX^3 = \sum_{n=1}^{W} \frac{i^3}{W} \sim \frac{W^3}{4}.
\]
The critical quantity in the Berry-Essen theorem is
\[
\frac{EX^3}{(EX^2)^{3/2}} \sim 1.
\]
In particular, for results at our required level of accuracy, the approximation with a Gaussian is justified as long as $k \gg 1$. The same question can be asked on the torus $\mathbb{Z}_N$ with $W \ll N$ and similar results hold true. We expect a ‘wrapping of the heat kernel around the torus’, however, operating on a much rougher level suffices.

Lemma 2. The asymptotics for the random walk continue to hold true on the torus $\mathbb{Z}_N$, i.e. for a jump process $0 = x_0, x_1, \ldots, x_k$ starting in $0$ with $k$ jumps uniformly distributed in
\[
\{-W, -W + 1, \ldots, 0, \ldots W - 1, W\}
\]
we have
\[
P\left(|x_k| \leq c\sqrt{kW}\right) = 1 - o(1),
\]
where $o(1)$ is to be understood with respect to $c$ becoming large.

Proof. For any random jump process in $\mathbb{Z}$, we can consider its projection
\[
\pi : \mathbb{Z} \to \mathbb{Z}_N.
\]
Every path in $\mathbb{Z}$ thus corresponds to a path in $\mathbb{Z}_N$. Since the statement is true on $\mathbb{Z}$, it is certainly true for $\mathbb{Z}_N$. Indeed, we could actually say more and prove the same result for a smaller class of random walks on $\mathbb{Z}_N$ (those, which do not end up being close to the origin by ‘going several times around the circle’).
2.3. Proof of the Theorem.

Proof. Suppose \( W \sim N^\alpha \). Denote the (real) eigenvectors of the random matrix \( H \) by \( (v_i)_{i=1}^N \) with associated eigenvalues \( \lambda_i \) ordered in such a way that \( |\lambda_1| \leq |\lambda_2| \leq \cdots \leq |\lambda_N| \). Take the vector \( v_N \) and choose (by pigeonholing) a number \( 1 \leq i \leq N \) such that the \( i \)-th component of \( v_N \) is large, i.e.

\[
|\delta_i v_N| \geq \frac{1}{\sqrt{N}}.
\]

Assume w.l.o.g. that \( i = 0 \). Take the standard basis vector \( x = (1, 0, \ldots, 0) \) and represent it as a linear combination of eigenvectors

\[
(1, 0, \ldots, 0) = \sum_{i=1}^N a_i v_i.
\]

By construction \( \|a\|_{\ell^2} = 1 \) and \( |a_N| \geq N^{-1/2} \). For \( k \sim N^{4\alpha/5} \), we study the vector \( x \in \ell^2(\mathbb{Z}_N) \) with entries

\[
(x_1, \ldots, x_N) := H^k (1, 0, \ldots, 0)^t.
\]

We are interested in showing that the bulk of the \( \ell^2 \)-mass of this vector \( x \) is concentrated in an interval of size \( \sim \sqrt{kW} (\sim N^{7\alpha/5}) \) with large probability. The expectation \( E x_j^2 \) can be written as a sum over paths, i.e.

\[
E x_j^2 = E \left( \sum_{i_1, \ldots, i_{k-1}} H_{i_1,1} H_{i_2, i_1} \cdots H_{i_{k-1}, i_{k-2}} H_{j, i_{k-1}} \right)^2
\]

i.e. merely the number of possible paths of length \( k \) from 0 to \( j \), where any two consecutive sites \( i_k \) and \( i_{k+1} \) have distance at most \( W \) on \( \mathbb{Z}_N \). For some large constant \( c \) we define two random variables

\[
A = \sum_{|j| \leq c \sqrt{kW}} x_j^2 \quad \text{and} \quad B = \sum_{j=1}^N x_j^2.
\]

Expectation is linear and Lemma 2 implies that

\[
\frac{E A}{E B} \geq 1 - o(1),
\]

where \( o(1) \) is to be understood as \( c \to \infty \). Lemma 1 implies that then

\[
\frac{E A}{E B} \geq 1 - o(1),
\]

where \( o(1) \) is to be understood in the same sense. Then, however, for any \( \varepsilon > 0 \)

\[
P \left( \frac{A}{B} \geq 1 - \varepsilon \right) \geq 1 - o_{\varepsilon}(1).
\]

Therefore, for any fixed probability \( p \) arbitrarily close to 1 and \( W, N \) sufficiently large, we find a suitable \( c \) such that a positive fraction of the \( \ell^2 \)-mass of the vector \( (x_1, \ldots, x_N) \) is contained in a \( c \sqrt{kW} \) interval with probability \( p \). At the same time, we have from the decomposition into eigenvalues that

\[
(x_1, \ldots, x_N) = H^k (1, 0, \ldots, 0) = \sum_{i=1}^N a_i \lambda_i^k v_i,
\]

where only few contributions are actually large. Thanks to the elementary statement

\[
\forall t \in \mathbb{R} \quad \lim_{k \to \infty} \left( 1 - \frac{t}{k} \right)^k = e^{-t},
\]
we can conclude that all eigenvectors \( v_i \) with (the precise power of the logarithm is not very important now or in what follows)

\[
\left| \frac{\lambda_i}{\lambda_N} \right| \leq 1 - \frac{(\log N)^3}{N^{4\alpha/5}}
\]

become relatively unimportant as (remember \( |a_N| \geq N^{-1/2} \))

\[
\left| \frac{a_i \lambda_i^k}{a_N \lambda_N^k} \right| \leq \frac{1}{N^{2.5}}.
\]

In particular, the bulk of the \( \ell^2 \)-mass of \((x_1, \ldots, x_N)\) is carried by

\[
\Lambda = \left\{ v_i \mid \frac{|\lambda_i|}{|\lambda_N|} \geq 1 - \frac{(\log N)^3}{N^{4\alpha/5}} \right\}.
\]

From Sodin’s result [7, Theorem 1.1, Case 2], we know that

\[
|\Lambda| \lesssim (\log N)^5 N^{1 - \frac{10}{7\gamma + 1}}.
\]

If we denote the \( c\sqrt{KW} \sim cN^{7\alpha/5} \) interval around the origin by \( I \), then we have just shown that with positive probability (arbitrarily close to 1 depending on \( c \) in the way indicated above)

\[
1 \lesssim \left\| \sum_{v_i \in \Lambda} a_i v_i \right\|_{\ell^2(I)}^2 = \sum_{v_i \in \Lambda} a_i^2 \|v_i\|_{\ell^2(I)}^2 + \sum_{v_i, v_j \in \Lambda, v_i \neq v_j} a_i a_j \left( \sum_{x \in I} \delta_x v_i \delta_x v_j \right)
\]

Now we can argue by contradiction: if the statement was false, there would be no mass concentration on small scales,

\[
\sum_{v_i \in \Lambda} a_i^2 \|v_i\|_{\ell^2(I)}^2 \leq \sup_{v_i \in \Lambda} \|v_i\|_{\ell^2(I)}^2 \left( \sum_{v_i \in \Lambda} a_i^2 \right) \leq \sup_{v_i \in \Lambda} \|v_i\|_{\ell^2(I)}^2 \ll 1.
\]

and no strong interaction between two distinct eigenvectors: with Cauchy-Schwarz, our choice of \( \gamma \) and Hölder’s inequality

\[
\sum_{v_i, v_j \in \Lambda, v_i \neq v_j} a_i a_j \left( \sum_{x \in I} \delta_x v_i \delta_x v_j \right) \leq N^{\gamma + \frac{7}{7\gamma + 1} - 1} \sum_{v_i, v_j \in \Lambda} |a_i| |a_j|
\]

\[
\leq N^{\gamma + \frac{7}{7\gamma + 1} - 1} \left( \sum_{v_i \in \Lambda} |a_i| \right)^2
\]

\[
\leq N^{\gamma + \frac{7}{7\gamma + 1} - 1} |\Lambda|
\]

\[
\leq N^{\gamma + \frac{7}{7\gamma + 1} - 1} (\log N)^5 N^{1 - \frac{10}{7\gamma + 1}} \ll 1.
\]

\[\square\]

**Acknowledgements.** The author acknowledges and is grateful for helpful conversations with László Erdős, Antti Knowles and Benjamin Schlein.

**References**

[1] L. Erdős, and A. Knowles, Quantum diffusion and eigenfunction delocalization in a random band matrix model. Comm. Math. Phys. 303 (2011), no. 2, 509–554.

[2] L. Erdős, A. Knowles, H.-T. Yau and J. Yin, Delocalization and Diffusion Profile for Random Band Matrices, to appear in Comm. Math. Phys.

[3] 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), no. 3, 815–852.

[4] Y.V. Fyodorov and A.D. Mirlin, Scaling properties of localization in random band matrices: a \( \sigma \)-model approach, Phys. Rev. Lett. 67 (1991), 2405–2409.
[5] M. Rudelson & R. Vershynin, Delocalization of eigenvectors of random matrices with independent entries, arXiv:1306.2887.

[6] J. Schenker, Eigenvector localization for random band matrices with power law band width, Comm. Math. Phys. 290 (2009), 1065–1097.

[7] S. Sodin, The spectral edge of some random band matrices, Ann. Math. 172 (2010), no. 3, 2223–2251.

Mathematisches Institut, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany