Eigenvectors of random matrices: A survey

CrossMark

Sean O'Rourke [a], Van Vu [b,1], Ke Wang [c,d]

[a] *Department of Mathematics, University of Colorado at Boulder, Boulder, CO 80309, USA*
[b] *Department of Mathematics, Yale University, New Haven, CT 06520, USA*
[c] *Jockey Club Institute for Advanced Study, Hong Kong University of Science and Technology, Hong Kong, China*
[d] *Computing & Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125, USA*

A R T I C L E   I N F O

A B S T R A C T

Eigenvectors of large matrices (and graphs) play an essential role in combinatorics and theoretical computer science. The goal of this survey is to provide an up-to-date account on properties of eigenvectors when the matrix (or graph) is random.

Published by Elsevier Inc.

Contents

1. Introduction

Eigenvectors of large matrices (and graphs) play an essential role in combinatorics and theoretical computer science. For instance, many properties of a graph can be deduced or estimated from its eigenvectors. In recent years, many algorithms have been developed which take advantage of this relationship to study various problems including spectral clustering [68,85], spectral partitioning [50,60], PageRank [57], and community detection [52,53].

The goal of this survey is to study basic properties of eigenvectors when the matrix (or graph) is random. As this survey is written with combinatorics/theoretical computer science readers in mind, we try to formalize the results in forms which are closest to their interest and give references for further extensions. Some of the results presented in this paper are new with proofs included, while many others have appeared in very recent papers.

We focus on the following models of random matrices.

Definition 1.1 *(Wigner matrix).* Let ξ, ζ be real random variables with mean zero. We say W is a *Wigner matrix* of size n with atom variables ξ, ζ if $W = (w_{ij})_{i,j=1}^n$ is a random real symmetric $n \times n$ matrix that satisfies the following conditions.

- (independence) $\{w_{ij} : 1 \le i \le j \le n\}$ is a collection of independent random variables.
- (off-diagonal entries) $\{w_{ij} : 1 \le i < j \le n\}$ is a collection of independent and identically distributed (iid) copies of ξ.
- (diagonal entries) $\{w_{ii} : 1 \le i \le n\}$ is a collection of iid copies of ζ.

If ξ and ζ have the same distribution, we say W is a Wigner matrix with atom variable ξ. In this case, W is a real symmetric matrix whose entries on and above the diagonal are iid copies of ξ.

One can similarly define complex Hermitian Wigner matrices whose off-diagonal entries are complex-valued random variables. For the purposes of this survey, we focus on real symmetric Wigner matrices.

Throughout the paper, we consider various assumptions on the atom variables ξ and ζ. We will always assume that ξ is *non-degenerate*, namely that there is no value c such that $\mathbb{P}(\xi = c) = 1$.

Definition 1.2 *(K-bounded).* A random variable ξ is *K-bounded* if $|\xi| \le K$ almost surely. For combinatorial applications, the entries usually take on values $\{0, \pm 1\}$ and are 1-bounded. In general, we allow K to depend on n.

Occasionally, we will assume the atom variable ξ is symmetric. Recall that a random variable ξ is *symmetric* if ξ and $-\xi$ have the same distribution. The Rademacher random variable, which take the values ± 1 with equal probability is an example of a symmetric random variable.

Definition 1.3 *(Sub-exponential).* A random variable ξ is called *sub-exponential* with exponent $\alpha > 0$ if there exists a constant $\beta > 0$ such that

$$\mathbb{P}(|\xi| > t) \le \beta \exp(-t^\alpha / \beta) \quad \text{for all } t > 0. \tag{1}$$

If $\alpha = 2$, then ξ is called *sub-gaussian*, and $1/\beta$ is the *sub-gaussian moment* of ξ.

The prototypical example of a Wigner real symmetric matrix is the *Gaussian orthogonal ensemble* (GOE). The GOE is defined as a Wigner random matrix with atom variables ξ, ζ, where ξ is a standard normal random variable and ζ is a normal random variable with mean zero and variance 2. The GOE is widely-studied in random matrix theory and mathematical physics. However, due to its continuous nature, the GOE has little use in combinatorial applications.

A case of principal interest in combinatorial applications is when both ξ and ζ are Bernoulli random variables. Let $0 < p < 1$, and take ξ to be the random variable

$$\xi := \begin{cases} 1 - p, & \text{with probability } p, \\ -p, & \text{with probability } 1 - p. \end{cases} \tag{2}$$

In particular, ξ has mean zero by construction. Let $B_n(p)$ denote the $n \times n$ Wigner matrix with atom variable ξ (i.e. the entries on and above the diagonal are iid copies of ξ). We refer to $B_n(p)$ as a *symmetric Bernoulli matrix* (with parameter p). The most interesting case is when $p = 1/2$. In this case, $2B_n(p)$ is the random symmetric Rademacher matrix, whose entries are ± 1 with probability $1/2$.

In applications, one often considers the adjacency matrix of a random graph. We let $G(n, p)$ denote the Erdös–Rényi random graph on n vertices with edge density p. That is, $G(n, p)$ is a simple graph on n vertices such that each edge $\{i, j\}$ is in $G(n, p)$ with probability p, independent of other edges. We let $A_n(p)$ be the zero-one adjacency matrix of $G(n, p)$. $A_n(p)$ is not a Wigner matrix since its entries do not have mean zero.

For the sake of simplicity, we will sometimes consider random graphs with loops (thus the diagonals of the adjacency matrix are also random). Let $\tilde{G}(n, p)$ denote the Erdös–Rényi random graph with loops on n vertices with edge density p. That is, $\tilde{G}(n, p)$ is a graph on n vertices such that each edge $\{i, j\}$ (including the case when $i = j$) is in $\tilde{G}(n, p)$ with probability p, independent of other edges. We let $\tilde{A}_n(p)$ denote the zero-one adjacency matrix of $\tilde{G}(n, p)$. Technically, $\tilde{A}_n(p)$ is not a Wigner random matrix because its entries do not have mean zero. However, we can view $\tilde{A}_n(p)$ as a low rank deterministic perturbation of a Wigner matrix. That is, we can write $\tilde{A}_n(p)$ as

$$\tilde{A}_n(p) = p\mathbf{J}_n + B_n(p),$$

where $\mathbf{J}_n$ is the all-ones matrix. We also observe that $A_n(p)$ can be formed from $\tilde{A}_n(p)$ by replacing the diagonal entries with zeros. In this note, we focus on the case when p is a constant, independent of the dimension n, but will also give references to the case when p decays with n.

For an $n \times n$ Hermitian matrix M, we let $\lambda_1(M) \leq \cdots \leq \lambda_n(M)$ denote the ordered eigenvalues of M (counted with multiplicity) with corresponding unit eigenvectors $v_1(M), \ldots, v_n(M)$. It is important to notice that the eigenvectors of M are *not uniquely determined*. In general, we let $v_1(M), \ldots, v_n(M)$ denote any orthonormal basis of eigenvectors of M such that

$$Mv_i(M) = \lambda_i(M)v_i(M), \quad 1 \leq i \leq n.$$

On the other hand, it is well known that if the spectrum of M is *simple* (i.e. all eigenvalues have multiplicity one) then the unit eigenvectors $v_1(M), \ldots, v_n(M)$ are determined uniquely up to phase. In this case, to avoid ambiguity, we always assume that the eigenvectors are taken so that their first non-zero coordinate is positive. Theorem 1.4 below

shows that, with high probability, the eigenvalues of many matrices under discussion are simple and the coordinates of all eigenvectors are non-zero.

Theorem 1.4 *([81]). Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ.*

- *If ξ is non-degenerate, then, for every $\alpha > 0$, there exists $C > 0$ (depending on α and ξ) such that the spectrum of W is simple with probability at least $1 - Cn^{-\alpha}$.*
- *Moreover, if ξ and ζ are sub-gaussian, then, for every $\alpha > 0$, there exists $C > 0$ (depending on α and ξ) such that every coordinate of every eigenvector of W is non-zero with probability at least $1 - Cn^{-\alpha}$.*

In addition, the conclusions above also hold for the matrices $A_n(p)$ and $\tilde{A}_n(p)$ when $p \in (0, 1)$ is fixed, independent of n.

Remark 1.5. In many cases, the bound $1 - Cn^{-\alpha}$ appearing in Theorem 1.4 can be improved to $1 - C \exp(n^{-c})$ for some constants $C, c > 0$ (see [54]).

Consequently, in many theorems we can assume that the spectrum is simple and the eigenvectors are uniquely defined (using the convention that the first coordinate is positive).

1.1. Overview and outline

Although this survey examines several models of random matrices, we mostly focus on Wigner matrices, specifically Wigner matrices whose atom variables have light tails (e.g. sub-exponential atom variables). In this case, the main message we would like to communicate is that an eigenvector of a Wigner matrix behaves like a random vector uniformly distributed on the unit sphere. While this concept is natural and intuitive, it is usually non-trivial to prove quantitatively. We present several estimates to quantify this statement from different aspects. In particular, we address

- the joint distribution of many coordinates (Section 3),
- the largest coordinate, i.e. the ℓ^∞-norm (Section 4),
- the smallest coordinate (Section 4),
- the ℓ^p-norm (Section 5),
- the amount of mass on a subset of coordinates (Section 5).

For comparison, we also discuss other models of random matrices, such as heavy-tailed and band random matrices, whose eigenvectors do not behave like random vectors uniformly distributed on the unit sphere.

The paper is organized as follows. In Section 2, we present results for the eigenvectors of a matrix drawn from the GOE (defined above). This is the "ideal" situation, where

the eigenvectors are indeed uniformly distributed on the unit sphere, thanks to the rotational invariance of the ensemble. Next, in Section 3, we discuss *universality* results which give a direct comparison between eigenvectors of general Wigner matrices with those of the GOE. In Section 4, we present bounds on the largest coordinate (i.e. the ℓ^∞-norm) and smallest coordinate of eigenvectors of Wigner matrices. Section 5 gives more information about the mass distribution on the coordinates of an eigenvector, such as the magnitude of the smallest coordinates, or the amount of mass one can have on any subset of coordinates of linear size. In Section 6, we present results for deterministic perturbations of Wigner matrices. In particular, this section contains results for the adjacency matrices $A_n(p)$ and $\tilde{A}_n(p)$. In Section 7, we change direction and review two ensembles of random matrices whose eigenvectors do not behave like random vectors uniformly distributed on the unit sphere. In Section 8 and Section 9, we discuss results concerning some other models of random matrices, which have not been mentioned in the introduction, such as random non-symmetric matrices or the adjacency matrix of a random regular graph. In the remaining sections, we represent proofs of the new results mentioned in the previous sections. The appendix contains a number of technical lemmas.

1.2. Notation

The *spectrum* of an $n \times n$ real symmetric matrix M is the multiset $\{\lambda_1(M), \dots, \lambda_n(M)\}$. We use the phrase *bulk* of the spectrum to refer to the eigenvalues $\lambda_i(M)$ with $\varepsilon n \leq i \leq (1-\varepsilon)n$, where ε is a small positive constant. The remaining eigenvalues form the *edge* of the spectrum.

For a vector $v = (v_i)_{i=1}^n \in \mathbb{R}^n$, we let

$$\|v\|_{\ell^p} := \left(\sum_{i=1}^n |v_i|^p \right)^{1/p}$$

denote the ℓ^p-norm of v. We let $\|v\| := \|v\|_{\ell^2}$ be the Euclidean norm of v. For $S \subset [n] := \{1, \dots, n\}$, we denote

$$\|v\|_S := \left(\sum_{i \in S} |v_i|^2 \right)^{1/2}.$$

It follows that $\|v\|_{[n]} = \|v\|$. Let

$$\|v\|_{\ell^\infty} := \max_{1 \leq i \leq n} |v_i|$$

denote the ℓ^∞-norm of v. We introduce the notation

$$\|v\|_{\min} := \min_{1 \leq i \leq n} |v_i| \tag{3}$$

to denote the minimal coordinate (in magnitude) of v; notice that this is *not* a norm. For two vectors $u, v \in \mathbb{R}^n$, we let $u \cdot v = u^{\mathrm{T}}v$ be the dot product between u and v. S^{n-1} is the unit sphere in $\mathbb{R}^n$.

For an $m \times n$ matrix M with real entries, we let $\|M\|$ denote the spectral norm of M:

$$\|M\| := \max_{v \in S^{n-1}} \|Mv\|.$$

We let E^c denote the complement of the event E. For a set S, $|S|$ is the cardinality of S. For two random variables (or vectors) ξ and ζ, we write $\xi \sim \zeta$ if ξ and ζ have the same distribution. We let $N(\mu, \sigma^2)$ denote the normal distribution with mean μ and variance σ^2. In particular, $\xi \sim N(0, 1)$ means ξ is a standard normal random variable. In addition, $|N(0, 1)|$ is the distribution which arises when one takes the absolute value of a standard normal random variable.

We use asymptotic notation under the assumption that n tends to infinity. We write $X = o(Y)$ if $|X| \leq c_n Y$ for some c_n which converges to zero as $n \to \infty$. In particular, $o(1)$ denotes a term which tends to zero as n tends to infinity.

2. A toy case: The Gaussian orthogonal ensemble

The GOE (defined above) is a special example of a Wigner matrix, enjoying the property that it is invariant under orthogonal transformations. By the spectral theorem, any $n \times n$ real symmetric matrix M can be decomposed as $M = UDU^{\mathrm{T}}$, where U is an orthogonal matrix whose columns are the eigenvectors of M and D is a diagonal matrix whose entries are the eigenvalues of M. However, if M is drawn from the GOE, the property of being invariant under orthogonal transformations implies that U and D are independent. It follows from [2, Section 2.5.1] that the eigenvectors $v_1(M), \ldots, v_n(M)$ are uniformly distributed on

$$S_+^{n-1} := \{x = (x_1, \ldots, x_n) \in S^{n-1} : x_1 > 0\},$$

and $(v_1(M), \ldots, v_n(M))$ is distributed like a sample of Haar measure on the orthogonal group $O(n)$, with each column multiplied by a norm one scalar so that the columns all belong to S_+^{n-1}. Additionally, the eigenvalues $(\lambda_1(M), \ldots, \lambda_n(M))$ have joint density

$$p_n(\lambda_1, \ldots, \lambda_n) := \begin{cases} \frac{1}{Z_n} \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j| \prod_{i=1}^n e^{-\lambda_i^2/4}, & \text{if } \lambda_1 \leq \cdots \leq \lambda_n, \\ 0, & \text{otherwise,} \end{cases}$$

where Z_n is a normalization constant. We refer the interested reader to [2, Section 2.5.1] for a further discussion of these results as well as additional references.

In the following statements, we gather information about a unit eigenvector v of a $n \times n$ matrix drawn from the GOE. As discussed above, this is equivalent to studying a random vector v uniformly distributed on S_+^{n-1}. In fact, since all of the properties

discussed below are invariant under scaling by a norm one scalar, we state all of the results in this section for a unit vector v uniformly distributed over the unit sphere S^{n-1} of $\mathbb{R}^n$.

Theorem 2.1 *(Smallest and largest coordinates). Let v be a random vector uniformly distributed on the unit sphere S^{n-1}. Then, for any $C > 1$, with probability at least $1 - 2n^{1-C} - \exp\left(-\frac{(C-1)^2}{4C^2}n\right)$,*

$$\|v\|_{\ell^\infty} \leq \sqrt{\frac{2C^3 \log n}{n}}. \tag{4}$$

In addition, for $n \geq 2$, any $0 \leq c < 1$, and any $a > 1$,

$$\|v\|_{\min} \geq \frac{c}{a}\frac{1}{n^{3/2}} \tag{5}$$

with probability at least $\exp\left(-2c\right) - \exp\left(-\frac{a^2 - \sqrt{2a^2-1}}{2}n\right)$.

Next, we obtain the order of the ℓ^p-norm of such a vector.

Theorem 2.2 *(ℓ^p-norm). Let v be a random vector uniformly distributed on the unit sphere S^{n-1}. Then, for any $p \geq 1$, there exists $c_p > 0$, such that almost surely*

$$\|v\|_{\ell^p}^p = n^{1-p/2}c_p + o(n^{1-p/2}).$$

We now consider the distribution of mass over the coordinates of such a vector. Recall that the beta distribution with shape parameters $\alpha, \beta > 0$ is the continuous probability distribution supported on the interval $[0, 1]$ with probability density function

$$f(x; \alpha, \beta) := \begin{cases} \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}, & \text{if } 0 \leq x \leq 1 \\ 0, & \text{otherwise,} \end{cases}$$

where $B(\alpha, \beta)$ is the normalization constant. A random variable ξ beta-distributed with shape parameters $\alpha, \beta > 0$ will be denoted by $\xi \sim \text{Beta}(\alpha, \beta)$.

Theorem 2.3 *(Distribution of mass). Let S be a proper nonempty subset of $[n]$, and let v be a random vector uniformly distributed on the unit sphere S^{n-1}. Then $\|v\|_S^2$ is distributed according to the beta distribution*

$$\|v\|_S^2 \sim \text{Beta}\left(\frac{|S|}{2}, \frac{n-|S|}{2}\right).$$

In particular, this implies that $\|v\|_S^2$ has mean $\frac{|S|}{n}$ and variance $\frac{|S|(n-|S|)}{n^2(n/2+1)}$.

As a corollary of Theorem 2.3, we obtain the following central limit theorem.

Theorem 2.4 *(Central limit theorem). Fix $\delta \in (0,1)$. For each $n \geq 2$, let $S_n \subset [n]$ with $|S_n| = \lfloor \delta n \rfloor$, and let v_n be a random vector uniformly distributed on the unit sphere S^{n-1}. Then*

$$\sqrt{\frac{n^3}{2|S_n|(n-|S_n|)}} \left(\|v_n\|_{S_n}^2 - \frac{|S_n|}{n} \right) \longrightarrow N(0,1)$$

in distribution as $n \to \infty$.

We also have the following concentration inequality.

Theorem 2.5 *(Concentration). Let $S \subset [n]$, and let v be a random vector uniformly distributed on the unit sphere S^{n-1}. Then, for any $t > 0$,*

$$\left| \|v\|_S^2 - \frac{|S|}{n} \right| \leq \frac{8}{n} \left(\sqrt{nt} + t \right)$$

with probability at least $1 - \exp(-cn) - 4\exp(-t)$, where $c > 0$ is an absolute constant.

Remark 2.6. The proof of Theorem 2.5 actually reveals that, for any $t > 0$,

$$\left| \|v\|_S^2 - \frac{|S|}{n} \right| \leq \frac{4}{n} \left(\sqrt{|S|t} + t \right) + 4\frac{|S|}{n^2} \left(\sqrt{nt} + t \right)$$

with probability at least $1 - \exp(-cn) - 4\exp(-t)$.

We now consider the maximum and minimum order statistics

$$\max_{S \subset [n] : |S| = \lfloor \delta n \rfloor} \|v\|_S \quad \text{and} \quad \min_{S \subset [n] : |S| = \lfloor \delta n \rfloor} \|v\|_S$$

for some $0 < \delta < 1$. Recall that the χ^2-distribution with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables. Let F be the cumulative distribution function of the χ^2-distribution with one degree of freedom. Following the notation in [19], let Q denote the quantile function of F. That is,

$$Q(s) := \inf\{x \in \mathbb{R} : F(x) \geq s\}, \quad 0 < s \leq 1, \quad Q(0) := \lim_{s \searrow 0} Q(s). \tag{6}$$

Define

$$H(s) := -Q(1-s), \quad 0 \leq s < 1. \tag{7}$$

Theorem 2.7 *(Extreme order statistics). For each $n \geq 1$, let v_n be a random vector uniformly distributed on the unit sphere S^{n-1}. Then, for any fixed $0 < \delta < 1$,*

$$\max_{S \subset [n]:|S|=\lfloor \delta n \rfloor} \|v_n\|_S^2 \longrightarrow -\int_0^\delta H(u)\, du$$

and

$$\min_{S \subset [n]:|S|=\lfloor \delta n \rfloor} \|v_n\|_S^2 \longrightarrow -\int_{1-\delta}^1 H(u)\, du$$

in probability as $n \to \infty$.

Remark 2.8. The integrals on the right-hand side involving H can be rewritten in terms of the standard normal distribution. Indeed, from (32) and integration by parts, we find

$$-\int_0^\delta H(u)\, du = 2 \int_{\Phi^{-1}(1-\frac{\delta}{2})}^\infty x^2 \Phi'(x)\, dx$$

$$= \delta + \sqrt{\frac{2}{\pi}} \Phi^{-1}\left(1 - \frac{\delta}{2}\right) \exp\left(-\frac{1}{2}\left[\Phi^{-1}\left(1 - \frac{\delta}{2}\right)\right]^2\right),$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. We also get

$$-\int_{1-\delta}^1 H(u)\, du = 1 + \int_0^{1-\delta} H(u)\, du$$

$$= \delta - \sqrt{\frac{2}{\pi}} \Phi^{-1}\left(\frac{1+\delta}{2}\right) \exp\left(-\frac{1}{2}\left[\Phi^{-1}\left(\frac{1+\delta}{2}\right)\right]^2\right).$$

Remark 2.9. Using the expressions in Remark 2.8, one can show that, as δ tends to zero,

$$-\int_0^\delta H(u)\, du = \Theta\left(\delta \log(\delta^{-1})\right)$$

and

$$-\int_{1-\delta}^1 H(u)\, du = \Theta(\delta^3).$$
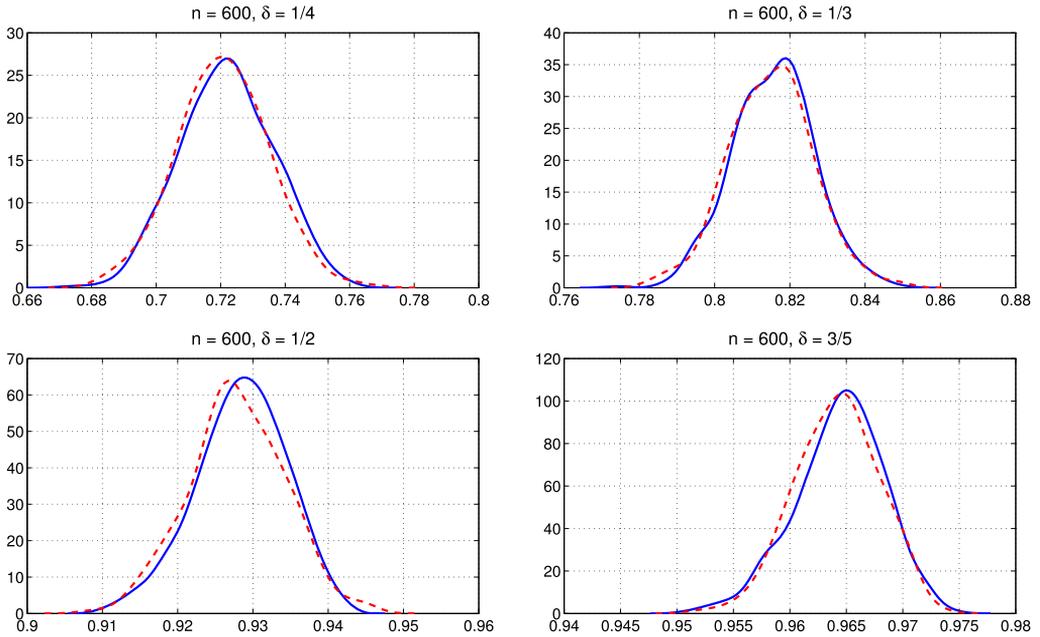
Fig. 1. The probability density function of the distribution of $\max_{S\subset[n]:|S|=\lfloor\delta n\rfloor}\|v\|_S^2$ when $n=600$, based on 500 samples. The blue curve is for v, a unit eigenvector of a matrix drawn from the GOE, while the red dashed curve corresponds to a unit eigenvector of a random symmetric Bernoulli matrix.

In other words, Theorem 2.7 implies that the smallest δn coordinates of an eigenvector contribute only $\Theta(\delta^3)$ fraction of the mass.

Fig. 1 depicts numerical simulations of the distribution of $\max_{S\subset[n]:|S|=\lfloor\delta n\rfloor}\|v_n\|_S^2$ when $n=600$. The simulation shows that $\max_{S\subset[n]:|S|=\lfloor\delta n\rfloor}\|v_n\|_S^2$ is highly concentrated at the value $g(\delta):=-\int_0^\delta H(u)\,du$ as indicated in Theorem 2.7. Indeed, numerical calculations show

$$g(1/4)\approx 0.7236069618,\quad g(1/3)\approx 0.8167557098,$$

$$g(1/2)\approx 0.9286740823,\quad g(3/5)\approx 0.9646603703.$$

Our next result shows that these extreme order statistics concentrate around their expectation, even for relatively small values of n.

Theorem 2.10 *(Concentration of the extreme order statistic). Let v be a random vector uniformly distributed on the unit sphere S^{n-1}. Then, for any $1\le m\le n$ and every $t\ge 0$,*

$$\mathbb{P}\Big(\Big|\max_{S\subset[n]:|S|=m}\|v\|_S-\mathbb{E}\max_{S\subset[n]:|S|=m}\|v\|_S\Big|>t\Big)\le C\exp(-cnt^2)\tag{8}$$

and

$$\mathbb{P}\left(\left|\min_{S\subset[n]:|S|=m}\|v\|_S - \mathbb{E}\min_{S\subset[n]:|S|=m}\|v\|_S\right| > t\right) \leq C\exp(-cnt^2) \tag{9}$$

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

We prove these results in Section 10.

3. Direct comparison theorems

In many cases, one can compare the eigenvectors of a Wigner matrix directly to those of the GOE. In the random matrix theory literature, such results are often referred to as *universality results.*

Let W be a Wigner matrix with atom variables ξ, ζ. We will require that ξ and ζ satisfy a few moment conditions. In particular, we say ξ, ζ are from the class $\mathcal{M}_4$ if

- ξ and ζ are sub-exponential random variables,
- $\mathbb{E}(\xi) = \mathbb{E}(\zeta) = \mathbb{E}(\xi^3) = 0$, and
- $\mathbb{E}(\xi^2) = 1, \mathbb{E}(\xi^4) = 3, \mathbb{E}(\zeta^2) = 2$.

These conditions imply that the first four moments of the off-diagonal entries of W match those of the GOE, and the first 2 moments of the diagonal entries of W match those of the GOE.

Let $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ be two sequences of real random variables. In order to show that they have (asymptotically) the same distribution, it suffices to show that

$$\mathbb{P}(X_n \leq t) - \mathbb{P}(Y_n \leq t) = o(1)$$

for all t. Notice that

$$\mathbb{P}(X_n \leq t) = \mathbb{E}f_t(X_n),$$

where $f_t(X_n)$ is the indicator function of the event $\{X_n \leq t\}$. In practice, it is useful to smoothen f_t a little bit to obtain a function with bounded derivatives, at the cost of an extra error term. We are going to use this strategy in the next few theorems, which allow us to compare X_n with Y_n by bounding $\mathbb{E}F(X_n) - \mathbb{E}F(Y_n)$ for a large set of test functions F.

One of the first universality results for eigenvectors is the following result.

Theorem 3.1 *(Eigenvector coefficients of Wigner matrices, [79]). Given $C > 0$ and random variables ξ, ζ from the class $\mathcal{M}_4$, there are constants $\delta, C_0 > 0$ such that the following statement holds. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ and with unit eigenvectors $v_1, \ldots, v_n$. Let $v_i(j)$ denote the jth entry of v_i. For $1 \leq i, j \leq n$, let $Z_{i,j}$ be independent random variables with $Z_{i,j} \sim N(0,1)$ for $j > 1$ and $Z_{i,1} \sim |N(0,1)|$.*

Let $1 \leq k \leq n^{\delta}$, and let $1 \leq i_1 < \ldots < i_k \leq n$ and $1 \leq j_1 < \ldots < j_k \leq n$ be indices. Then

$$\left| \mathbb{E} F \left((\sqrt{n} v_{i_a}(j_b))_{1 \leq a,b \leq k} \right) - \mathbb{E} F \left((Z_{i_a,j_b})_{1 \leq a,b \leq k} \right) \right| \leq C_0 n^{-\delta} \tag{10}$$

whenever $F : \mathbb{R}^{k^2} \to \mathbb{R}$ is a smooth function obeying the bounds

$$|F(x)| \leq C$$

and

$$|\nabla^j F(x)| \leq C n^{\delta}$$

for all $x \in \mathbb{R}^{k^2}$ and $0 \leq j \leq 5$.

Theorem 3.1 is essentially the first part of [79, Theorem 7]. The original theorem has $o(1)$ on the right-hand side of (10), but one can obtain the better bound $C_0 n^{-\delta}$ using the same proof. The essence of this theorem is that any set of at most n^{δ} coordinates (which may come from different eigenvectors) behaves like a set of iid Gaussian random variables. Because of our normalization, requiring the first non-zero coordinate of the eigenvectors of W to be positive, we cannot expect this coordinate to be close to a normal random variable. However, Theorem 3.1 shows that it is close in distribution to the absolute value of a standard normal random variable. Similar results were also obtained in [45] for the case when $k = O(1)$.

Another way to show that an eigenvector v behaves like a random vector u uniformly distributed on the unit sphere is to fix a unit vector a, and compare the distribution of the inner product $v \cdot a$ with $u \cdot a$. It is easy to show that $u \cdot a$ satisfies the central limit theorem. Namely, if u_n is a random vector uniformly distributed on the unit sphere S^{n-1} and $\{a_n\}$ is a sequence of unit vectors with $a_n \in S^{n-1}$, then

$$\sqrt{n} u_n \cdot a_n \longrightarrow N(0, 1)$$

in distribution as $n \to \infty$. We refer the reader to [41] and [79, Proposition 25] for details and a proof of this statement. The following is a consequence of [79, Theorem 13].

Theorem 3.2 (Theorem 13 from [79]). *Let ξ, ζ be random variables from the class $\mathcal{M}_4$, and assume ξ is a symmetric random variable. For each $n \geq 1$, let W_n be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Let $\{a_n\}$ be a sequence of unit vectors with $a_n \in S^{n-1}$ such that $\lim_{n \to \infty} \|a_n\|_{\ell^{\infty}} = 0$, and let $\{i_n\}$ be a sequence of indices with $i_n \in [n]$. Then*

$$\sqrt{n} v_{i_n}(W_n) \cdot a_n \longrightarrow N(0, 1)$$

in distribution as $n \to \infty$.

In a recent paper [15], it was showed that one can remove the assumption that ξ, ζ are from the class $\mathcal{M}_4$, with an extra restriction on the index i_n.

Theorem 3.3 *(Theorem 1.2 from [15]). Let ξ, ζ be sub-exponential random variables with mean zero, and assume ξ has unit variance. For each $n \geq 1$, let W_n be an $n \times n$ Wigner matrix with atom variables ξ, ζ. In addition, let $\{a_n\}$ be a sequence of unit vectors with $a_n \in S^{n-1}$. Then there exists $\tilde{\delta} > 0$ such that, for any fixed integer $m \geq 1$ and any*

$$I_n \subset T_n := \left([1, n^{1/4}] \cup [n^{1-\tilde{\delta}}, n - n^{1-\tilde{\delta}}] \cup [n - n^{1/4}, n] \right) \cap \mathbb{N}$$

with $|I_n| = m$,

$$\sqrt{n}(|a_n \cdot v_i(W_n)|)_{i \in I_n} \longrightarrow (|Z_i|)_{i=1}^m$$

in distribution as $n \to \infty$, where $Z_1, \ldots, Z_m$ are iid standard normal random variables.

One immediately obtains the following corollary

Corollary 3.4 *(Corollary 1.3 from [15]). Let W_n, $\tilde{\delta}$, and T_n be as in Theorem 3.3. For each $n \geq 1$, let $i_n \in T_n$, and let $v_{i_n}^{(n)}(j)$ be the jth coordinate of $v_{i_n}(W_n)$. Assume l is a fixed positive integer. Then, for any $J_n \subset [n]$ with $|J_n| = l$,*

$$\sqrt{n} \left(\left| v_{i_n}^{(n)}(j) \right| \right)_{j \in J_n} \longrightarrow (|Z_j|)_{j=1}^l$$

in distribution as $n \to \infty$, where $Z_1, \ldots, Z_l$ are iid standard normal random variables.

Remark 3.5. Both Theorem 3.3 and Corollary 3.4 hold for more general ensembles of random matrices than Wigner matrices. Indeed, the results in [15] hold for so-called generalized Wigner matrices, where the entries above the diagonal are only required to be independent, not necessarily identically distributed; see [15, Definition 1.1] for details.

4. Extremal coordinates

In this section, we investigate the largest and smallest coordinates (in absolute value) of an eigenvector of a Wigner matrix.

4.1. The largest coordinate

In view of Theorem 2.1, it is natural to conjecture that, for any eigenvector v_j,

$$\|v_j\|_{\ell^\infty} = O \left(\sqrt{\frac{\log n}{n}} \right) \tag{11}$$

with high probability. The first breakthrough was made in [28, Theorem 5.1], which provides a bound of the form $\frac{\log^{O(1)} n}{\sqrt{n}}$ for a large proportion of eigenvectors v_j, under some technical conditions on the distribution of the entries. This result was extended to all eigenvectors under a weaker assumption in [75,76], and many newer papers [23, 37,27,28,30,31,33,26,32,34,35,80,86] give further strengthening and generalizations. In particular, the optimal bound in (11) was recently obtained in [86].

Theorem 4.1 *(Optimal upper bound; Theorem 6.1 from [86]). Let ξ be a sub-gaussian random variable with mean zero and unit variance. Then, for any $C_1 > 0$ and any $0 < \varepsilon < 1$, there exists a constant $C_2 > 0$ such that the following holds. Let W be an $n \times n$ Wigner matrix with atom variable ξ.*

- *(Bulk case) For any $\varepsilon n \le i \le (1 - \varepsilon)n$,*

$$\|v_i(W)\|_{\ell^\infty} \le C_2 \sqrt{\frac{\log n}{n}}$$

  with probability at least $1 - n^{-C_1}$.
- *(Edge case) For $1 \le i \le \varepsilon n$ or $(1 - \varepsilon)n \le i \le n$,*

$$\|v_i(W)\|_{\ell^\infty} \le C_2 \frac{\log n}{\sqrt{n}}$$

  with probability at least $1 - n^{-C_1}$.

Remark 4.2. Theorem 4.1 was proved in [86] under the stronger assumption that ξ is K-bounded for some constant K. One can easily obtain the more general result here by observing that [86, Lemma 1.2] holds under the sub-gaussian assumption, as a special case of a recent result, [65, Theorem 2.1]. The rest of the proof remains unchanged.

Using Theorem 3.1, one can derive a matching lower bound.

Theorem 4.3 *(Matching lower bound). Let ξ, ζ be random variables from the class $\mathcal{M}_4$. Then there exists a constant $c > 0$ such that the following holds. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. For any $1 \le i \le n$,*

$$\|v_i(W)\|_{\ell^\infty} \ge c \sqrt{\frac{\log n}{n}}$$

with probability $1 - o(1)$.

Open question. Prove the optimal bound in (11) for all eigenvectors (including the edge case).

Open question. Is the limiting distribution of $\|v_i\|_{\ell^\infty}$ universal, or does it depend on the atom variables ξ and ζ?

One can deduce a slightly-weaker version of Theorem 4.1 when the entries are only assumed to be sub-exponential with exponent $0 < \alpha < 2$ by applying [86, Theorem 6.1] and a truncation argument. We apply such an argument in Section 12 to obtain the following corollary.

Corollary 4.4. *Let ξ be a symmetric sub-exponential random variable with exponent $\alpha > 0$. Assume further that ξ has unit variance. Then, for any C_1 and $0 < \varepsilon < 1$, there is a constant C_2 such that the following holds. Let W be an $n \times n$ Wigner matrix with atom variable ξ.*

- *(Bulk case) For any $\varepsilon n \le i \le (1 - \varepsilon)n$,*

$$\|v_i(W)\|_{\ell^\infty} \le C_2 \sqrt{\frac{\log^{1+2/\alpha} n}{n}}$$

  with probability at least $1 - n^{-C_1}$.
- *(Edge case) For $1 \le i \le \varepsilon n$ or $(1 - \varepsilon)n \le i \le n$,*

$$\|v_i(W)\|_{\ell^\infty} \le C_2 \frac{\log^{1+2/\alpha} n}{\sqrt{n}}$$

  with probability at least $1 - n^{-C_1}$.

Corollary 4.4 falls short of the optimal bound in (11); it remains an open question to obtain the optimal bound when the entries are not sub-gaussian.

Next, we discuss a generalization. Notice that $\|v\|_{\ell^\infty} = \max_{1 \le i \le n} |v \cdot e_i|$, where $e_1, \ldots, e_n$ are the standard basis vectors. What happens if we consider the inner product $v \cdot u$, for any fixed unit vector u? The theorems in the previous section show that, under certain technical assumptions, $\sqrt{n} v \cdot u$ is approximately Gaussian, which implies that $|v \cdot u|$ is typically of order $n^{-1/2}$. The following result gives a strong deviation bound.

Theorem 4.5 *(Isotropic delocalization, Theorem 2.16 from [12]). Let ξ and ζ be zero-mean sub-exponential random variables, and assume that ξ has unit variance. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Then, for any $C_1 > 0$ and $0 < \varepsilon < 1/2$, there exists $C_2 > 0$ (depending only on C_1, ε, ξ, and ζ) such that*

$$\sup_{1 \le i \le n} |v_i \cdot u| \le \frac{n^\varepsilon}{\sqrt{n}},$$

for any fixed unit vector $u \in S^{n-1}$, with probability at least $1 - C_2 n^{-C_1}$.

Remark 4.6. Theorem 4.5 and the results in [12] actually hold for a larger class of so-called generalized Wigner matrices whose entries have bounded moments; see [12, Section 2.2] for details.

4.2. The smallest coordinate

We now turn our attention to the smallest coordinate of a given eigenvector. To this end, we recall the definition of $\|\cdot\|_{\min}$ given in (3).

Theorem 4.7 (*Individual coordinates: Lower bound*). *Let ξ, ζ be sub-gaussian random variables with mean zero, and assume ξ has unit variance. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Let $v_1, \ldots, v_n$ denote the unit eigenvectors of W, and let $v_i(j)$ denote the jth coordinate of v_i. Then there exist constants $C, c, c_0, c_1 > 0$ (depending only on ξ, ζ) such that, for any $n^{-c_0} < \alpha < c_0$ and $\delta \geq n^{-c_0/\alpha}$,*

$$\sup_{1 \leq i,j \leq n} \mathbb{P}\left(|v_i(j)| \leq \frac{\delta}{\sqrt{n}(\log n)^{c_1}} \right) \leq C \frac{\delta}{\sqrt{\alpha}} + C \exp(-c \log^2 n).$$

By the union bound, we immediately obtain the following corollary.

Corollary 4.8. *Let ξ, ζ be sub-gaussian random variables with mean zero, and assume ξ has unit variance. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Let $v_1, \ldots, v_n$ denote the unit eigenvectors of W, and let $v_i(j)$ denote the jth coordinate of v_i. Then there exist constants $C, c, c_0, c_1 > 0$ (depending only on ξ, ζ) such that, for any $n^{-c_0} < \alpha < c_0$ and $\delta \geq n^{-c_0/\alpha}$,*

$$\sup_{1 \leq i \leq n} \mathbb{P}\left(\|v_i(W)\|_{\min} \leq \frac{\delta}{n^{3/2}(\log n)^{c_1}} \right) \leq C \frac{\delta}{\sqrt{\alpha}} + C \exp(-c \log^2 n).$$

In particular, Corollary 4.8 implies that, with high probability,

$$\|v_i(W)\|_{\min} = \Omega\left(\frac{1}{n^{3/2}(\log n)^{c_1}} \right).$$

In view of Theorem 2.1, this is optimal up to logarithmic factors.

Open question. Is the limiting distribution of $\|v_j(W)\|_{\min}$ universal, or does it depend on the atom variables ξ and ζ?

We prove Corollary 4.4 and Theorem 4.7 in Section 12.

5. No-gaps delocalization

The results in the previous section address how much mass can be contained in a single coordinate. We next turn to similar estimates for the amount of mass contained

on a number of coordinates. In particular, the following results assert that any subset of coordinates of linear size must contain a non-negligible fraction of the vector's ℓ^2-norm. Following Rudelson and Vershynin [66], we refer to this phenomenon as *no-gaps delocalization*.

Using Corollary 3.4, we obtain the following analogue of Theorem 2.7. Recall the function H defined in (7).

Theorem 5.1. *Let W_n, $\tilde{\delta}$, and T_n be as in Theorem 3.3. For each n, let $k_n \in T_n$. Let v_{k_n} be a unit eigenvector of W_n corresponding to $\lambda_{k_n}(Y_n)$. Then, for any fixed $0 < \delta < 1$,*

$$\max_{S \subset [n]:|S|=\lfloor \delta n \rfloor} \|v_{k_n}\|_S^2 \longrightarrow -\int_0^\delta H(u)\,du$$

and

$$\min_{S \subset [n]:|S|=\lfloor \delta n \rfloor} \|v_{k_n}\|_S^2 \longrightarrow -\int_{1-\delta}^1 H(u)\,du$$

in probability as $n \to \infty$, where H is defined in (7).

We prove Theorem 5.1 in Section 13. The integrals in the right-hand side involving H can be expressed in a number of different ways; see Remarks 2.8 and 2.9 for details. In particular, Remark 2.9 shows that, for δ sufficiently small, the first integral is $\Theta(\delta \log(\delta^{-1}))$ and the second is $\Theta(\delta^3)$.

While this survey was written, Rudelson and Vershynin posted the following theorem, which addresses the lower bound for more general models of random matrices. Their proof is more involved and very different from that of Theorem 5.1.

Theorem 5.2 *(Theorem 1.5 from [66]).* *Let ξ be a real random variable which satisfies*

$$\sup_{u \in \mathbb{R}} \mathbb{P}(|\xi - u| \leq 1) \leq 1 - p \quad and \quad \mathbb{P}(|\xi| > K) \leq p/2$$

for some $K, p > 0$. Let W be an $n \times n$ Wigner matrix with atom variable ξ. Let $\kappa \geq 1$ be such that the event $\mathcal{B}_\kappa := \{\|W\| \leq \kappa\sqrt{n}\}$ holds with probability at least $1/2$. Let $\delta \geq 1/n$ and $t \geq c_1 \delta^{-7/6} n^{-1/6} + e^{-c_2/\sqrt{\delta}}$. Then, conditionally on $\mathcal{B}_\kappa$, the following holds with probability at least $1 - (c_3 t)^{\delta n}$: every eigenvector v of W satisfies

$$\min_{S \subset [n]:|S| \geq \delta n} \|v\|_S \geq (\delta t)^6.$$

Here, the constants $c_1, c_2, c_3 > 0$ depend on p, K, and κ.

Remark 5.3. The results in [66] hold for even more general ensembles of random matrices than what is stated above; see [66, Assumption 1.1] for details. For many atom variables ξ, ζ, the spectral norm $\|W\|$ is strongly concentrated (see Lemma 11.3 or [74, Corollary 2.3.6] for examples), so the event $\mathcal{B}_\kappa$ holds with high probability, provided κ is sufficiently large.

Compared to Theorem 5.1, the estimate in Theorem 5.2 is not optimal. On the other hand, the probability bound in Theorem 5.2 is stronger and thus the estimate is more applicable. It would be desirable to have a common strengthening of these two theorems. This can be done by achieving an extension of Theorem 2.10.

Open question. Extend Theorem 2.10 to eigenvectors of Wigner matrices with non-gaussian entries.

We conclude this section with the following corollary, which is comparable to Theorem 2.2.

Corollary 5.4 (*ℓ^p-norm*). *Let ξ, ζ be real sub-gaussian random variables with mean zero, and assume ξ has unit variance. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Then, for any $1 \le p \le 2$, there exist constants $C, c, C_0, c_0 > 0$ (depending only on p and the sub-gaussian moments of ξ and ζ) such that*

$$c_0 n^{1/p-1/2} \le \min_{1 \le j \le n} \|v_j(W)\|_{\ell^p} \le \max_{1 \le j \le n} \|v_j(W)\|_{\ell^p} \le C_0 n^{1/p-1/2}$$

with probability at least $1 - C\exp(-cn)$.

One-sided bounds for $\|v_j(W)\|_{\ell^p}$ were previously obtained in [28]. We give a proof of Corollary 5.4 in Section 13.

6. Random symmetric matrices with non-zero mean

In this section, we consider random symmetric matrices with nonzero mean, which includes the adjacency matrices $A_n(p)$ and $\tilde{A}_n(p)$. In view of Definition 1.1, we do not refer to these matrices as Wigner matrices. However, such random symmetric matrices can be written as deterministic perturbations of Wigner matrices, and we will often take advantage of this fact.

It has been observed that the unit eigenvector corresponding to the largest eigenvalue of the adjacency matrix $A_n(p)$ looks like the normalized all-ones vector $\frac{1}{\sqrt{n}}\mathbf{1}_n$, since the degree of the vertices are approximately the same (assuming $p \gg \log n/n$). For the rest of the spectrum, we expect the corresponding eigenvectors to be uniformly distributed on the unit sphere in $\mathbb{R}^n$.

6.1. The largest eigenvector

The following result describes the unit eigenvector corresponding to the largest eigenvalue of $A_n(p)$

Theorem 6.1 *(Theorem 1 from [51]). Let $A_n(p)$ be the adjacency matrix of the random graph $G(n, p)$. Let v_n be a unit eigenvector corresponding to the largest eigenvalue of $A_n(p)$. For $p \geq \frac{\log^6 n}{n}$,*

$$\left\| v_n - \frac{1}{\sqrt{n}} \mathbf{1}_n \right\|_{\ell^\infty} \leq C \frac{\log n}{\log(np)} \frac{1}{\sqrt{n}} \sqrt{\frac{\log n}{np}}$$

with probability $1 - o(1)$, for some constant $C > 0$.

6.2. Extremal coordinates

Recall that $\tilde{G}(n, p)$ is the Erdös–Rényi random graph with loops on n vertices and edge density p. We have the following analogue of Theorem 4.1.

Theorem 6.2 *(Theorem 2.16 from [37]). Let $\tilde{A}_n(p)$ be the adjacency matrix of the random graph $\tilde{G}(n, p)$, and let $\tilde{v}_i$ be the unit eigenvector corresponding to the ith smallest eigenvalue. Fix $1 + \varepsilon_0 \leq \alpha \leq C_0 \log \log n$ for some constants $\varepsilon_0, C_0 > 0$. Assume $c' \geq p \geq (\log n)^{6\alpha}/n$ for some constant c'. Then there exist constants $C, c > 0$ (depending on ε_0, C_0, and c') such that*

$$\max_{1 \leq i \leq n-1} \| \tilde{v}_i \|_{\ell^\infty} \leq \frac{(\log n)^{4\alpha}}{\sqrt{n}}$$

and

$$\left\| \tilde{v}_n - \frac{1}{\sqrt{n}} \mathbf{1}_n \right\|_{\ell^\infty} \leq C \frac{(\log n)^\alpha}{\sqrt{n}} \sqrt{\frac{1}{np}}$$

with probability at least $1 - C \exp(-c(\log n)^\alpha)$.

An upper bound for the ℓ^∞-norm of the form $\frac{\log^C n}{\sqrt{n}}$ was originally proven in [82] for fixed values of p. Theorem 6.2 above is an extension which applies to a wider range of values of p. More generally, the results in [37] also apply to perturbed Wigner matrices.

Theorem 6.3 *(Theorem 2.16 from [37]). Let ξ be a sub-exponential random variable with mean zero and unit variance. Let W be the $n \times n$ Wigner matrix with atom variable ξ. Fix $\mu \in \mathbb{R}$, and consider the matrix $M := W + \mu \mathbf{J}$, where $\mathbf{J}$ is the all-ones matrix. Fix $1 + \varepsilon_0 \leq \alpha \leq C_0 \log \log n$ for some constants $\varepsilon_0, C_0 > 0$. Then there exist $C, c > 0$ (depending on ε_0, C_0, ξ, and μ) such that*

$$\max_{1 \le i \le n-1} \|v_i(M)\|_{\ell^\infty} \le \frac{(\log n)^{4\alpha}}{\sqrt{n}}$$

and

$$\left\| v_n(M) - \frac{1}{\sqrt{n}} \mathbf{1}_n \right\|_{\ell^\infty} \le C \frac{(\log n)^\alpha}{n}$$

with probability at least $1 - C \exp(-c(\log n)^\alpha)$.

We next consider the smallest coordinates of each eigenvector of the adjacency matrix $A_n(p)$. We prove the following analogue of Theorem 4.7

Theorem 6.4 (*Individual coordinates: Lower bound*). *Let $A_n(p)$ be the adjacency matrix of the random graph $G(n, p)$ for some fixed value of $p \in (0, 1)$. Let $v_1, \ldots, v_n$ be the unit eigenvectors of $A_n(p)$, and let $v_i(j)$ denote the jth coordinate of v_i. Then, for any $\alpha > 0$, there exist constants $C, c_1 > 0$ (depending on p, α) such that, for any $\delta > n^{-\alpha}$,*

$$\sup_{1 \le i \le n-1} \sup_{1 \le j \le n} \mathbb{P}\left(|v_i(j)| \le \frac{\delta}{\sqrt{n}(\log n)^{c_1}} \right) \le Cn^{o(1)}\delta + o(1).$$

Here, the rate of convergence to zero implicit in the $o(1)$ terms depends on p, α.

Notice that Theorem 6.4 does not address the eigenvector corresponding to the largest eigenvalue of $A_n(p)$. Bounds for this eigenvector can easily be obtained using Theorem 6.1 and are left as an exercise.

6.3. No-gaps delocalization

In [21], the following result is developed as a tool in the authors' study of nodal domain for the eigenvectors of $G(n, p)$.

Theorem 6.5 (*Theorem 3.1 from [21]*). *Let $A_n(p)$ be the adjacency matrix of the random graph $G(n, p)$, and let v_i be the unit eigenvector corresponding to the ith smallest eigenvalue. For every $p \in (0, 1)$ and every $\varepsilon > 0$, there exist $C, c, \eta > 0$ (depending on ε and p) such that, for every fixed subset $S \subset [n]$ of size $|S| \ge (\frac{1}{2} + \varepsilon)n$,*

$$\min_{1 \le i \le n-1} \|v_i\|_S \ge \eta$$

with probability at least $1 - C \exp(-cn)$.

Theorem 6.5 is not nearly as strong as Theorem 5.2 in Section 5, which holds for all subsets $S \subset [n]$ of specified size. However, the results from [66] do not apply directly to the adjacency matrix $A_n(p)$. In particular, Theorem 5.2 only applies when the spectral

norm of the matrix is $O(\sqrt{n})$ with high probability. However, the largest eigenvalue of $A_n(p)$ is of order $\Theta(n)$.

We conclude this section by discussing some variations which do not require the spectral norm to be $O(\sqrt{n})$. Specifically, we consider eigenvectors of matrices of the form $M := W + J$, where W is a Wigner matrix and J is a real symmetric deterministic low-rank matrix. When W is drawn from the GOE, it suffices to consider the case that J is diagonal (since W is invariant under conjugation by orthogonal matrices). Thus, we begin with the case when J is diagonal.

Theorem 6.6 *(Diagonal low-rank perturbations). Let ξ, ζ be real sub-gaussian random variables with mean zero, and assume ξ has unit variance. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ; let J be $n \times n$ deterministic, diagonal real symmetric matrix with rank k. Let $M := W + J$, and consider $\lambda_j(M) \in [\lambda_1(W), \lambda_n(W)]$ and its corresponding unit eigenvector $v_j(M)$. For any constant $\delta \in (0,1)$ and any fixed set $S \subset [n]$ with size $|S| = \lfloor \delta n \rfloor$, there exist constants $C, c > 0$ and $0 < \eta_1 \leq \eta_2 < 1$ (depending only on δ, k, and the sub-gaussian moments of ξ and ζ) such that*

$$\eta_2 \geq \|v_j(M)\|_S \geq \eta_1$$

with probability at least $1 - C \exp\left(-c(\log n)^{c \log \log n}\right)$.

Theorem 6.6 does not handle the eigenvectors corresponding to the extreme eigenvalues of M due to the condition $\lambda_j(M) \in [\lambda_1(W), \lambda_n(W)]$. In particular, it is possible for the eigenvectors corresponding to the largest and smallest eigenvalues to be *localized* for certain choices of the diagonal matrix J. A localized eigenvector is an eigenvector whose mass in concentrated on only a few coordinates. We present an example of this phenomenon in the following theorem.

Theorem 6.7. *Let ξ, ζ be real sub-gaussian random variables with mean zero and unit variance. Then there exist constants $C, c > 0$ such that the following holds. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Let $J = \sqrt{n}\theta e_k e_k^{\mathrm{T}}$ for some $\theta > C$ and $1 \leq k \leq n$, where $e_1, \ldots, e_n$ is the standard basis of $\mathbb{R}^n$. Set $M := W + J$. Then*

$$|v_n(M) \cdot e_k|^2 \geq 1 - \frac{C^2}{\theta^2}$$

with probability at least $1 - C \exp(-cn)$.

Specifically, Theorem 6.7 implies that for θ large enough (say, $\theta = \Omega(\log n)$), most of the mass of the eigenvector $v_n(M)$ is concentrated on the kth coordinate. This is in contrast to Theorem 6.6, which shows that, with high probability, the other eigenvectors cannot concentrate on a single coordinate. Theorem 6.7 is part of a large collection of results concerning the extreme eigenvalues and eigenvectors of perturbed random

matrices. We refer the reader to [6,9,46,56,59,61] and references therein for many other results.

Theorem 6.7 follows as a simple corollary of a slightly-modified version of the Davis–Kahan $\sin \theta$ Theorem (see, for instance, [56, Theorem 4]) and a bound on the spectral norm of a Wigner matrix (Lemma 11.3); we leave the details as an exercise.

Theorems 6.6 and 6.7 both deal with diagonal perturbations. For more general perturbations, we have the following result.

Theorem 6.8 *(General low-rank perturbations). Let ξ, ζ be real sub-gaussian random variables with mean zero, and assume ξ has unit variance. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ; let J be $n \times n$ deterministic real symmetric matrix with rank k. Let $\varepsilon_1, \varepsilon_0 > 0$. Let $M := W + J$, and consider $\lambda_j(M)$ such that $\varepsilon_1 n \leq j \leq (1 - \varepsilon_1)n$ and its corresponding unit eigenvector $v_j(M)$. For any constant $\delta \in (0, 1)$ and any fixed set $S \subset [n]$ with size $|S| = \lfloor \delta n \rfloor$, there exist constants $C, c > 0$ and $0 < \eta < 1$ (depending only on δ, k, ε_0, ε_1, and the sub-gaussian moments of ξ and ζ) such that*

$$\mathbb{P} \left(\frac{1}{n^{1-\varepsilon_0}} \leq \|v_j(M)\|_S \leq \eta \right) \leq C \exp \left(-c(\log n)^{c \log \log n} \right).$$

Remark 6.9. Theorem 6.8 shows that, with high probability, either $\|v_j(M)\|_S \geq \eta$ or $\|v_j(M)\|_S \leq \frac{1}{n^{1-\varepsilon_0}}$. Based on the previous results, we do not expect the later case to be a likely event. However, it appears additional structural information about J (for instance, that J is diagonal as in Theorem 6.6) is required to eliminate this possibility. See the proof of Theorem 6.8 for additional details.

In the case when J has rank one, we have the following.

Theorem 6.10 *(Rank one perturbations). Let ξ, ζ be real sub-gaussian random variables with mean zero and unit variance, and fix $\varepsilon_1 > 0$. Then there exist constants $C, c > 0$ such that the following holds. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Suppose $J = \theta u u^{\mathrm{T}}$, where $\theta \in \mathbb{R}$ and $u \in \mathbb{R}^n$ is a unit vector. Set $M := W + J$. Then, for any integer j with $\varepsilon_1 n \leq j \leq (1 - \varepsilon_1)n$,*

$$|v_j(M) \cdot u| \leq \frac{C(\log n)^{c \log \log n}}{|\theta|}$$

with probability $1 - o(1)$.

When $\theta := \mu n$ and $u := n^{-1/2} \mathbf{1}_n$, J becomes the matrix in which every entry takes the value μ. In this case, the entries of $W + J$ have mean μ instead of mean zero. Thus, applying Theorem 6.10, we immediately obtain the following corollary.

Corollary 6.11 *(Wigner matrices with non-zero mean). Let ξ, ζ be real sub-gaussian random variables with mean zero and unit variance, and fix $\varepsilon_1 > 0$. Then there exist*

constants $C, c > 0$ such that the following holds. Let $\mu \in \mathbb{R}$ with $\mu \neq 0$. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Set $M := W + \mu \mathbf{J}$, where $\mathbf{J}$ is the all-ones matrix. Then, for any integer j with $\varepsilon_1 n \leq j \leq (1 - \varepsilon_1)n$,

$$|v_j(M) \cdot \mathbf{1}_n| \leq \frac{C(\log n)^{c \log \log n}}{|\mu| \sqrt{n}}$$

with probability $1 - o(1)$.

We prove Theorems 6.4, 6.6, 6.8 and 6.10 in Section 14.

7. Localized eigenvectors: Heavy-tailed and band random matrices

As we saw in Theorem 6.7, the eigenvectors corresponding to the extreme eigenvalues of a perturbed Wigner matrix can be *localized*, meaning that most of the mass is contained on only a few coordinates. For instance, in Theorem 6.7, most of the mass was contained on a single coordinate. We now discuss a similar phenomenon for heavy-tailed and band random matrices.

7.1. Heavy-tailed random matrices

Most of the results from the previous sections required the atom variables ξ, ζ to be sub-exponential or sub-gaussian. In particular, these conditions imply that ξ and ζ have finite moments of all orders. In other words, the atom variables have very light tails. We now consider the case when the atom variables have heavy tails, such as when ξ and ζ have only one or even zero finite moments. In this case, the eigenvectors corresponding to the largest eigenvalues behave very differently than predicted by the results above.

Theorem 7.1 *(Theorem 1.1 from [10]). Let ξ be a real random variable satisfying*

$$\mathbb{P}(|\xi| \geq x) = L(x) x^{-\alpha}$$

for all $x > 0$, where $0 < \alpha < 2$ and $L : (0, \infty) \to (0, \infty)$ is a slowly varying function, i.e., for all $t > 0$,

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1.$$

For each $n \geq 1$, let W_n be an $n \times n$ Wigner matrix with atom variable ξ. Fix an integer $k \geq 0$. Then, for every fixed $\varepsilon > 0$,

$$\|v_{n-k}(W_n)\|_{\ell^\infty} \geq \frac{1}{\sqrt{2}} - \varepsilon$$

with probability $1 - o(1)$. *In addition,*

$$\min_{S \subset [n]:|S|=n-2} \|v_{n-k}(W_n)\|_S \longrightarrow 0$$

in probability as $n \to \infty$.

Remark 7.2. Theorem 7.1 also holds when $2 \le \alpha < 4$ provided the atom variable ξ is symmetric; see [10, Theorem 1.1] for details.

Theorem 7.1 shows that the eigenvectors corresponding to the largest eigenvalues of W_n are concentrated on at most two coordinates. This is considerably different than the cases discussed above when ξ has light tails.

Let us try to explain this phenomenon based on the tail behavior of ξ. It is well known that the largest entry of an $n \times n$ Wigner matrix with sub-gaussian entries is $O(\sqrt{\log n})$ with high probability. However, when the tails are heavy, the maximum entry of W_n can be significantly larger. It was observed by Soshnikov [69,70] that, in this case, the largest eigenvalues behave like the largest entries of the matrix. Intuitively, the eigenvector corresponding to the largest eigenvalue of W_n should localize on the coordinates which match the largest entry. Since W_n is symmetric, the largest entry can appear at most twice. Hence, we expect this eigenvector to be concentrated on at most two coordinates. This heuristic has led to a number of results regarding the eigenvalues and eigenvectors of heavy-tailed Wigner matrices; we refer the reader to [4,8,10,13,14,18,69,70] and references therein for further details and additional results.

7.2. Random band matrices

The standard basis elements $e_1, \ldots, e_n$ of $\mathbb{R}^n$ are always eigenvectors of an $n \times n$ diagonal matrix. In other words, the eigenvectors of a diagonal matrix are localized. Band matrices generalize diagonal matrices by only allowing the entries on and near the diagonal to be non-zero while requiring the other entries, away from the diagonal, to be zero.

We can form random band matrices from Wigner matrices. Indeed, let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. For simplicity, let us assume that ξ and ζ are sub-gaussian random variables. We can form an $n \times n$ random band matrix T from W with band width $L \ge 1$ by replacing the (i,j)-entry of W by zero if and only if $|i-j| \ge L$. Hence, the (i,j)-entry of T is just the (i,j)-entry of W when $|i-j| < L$. A random band matrix with width n is a Wigner matrix, and a random band matrix with band width 1 is a diagonal matrix. Thus, we expect a transition in the eigenvector behavior when the band width L interpolates between 1 and n. Indeed, it is conjectured that for L significantly smaller than $\sqrt{n}$, the eigenvectors will be localized (with localization length on the order of L^2). On the other hand, for L sufficiently larger than $\sqrt{n}$, it is expected

that the eigenvectors of T behave more like the eigenvectors of W. Some partial results in this direction have been established in [24,25,38,67].

While random band matrices can be constructed from Wigner matrices, they are, in general, not Wigner matrices, and we will not focus on them here. We refer the interested reader to [10,24,25,38,67,71] and references therein for results concerning the spectral properties of random band matrices. In the discussion above, we have focused on the case when the atom variables ξ and ζ are sub-gaussian. However, Theorem 7.1 can be extended to random band matrices constructed from heavy-tailed Wigner matrices; see [10] for details.

8. Singular vectors and eigenvectors of non-Hermitian matrices

In this section, we consider the singular vectors and eigenvectors of non-Hermitian random matrices.

Let M be a $p \times n$ matrix with real entries. Recall that the singular values of M are the square roots of the eigenvalues of MM^{T}. The left singular vectors are the eigenvectors of MM^{T}, and the right singular vectors are the eigenvectors of $M^{\mathrm{T}}M$. Following our previously introduced notation, we will write

$$\sqrt{\lambda_1(MM^{\mathrm{T}})}, \ldots, \sqrt{\lambda_p(MM^{\mathrm{T}})}$$

to denote the singular values and

$$v_1(MM^{\mathrm{T}}), \ldots, v_p(MM^{\mathrm{T}})$$

to denote the left singular vectors of M.

Let $M = M_{p,n} = (\zeta_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ be a random matrix (more specifically, a sequence of random matrices) whose entries are independent real random variables with mean zero and unit variance. Assume $\lim_{n \to \infty} p/n = y$ for some $y \in (0, 1]$. The delocalization properties of the singular vectors of $M_{p,n}$ have been explored in [17,58,78,86,87]. The optimal bound of $O(\sqrt{\log n/n})$ for the ℓ^∞-norm was obtained recently in [86].

Theorem 8.1 (*Delocalization of singular vectors, Theorem B.3 from [86]*). *Let ζ be a sub-gaussian random variable with mean zero and unit variance. Then, for any $C_1 > 0$ and any $0 < \varepsilon < 1$, there exists a constant $C_2 > 0$ such that the following holds. Assume the entries of $M_{p,n} = (\zeta_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ are iid copies of ζ. Let $a := (1 - \sqrt{y})^2$ and $b := (1 + \sqrt{y})^2$.*

- (*Bulk case*) *For any $1 \leq i \leq n$ such that $\frac{1}{n}\lambda_i(M_{p,n}M_{p,n}^{\mathrm{T}}) \in [a + \varepsilon, b - \varepsilon]$, there is a corresponding left singular vector $v_i(M_{p,n}M_{p,n}^{\mathrm{T}})$ such that*

$$\|v_i(M_{p,n}M_{p,n}^{\mathrm{T}})\|_{\ell^\infty} \leq C_2 \sqrt{\frac{\log n}{n}}$$

with probability at least $1 - n^{-C_1}$. The same also holds for right singular vectors.

- *(Edge case) For any $1 \le i \le n$ such that $\frac{1}{n}\lambda_i(M_{p,n}M_{p,n}^{\mathrm{T}}) \in [a-\varepsilon, a+\varepsilon] \cup [b-\varepsilon, b+\varepsilon]$ if $a \ne 0$ and $\frac{1}{n}\lambda_i(M_{p,n}M_{p,n}^{\mathrm{T}}) \in [4-\varepsilon, 4]$ if $a = 0$, there is a corresponding left singular vector $v_i(M_{n,p}M_{n,p}^{\mathrm{T}})$ such that*

$$\|v_i(M_{n,p}M_{n,p}^{\mathrm{T}})\|_{\ell^\infty} \le C_2 \frac{\log n}{\sqrt{n}}$$

with probability at least $1 - n^{-C_1}$. The same also holds for right singular vectors.

Similar to Theorem 4.1 this was first proved under the stronger assumption that the entries of the matrix are bounded, but one can obtain this version using the same argument as in Remark 4.2. The analogue of Theorem 8.1 for the eigenvectors of $M_{n,n}$ was recently proved in [64], using a completely different method.

Theorem 8.2 *(Delocalization of eigenvectors; Theorem 1.1 from [64]). Let $M = (\zeta_{ij})_{1 \le i,j \le n}$ be an $n \times n$ random matrix whose entries are independent real random variables with mean zero, variance at least one, and*

$$\sup_{p \ge 1} p^{-1/2} (\mathbb{E}|\zeta_{ij}|^p)^{1/p} \le K$$

for all $1 \le i,j \le n$. Then, for any $t \ge 2$, with probability at least $1 - n^{1-t}$, all unit eigenvectors v of M satisfy

$$\|v\|_{\ell^\infty} \le \frac{Ct^{3/2}\log^{9/2} n}{\sqrt{n}},$$

where C depends only on K.

Remark 8.3. The above result holds for more general matrix ensembles, e.g. random matrices with independent sub-exponential entries; see [64, Corollary 1.5] for details.

9. Random regular graphs

We now turn to the eigenvectors of the adjacency matrix of random regular graphs. Recall that a regular graph is a simple graph where each vertex has the same degree, and a d-regular graph is a regular graph with vertices of degree d. It is well-known that a d-regular graph on n vertices exists if and only if $n \ge d + 1$ and nd is even. Let $G_{n,d}$ denote a random d-regular graph chosen uniformly from all d-regular graphs on n vertices. It is easy to see that the adjacency matrix of $G_{n,d}$ has a trivial eigenvector $\frac{1}{\sqrt{n}}\mathbf{1}_n$ corresponding to the eigenvalue d. Further, it has been conjectured that every non-trivial unit eigenvector behaves like a uniform vector on the unit sphere.

In the combinatorics/computer science literature, the most interesting case to consider is when d is a constant. This also seems to be the most difficult case to study. In this case, the strongest delocalization result known to the authors is the following, which is a corollary of [16, Theorem 1].

Theorem 9.1 *(Theorem 1 from [16]). Let d be fixed and $\varepsilon > 0$. Then there is a constant $\delta > 0$ (depending on d and ε) such that the following holds. With probability $1 - o(1)$, for any unit eigenvector v of the adjacency matrix of $G_{n,d}$, any subset $S \subset [n]$ satisfying*

$$\|v\|_S^2 > \varepsilon$$

must be of size $|S| \geq n^\delta$.

Another form of delocalization result for fixed d is considered in [1]. From [22], one has the following result for the eigenvectors in the case when d grows slowly with the vertex size n.

Theorem 9.2 *(Theorem 3 from [22]). Fix $\delta > 0$. Let $d := d_n = (\log n)^\gamma$ for $\gamma > 0$. Let $\eta_n := (r_n - r_n^{-1})/2$ where $r_n := \exp(d_n^{-\alpha})$ for some $0 < \alpha < \min(1, \gamma^{-1})$. Let $T_n \subset [n]$ be a deterministic set of size $L_n = o(\eta_n^{-1})$. Let Ω_n be the event that some unit eigenvector v of the adjacency matrix of $G_{n,d}$ satisfies $\|v\|_{T_n}^2 \geq 1 - \delta$. Then, for all sufficiently large n,*

$$\mathbb{P}(\Omega_n^c) \geq e^{-L_n \eta_n / d_n} \left(1 - o(d_n^{-1})\right) = 1 - o(d_n^{-1}).$$

Finally, let us mention the following recent result from [7], which provides a near optimal bound when d grows sufficiently fast with n.

Theorem 9.3 *(Corollary 1.2 from [7]). There exist constants $C, C' > 0$ such that the following holds. Let $C^{-1} \log^4 n \leq d \leq C n^{2/3} \log^{-4/3} n$. Then any unit eigenvector v of the adjacency matrix of $G_{n,d}$ satisfies*

$$\|v\|_{\ell^\infty} \leq C' \frac{\log^2 n}{\sqrt{n}}$$

with probability at least $1 - \exp(-2 \log^2 n \cdot \log \log n)$.

Remark 9.4. More generally, the results in [7] hold for other models of random regular graphs besides the uniform model $G_{n,d}$ discussed here; see [7, Section 1.2] for details.

10. Proofs for the Gaussian orthogonal ensemble

In order to prove the results in Section 2, we will need the following characterization of a unit vector uniformly distributed on the unit sphere S^{n-1}.

Lemma 10.1. *Let v be a random vector uniformly distributed on the unit sphere S^{n-1}. Then v has the same distribution as*

$$\left(\frac{\xi_1}{\sqrt{\sum_{j=1}^{n} \xi_j^2}}, \ldots, \frac{\xi_n}{\sqrt{\sum_{j=1}^{n} \xi_j^2}} \right)$$

where $\xi_1, \ldots, \xi_n$ are iid standard normal random variables.

Proof. The claim follows from the fact that the Gaussian vector $(\xi_k)_{k=1}^{n}$ is rotationally invariant. We refer the reader to [44] for further details and other interesting results regarding entries of uniformly distributed unit vectors, and, more generally, results concerning entries of orthogonal matrices distributed according to Haar measure. ☐

We now prove Theorem 2.1. In order to do so, we will need the following result from [47].

Lemma 10.2 *(Lemma 1 from [47]). Suppose X is a χ^2-distributed with k degrees of freedom. Then, for all $t > 0$,*

$$\mathbb{P}(X - k \geq 2\sqrt{kt} + 2t) \leq \exp(-t)$$

and

$$\mathbb{P}(k - X \geq 2\sqrt{kt}) \leq \exp(-t).$$

Proof of Theorem 2.1. In view of Lemma 10.1, it suffices to assume that

$$v := \left(\frac{\xi_1}{\sqrt{\sum_{j=1}^{n} \xi_j^2}}, \ldots, \frac{\xi_n}{\sqrt{\sum_{j=1}^{n} \xi_j^2}} \right),$$

where $\xi_1, \ldots, \xi_n$ are iid standard normal random variables.

We first verify (4). Let $C > 1$ and $c_1 = 1/C < 1$. Define the events

$$\Omega_1 := \left\{ \sqrt{\sum_{j=1}^{n} \xi_j^2} \geq c_1 \sqrt{n} \right\}$$

and

$$\Omega_2 := \left\{ \max_{1 \leq i \leq n} |\xi_i| \leq \sqrt{2C \log n} \right\}.$$

In order to verify (4), it suffices to show $\Omega_1 \cap \Omega_2$ holds with probability at least $1 - 2n^{1-C} - \exp(-\frac{(C-1)^2}{4C^2}n)$.

As $\sum_{j=1}^{n} \xi_j^2$ is χ^2-distributed with n degrees of freedom, Lemma 10.2 implies that

$$\mathbb{P}(\Omega_1^c) = \mathbb{P}\left(\sum_{j=1}^{n} \xi_j^2 < c_1^2 n\right) \leq \exp\left(-\frac{(1-c_1^2)^2}{4}n\right) \leq \exp\left(-\frac{(1-c_1)^2}{4}n\right).$$

Since ξ_1 is a standard normal random variable, it follows that, for every $t \geq 0$,

$$\mathbb{P}(|\xi_1| > t) \leq 2e^{-t^2/2}; \tag{12}$$

this bound can be deduced from the exponential Markov inequality; see also [84, Section 5.2.3]. Thus, from (12), we have

$$\mathbb{P}(\Omega_2^c) \leq n\mathbb{P}(|\xi_1| > \sqrt{2C\log n}) \leq 2n\exp(-C\log n) = 2n^{1-C}.$$

Combining the bounds above yields

$$\mathbb{P}(\Omega_1^c \cup \Omega_2^c) \leq \exp\left(-\frac{(1-c_1)^2}{4}n\right) + 2n^{1-C}$$

and the proof of (4) is complete.

We now verify (5). Let $0 \leq c < 1$ and $a > 1$. Define the events

$$\Omega_3 := \left\{ \sqrt{\sum_{j=1}^{n} \xi_j^2} \leq a\sqrt{n} \right\}$$

and

$$\Omega_4 := \left\{ \min_{1 \leq i \leq n} |\xi_i| \geq \frac{c}{n} \right\}.$$

It suffices to show that $\Omega_3 \cap \Omega_4$ holds with probability at least

$$\exp(-2c) - \exp\left(-\frac{a^2 - \sqrt{2a^2 - 1}}{2}n\right).$$

From Lemma 10.2, we again find

$$\mathbb{P}(\Omega_3^c) \leq \exp\left(-\frac{(\sqrt{2a^2-1}-1)^2}{4}n\right) = \exp\left(-\frac{a^2 - \sqrt{2a^2-1}}{2}n\right). \tag{13}$$

Since ξ_1 is a standard normal random variable, we have

$$\mathbb{P}\left(|\xi_1| \leq \frac{c}{n}\right) = \frac{2}{\sqrt{2\pi}} \int_0^{cn^{-1}} e^{-t^2/2} dt \leq \int_0^{cn^{-1}} dt = \frac{c}{n}.$$

Hence, we obtain

$$\mathbb{P}(\Omega_4) = \left(1 - \mathbb{P}\left(|\xi_1| < \frac{c}{n}\right)\right)^n$$

$$\geq \left(1 - \frac{c}{n}\right)^n$$

$$= \exp\left(n \log\left(1 - \frac{c}{n}\right)\right).$$

By expanding the Taylor series for $\log(1-x)$, it follows that

$$\log\left(1 - \frac{c}{n}\right) \geq -\frac{c}{n} - \frac{c^2}{n^2}\frac{1}{1-\frac{c}{n}},$$

and hence

$$\mathbb{P}(\Omega_4) \geq \exp(-c)\exp\left(-\frac{c^2}{n-c}\right) \geq \exp(-2c), \tag{14}$$

for $n \geq 2$.

Since

$$\mathbb{P}(\Omega_3 \cap \Omega_4) \geq 1 - \mathbb{P}(\Omega_3^c) - \mathbb{P}(\Omega_4^c) = \mathbb{P}(\Omega_4) - \mathbb{P}(\Omega_3^c),$$

we apply (13) and (14) to conclude that

$$\mathbb{P}(\Omega_3 \cap \Omega_4) \geq \exp(-2c) - \exp\left(-\frac{a^2 - \sqrt{2a^2 - 1}}{2} n\right),$$

as desired. □

We now prove Theorem 2.2 using Lemma 10.1 and the law of large numbers.

Proof of Theorem 2.2. In view of Lemma 10.1, it suffices to assume that

$$v := \left(\frac{\xi_1}{\sqrt{\sum_{j=1}^n \xi_j^2}}, \ldots, \frac{\xi_n}{\sqrt{\sum_{j=1}^n \xi_j^2}}\right),$$

where $\xi_1, \ldots, \xi_n$ are iid standard normal random variables.

Fix $p \geq 1$, and define $c_p := \mathbb{E}|\xi_1|^p$. Then

$$\frac{n^{p/2}}{n}\|v\|_{\ell^p}^p - c_p = \left(\left(\frac{n}{\sum_{j=1}^n |\xi_j|^2} \right)^{p/2} - 1 \right) \frac{1}{n} \sum_{i=1}^n |\xi_i|^p + \frac{1}{n} \sum_{i=1}^n \left(|\xi_i|^p - c_p \right).$$

By the law of large numbers, it follows that almost surely

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n |\xi_i|^p = c_p \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^n |\xi_j|^2 = 1.$$

Hence, we conclude that almost surely

$$\lim_{n\to\infty} \left| \frac{n^{p/2}}{n} \|v\|_{\ell^p}^p - c_p \right| = 0,$$

and the claim follows. $\square$

Let $S \subset [n]$. It follows from Lemma 10.1 that $\|v\|_S^2$ has the same distribution as

$$\frac{\sum_{i=1}^{|S|} \xi_i^2}{\sum_{j=1}^n \xi_j^2}, \tag{15}$$

where $\xi_1, \ldots, \xi_n$ are iid standard normal random variables. In particular, we observe that, for any $1 \le k \le n$, $\sum_{j=1}^k \xi_j^2$ is χ^2-distributed with k degrees of freedom. Thus, the random variable in (15) can be expressed as $\frac{X}{X+W}$ where X and W are independent χ^2-distributed random variables with $|S|$ and $n - |S|$ degrees of freedom respectively. Theorem 2.3 follows from computing the distribution of this ratio; see [72] for details.

Theorem 2.4 is an immediate consequence of Theorem 2.3 and the following lemma.

Lemma 10.3. *Let $(\alpha_n)_{n=1}^\infty$ and $(\beta_n)_{n=1}^\infty$ be sequences of positive integers which satisfy*

(i) $\alpha_n \to \infty$ *and* $\beta_n \to \infty$ *as* $n \to \infty$,
(ii) $\frac{\alpha_n}{\alpha_n + \beta_n}$ *converges to a limit in* $[0, 1]$ *as* $n \to \infty$.

For each $n \ge 1$, let $X_n \sim \text{Beta}\left(\frac{\alpha_n}{2}, \frac{\beta_n}{2} \right)$. Then

$$\sqrt{\frac{(\alpha_n + \beta_n)^3}{2\alpha_n \beta_n}} \left(X_n - \frac{\alpha_n}{\alpha_n + \beta_n} \right) \longrightarrow N(0,1)$$

in distribution as $n \to \infty$.

We present the proof of Lemma 10.3 in Appendix A. We now prove Theorem 2.5.

Proof of Theorem 2.5. Let $m := |S|$. In view of Lemma 10.1, it suffices to bound

$$\left| \frac{\sum_{i=1}^{m} \xi_i^2}{\sum_{j=1}^{n} \xi_j^2} - \frac{m}{n} \right|,$$

where $\xi_1, \ldots, \xi_n$ are iid standard normal random variables. Fix $t > 0$, and define the events

$$\Omega_1 := \left\{ \sum_{j=1}^{n} \xi_j^2 \geq \frac{n}{2} \right\},$$

$$\Omega_2 := \left\{ \left| \sum_{j=1}^{n} \xi_j^2 - n \right| \leq 2\sqrt{nt} + 2t \right\},$$

$$\Omega_3 := \left\{ \left| \sum_{i=1}^{m} \xi_i^2 - m \right| \leq 2\sqrt{mt} + 2t \right\}.$$

By the union bound and Lemma 10.2, we have

$$\mathbb{P}(\Omega_1^c \cup \Omega_2^c \cup \Omega_3^c) \leq \exp(-cn) + 4\exp(-t)$$

for some absolute constant $c > 0$.

On the event $\Omega_1 \cap \Omega_2 \cap \Omega_3$, we observe that

$$
\begin{aligned}
\left| \frac{\sum_{i=1}^{m} \xi_i^2}{\sum_{j=1}^{n} \xi_j^2} - \frac{m}{n} \right|
&\leq \frac{\left| \sum_{i=1}^{m} \xi_i^2 - m \right|}{\sum_{j=1}^{n} \xi_j^2} + \left| \frac{m}{\sum_{j=1}^{n} \xi_j^2} - \frac{m}{n} \right| \\
&\leq \frac{2}{n} \left| \sum_{i=1}^{m} \xi_i^2 - m \right| + \frac{m}{n} \left| \frac{n}{\sum_{j=1}^{n} \xi_j^2} - 1 \right| \\
&\leq \frac{2}{n} \left(2\sqrt{mt} + 2t \right) + 2\frac{m}{n^2} \left| n - \sum_{j=1}^{n} \xi_j^2 \right| \\
&\leq \frac{4}{n} \left(\sqrt{mt} + t \right) + 4\frac{m}{n^2} \left(\sqrt{nt} + t \right).
\end{aligned}
$$

Since $m \leq n$, the claim follows. $\square$

We conclude this section with the proofs of Theorems 2.7 and 2.10.

Proof of Theorem 2.7. In view of Lemma 10.1, it suffices to consider the vector

$$v_n := \frac{1}{\sqrt{\sum_{j=1}^{n} \xi_j^2}} (\xi_1, \ldots, \xi_n),$$

where $\xi_1, \xi_2, \ldots$ are iid standard normal random variables. Thus, $\xi_1^2, \xi_2^2, \ldots$ are iid χ^2-distributed random variables with one degree of freedom. Recall that F is the cu-

mulative distribution function of ξ_1^2, Q is the quantile function of F defined in (6), and H is defined in (7).

Let $\xi_{(n,1)}^2 \leq \cdots \leq \xi_{(n,n)}^2$ denote the order statistics based on the sample $\xi_1^2, \ldots, \xi_n^2$. Let $m_n := \lfloor \delta n \rfloor$. Define

$$S_{n,m_n} := \sum_{i=1}^{m_n} \xi_{(n,n-m_n+i)}^2$$

to be the partial sum of the largest m_n entries, and set

$$S_n := \sum_{i=1}^{n} \xi_i^2.$$

We observe that

$$\max_{S \subset [n]: |S|=m_n} \|v_n\|_S^2 = \frac{S_{n,m_n}}{S_n}. \tag{16}$$

Since F is a special case of the gamma distribution, it follows from [20, Theorem 1.1.8] and [19, Corollary 2] that

$$\frac{S_{n,m_n} - \mu_n}{\sqrt{n} a_n} \longrightarrow N(0,1) \tag{17}$$

in distribution as $n \to \infty$, where

$$\mu_n := -n \int\limits_{1/n}^{m_n/n} H(u) \, du - H\left(\frac{1}{n}\right)$$

and

$$a_n := \left(\int\limits_{1/n}^{m_n/n} \int\limits_{1/n}^{m_n/n} (\min(u,v) - uv) \, dH(u) dH(v) \right)^{1/2}.$$

Since

$$\lim_{x \to \infty} \frac{1 - F(x)}{\sqrt{\frac{2}{\pi}} x^{-1/2} \exp(-x/2)} = 1,$$

we apply [73, Proposition 1] to obtain that $H(\frac{1}{n}) \leq C \log n$ for some constant $C > 0$. Thus, we have

$$\lim_{n\to\infty} \frac{\mu_n}{n} = -\int_0^\delta H(u)\, du$$

and

$$\lim_{n\to\infty} a_n = \left(\int_0^\delta \int_0^\delta (\min(u,v) - uv)\, dH(u) dH(v) \right)^{1/2}.$$

From (17), we conclude that

$$\frac{S_{n,m_n}}{n} \longrightarrow -\int_0^\delta H(u)\, du$$

in probability as $n \to \infty$.

On the other hand, by the law of large numbers, we have

$$\frac{S_n}{n} \longrightarrow 1$$

in probability as $n \to \infty$. Thus, by Slutsky's theorem (see Theorem 11.4 in [42, Chapter 5]) and (16), we have

$$\max_{S\subset[n]:|S|=m_n} \|v_n\|_S^2 = \frac{S_{n,m_n}}{n} \cdot \frac{n}{S_n} \longrightarrow -\int_0^\delta H(u)\, du$$

in probability as $n \to \infty$.

For the minimum, we note that

$$\min_{S\subset[n]:|S|=m_n} \|v_n\|_S^2 = 1 - \max_{T\subset[n]:|T|=n-m_n} \|v_n\|_T^2,$$

and hence

$$1 - \max_{T\subset[n]:|T|=n-m_n} \|v_n\|_T^2 \longrightarrow 1 + \int_0^{1-\delta} H(u)\, du = -\int_{1-\delta}^1 H(u)\, du$$

in probability as $n \to \infty$. Here, we used the fact that $\int_0^1 Q(s)\, ds = -\int_0^1 H(u)\, du = 1$. $\square$

Proof of Theorem 2.10. We first observe that it suffices to prove (8). Indeed, (9) follows immediately from (8) by applying the identity

$$\max_{S\subset[n]:|S|=m} \|v\|_S + \min_{S\subset[n]:|S|=n-m} \|v\|_S = 1.$$

Let $1 \leq m \leq n$, and define the function $F : \mathbb{R}^n \to \mathbb{R}$ by

$$F(X) := \max_{S \subset [n]:|S|=m} \|X\|_S.$$

Clearly, $|F(v) - \mathbb{E}F(v)| \leq 2\|v\| = 2$. Thus, it suffices to show

$$\mathbb{P}\left(|F(v) - \mathbb{E}F(v)| > t\right) \leq C \exp(-ct^2 n)$$

for all $t \in [0, 2]$ (as opposed to all $t \geq 0$). In addition, by taking the absolute constant C sufficiently large, it suffices to prove

$$\mathbb{P}\left(|F(v) - \mathbb{E}F(v)| > t\right) \leq C \exp(-ct^2 n) \tag{18}$$

for all $t \in \left(24 n^{-1/2}, 2\right]$.

Clearly, F is 1-Lipschitz. Thus, by Lévy's lemma (see, for example, [49, Theorem 14.3.2]), we have, for all $0 \leq t \leq 1$,

$$\mathbb{P}\left(|F(v) - \mathbb{M}F(v)| > t\right) \leq 4 \exp\left(-t^2 n/2\right). \tag{19}$$

Here $\mathbb{M}F(v)$ denotes the median of the random variable $F(v)$. In addition, [49, Proposition 14.3.3] implies that

$$|\mathbb{M}F(v) - \mathbb{E}F(v)| \leq \frac{12}{\sqrt{n}}. \tag{20}$$

We will use (20) to replace the median appearing in (19) with expectation. Indeed, assume $24 n^{-1/2} < t \leq 2$. Then, from (20), we have

$$\mathbb{P}\left(|F(v) - \mathbb{E}F(v)| > t\right) \leq \mathbb{P}\left(|F(v) - \mathbb{M}F(v)| > t - |\mathbb{E}F(v) - \mathbb{M}F(v)|\right)$$

$$\leq \mathbb{P}\left(|F(v) - \mathbb{M}F(v)| > \frac{t}{2}\right).$$

The bound in (18) now follows by applying (19). □

11. Tools required for the remaining proofs

We now turn our attention to the remaining proofs. In order to better organize the arguments, we start by collecting a variety of deterministic and probabilistic tools we will require in subsequent sections.

11.1. Tools from linear algebra

The Courant–Fisher minimax characterization of the eigenvalues (see, for instance, [11, Chapter III]) states that

$$\lambda_i(M) = \min_V \max_{u \in V} u^* M u,$$

where M is a Hermitian $n \times n$ matrix, V ranges over i-dimensional subspaces of $\mathbb{C}^n$, and u ranges over unit vectors in V.

From this one can obtain *Cauchy's interlacing inequalities*:

$$\lambda_i(M_n) \leq \lambda_i(M_{n-1}) \leq \lambda_{i+1}(M_n) \tag{21}$$

for all $i < n$, where M_n is an $n \times n$ Hermitian matrix and M_{n-1} is the top $(n-1) \times (n-1)$ minor. One also has the following more precise version of the Cauchy interlacing inequality.

Lemma 11.1 *(Interlacing identity; Lemma 40 from [76]). Let*

$$M_n = \begin{pmatrix} M_{n-1} & X \\ X^* & m_{nn} \end{pmatrix}$$

be an $n \times n$ Hermitian matrix, where M_{n-1} is the upper-left $(n-1) \times (n-1)$ minor of M_n, $X \in \mathbb{C}^{n-1}$, and $m_{nn} \in \mathbb{R}$. Suppose that X is not orthogonal to any of the unit eigenvectors $v_j(M_{n-1})$ of M_{n-1}. Then we have

$$\sum_{j=1}^{n-1} \frac{|v_j(M_{n-1})^* X|^2}{\lambda_j(M_{n-1}) - \lambda_i(M_n)} = m_{nn} - \lambda_i(M_n) \tag{22}$$

for every $1 \leq i \leq n$.

The following lemma is needed when one wants to consider the coordinates of an eigenvector.

Lemma 11.2 *([28]; Lemma 41 from [76]). Let*

$$M_n = \begin{pmatrix} M_{n-1} & X \\ X^* & m_{nn} \end{pmatrix}$$

be an $n \times n$ Hermitian matrix, where M_{n-1} is the upper-left $(n-1) \times (n-1)$ minor of M_n, $X \in \mathbb{C}^{n-1}$, and $m_{nn} \in \mathbb{R}$. Let $\begin{pmatrix} v \\ x \end{pmatrix}$ be a unit eigenvector of $\lambda_i(M_n)$, where $v \in \mathbb{C}^{n-1}$ and $x \in \mathbb{C}$. Suppose that none of the eigenvalues of M_{n-1} are equal to $\lambda_i(M_n)$. Then

$$|x|^2 = \frac{1}{1 + \sum_{j=1}^{n-1} (\lambda_j(M_{n-1}) - \lambda_i(M_n))^{-2} |v_j(M_{n-1})^* X|^2},$$

where $v_j(M_{n-1})$ is a unit eigenvector corresponding to the eigenvalue $\lambda_j(M_{n-1})$.

11.2. Spectral norm

We will make use of the following bound for the spectral norm of a Wigner matrix with sub-gaussian entries.

Lemma 11.3 (*Spectral norm of a Wigner matrix; Lemma 5 from [55]*). *Let ξ, ζ be real sub-gaussian random variables with mean zero, and assume ξ has unit variance. Let W be an $n \times n$ Wigner random matrix with atom variables ξ, ζ. Then there exist constants $C_0, c_0 > 0$ (depending only on the sub-gaussian moments of ξ and ζ) such that $\|W\| \leq C_0 \sqrt{n}$ with probability at least $1 - C_0 \exp(-c_0 n)$.*

11.3. Local semicircle law

While Wigner's semicircle law (see, for example, [5, Theorem 2.5]) describes the global behavior of the eigenvalues of a Wigner matrix, the local semicircle law describes the fine-scale behavior of the eigenvalues. Many authors have proved versions of the local semicircle law under varying assumptions; we refer the reader to [23,36,37,29,27,28,30, 31,33,26,32,34,35,48,75–77,80] and references therein.

The local semicircle law stated below was proved by Lee and Yin in [48]. We let ρ_{sc} denote the density of the semicircle distribution defined by

$$\rho_{sc}(x) := \begin{cases} \frac{1}{2\pi}\sqrt{4 - x^2}, & |x| \leq 2, \\ 0, & |x| > 2. \end{cases} \tag{23}$$

Theorem 11.4 (*Theorem 3.6 from [48]*). *Let ξ, ζ be real sub-gaussian random variables with mean zero, and assume ξ has unit variance. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Let N_I be the number of eigenvalues of $\frac{1}{\sqrt{n}}W$ in the interval I. Then, there exist constants $C, c, c' > 0$ (depending on the sub-gaussian moments of ξ and ζ) such that, for any interval $I \subset \mathbb{R}$,*

$$\mathbb{P}\left(\left|N_I - n\int\limits_I \rho_{sc}(x)dx\right| \geq (\log n)^{c' \log\log n}\right) \leq C\exp\left(-c(\log n)^{c \log\log n}\right).$$

It will occasionally be useful to avoid the $(\log n)^{c' \log\log n}$ term present in Theorem 11.4 in favor of a bound of the form $(\log n)^{c'}$. In these cases, the following result will be useful.

Theorem 11.5. *Let ξ, ζ be real sub-gaussian random variables with mean zero, and assume ξ has unit variance. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Let N_I be the number of eigenvalues of $\frac{1}{\sqrt{n}}W$ in the interval I. Then, there exist constants $C, c, c' > 0$ (depending on the sub-gaussian moments of ξ and ζ) such that, for any interval $I \subset \mathbb{R}$,*

$$\mathbb{P}\left(N_I \geq (1.1)n \int_I \rho_{\mathrm{sc}}(x)dx + (\log n)^{c'}\right) \leq C\exp\left(-c\log^2 n\right).$$

Remark 11.6. The constant 1.1 appearing in Theorem 11.5 can be replaced by any absolute constant larger than one.

Theorem 11.5 follows from the arguments in [76,86] (see [76, Proposition 66] for details).

11.4. Smallest singular value

We will need the following result concerning the least singular value of a rectangular random matrix with iid entries. For an $N \times n$ matrix M, we let $\sigma_1(M) \geq \cdots \geq \sigma_n(M)$ denote the ordered eigenvalues of $\sqrt{M^*M}$. (Note that the non-zero eigenvalues of $\sqrt{MM^*}$ are the same as the non-zero eigenvalues of $\sqrt{M^*M}$.) In particular, $\sigma_1(M) = \|M\|$, and $\sigma_n(M)$ is called the smallest singular value of M. We refer the interested reader to [63] for a wonderful survey on the non-asymptotic theory of extreme singular values of random matrices with independent entires.

Theorem 11.7 *(Theorem 1.1 from [62]). Let B be an $N \times n$ random matrix, $N \geq n$, whose elements are independent copies of a mean zero sub-gaussian random variable ξ with unit variance. Then, for every $\varepsilon > 0$,*

$$\mathbb{P}\left(\sigma_n(B) \leq \varepsilon\left(\sqrt{N} - \sqrt{n-1}\right)\right) \leq (C\varepsilon)^{N-n+1} + e^{-cN},$$

where $C, c > 0$ depend only on the sub-gaussian moment of ξ.

Similar bounds for the smallest singular value were also obtained by Vershynin [83] under the assumption that ξ has $4 + \varepsilon$ finite moments instead of the sub-gaussian assumption above. We also refer the reader to [84] for bounds on the extreme singular values of random matrices with heavy-tailed rows.

11.5. Projection lemma

We will also need the following bound, which follows from the Hanson–Wright inequality (see, for example, [65, Theorem 1.1]).

Lemma 11.8 *(Projection lemma). Let ξ be a sub-gaussian random variable with mean zero and unit variance. Let B be a $n \times m$ random matrix whose entries are iid copies of ξ. Let H be a subspace of $\mathbb{R}^n$ of dimension d, and let P_H denote the orthogonal projection onto H. Then there exist constants $C, c > 0$ (depending only on the sub-gaussian moment of ξ) such that, for any unit vector $y \in \mathbb{R}^m$ and every $t \geq 0$,*

$$\mathbb{P}\left(\left|\|P_H By\|^2 - d\right| > t\right) \le C \exp\left(-c \min\left\{\frac{t^2}{d}, t\right\}\right).$$

Proof. Let $B_1, \dots, B_m$ denote the columns of B, and set $y = (y_i)_{i=1}^m$. Since P_H is an orthogonal projection, we write $P_H = \sum_{i=1}^d u_i u_i^{\mathrm{T}}$, where $\{u_1, \dots, u_d\}$ is an orthonormal basis of H.

Let X denote the mn-vector with iid entries given by

$$X := \begin{pmatrix} B_1 \\ \vdots \\ B_m \end{pmatrix}.$$

Define the $mn \times mn$ matrix $\mathcal{P} := \sum_{i=1}^d w_i w_i^{\mathrm{T}}$, where

$$w_i := \begin{pmatrix} y_1 u_i \\ \vdots \\ y_m u_i \end{pmatrix}.$$

It follows, from the definitions above, that $X^{\mathrm{T}} \mathcal{P} X = \|P_H By\|^2$. Moreover, since $\{w_1, \dots, w_d\}$ is an orthonormal set, $\mathcal{P}$ is an orthogonal projection. Thus, $\|\mathcal{P}\| = 1$ and

$$\operatorname{tr} \mathcal{P} = \|\mathcal{P}\|_2^2 = \operatorname{rank}(\mathcal{P}) = d.$$

Since the entries of X are iid copies of ξ, we find that

$$\mathbb{E} X^{\mathrm{T}} \mathcal{P} X = \operatorname{tr} \mathcal{P} = d.$$

So, by the Hanson–Wright inequality (see, for example, [65, Theorem 1.1]), we obtain, for all $t \ge 0$,

$$\mathbb{P}\left(\left|X^{\mathrm{T}} \mathcal{P} X - d\right| > t\right) \le C \exp\left(-c \min\left\{\frac{t^2}{d}, t\right\}\right),$$

where $C, c > 0$ depends only on the sub-gaussian moment of ξ. $\quad\square$

In the case when B is an $n \times 1$ matrix (i.e. a vector) and $y = 1$, we immediately obtain the following corollary.

Corollary 11.9. *Let ξ be a sub-gaussian random variable with mean zero and unit variance. Let X be a vector in $\mathbb{R}^n$ whose entries are iid copies of ξ. Let H be a subspace of $\mathbb{R}^n$ of dimension d, and let P_H denote the orthogonal projection onto H. Then there exist constants $C, c > 0$ (depending only on the sub-gaussian moment of ξ) such that, for every $t \ge 0$,*

$$\mathbb{P}\left(\left|\|P_H X\|^2 - d\right| > t\right) \le C \exp\left(-c \min\left\{\frac{t^2}{d}, t\right\}\right).$$

11.6. Deterministic tools and the equation $Ax = By$

We now consider the equation $Ax = By$, where x and y are vectors, A is a rectangular matrix, and B is a Hermitian matrix. If $\|x\|$ is small, then $\|Ax\| = \|By\|$ is also relatively small. Intuitively, then, it must be the case that the vector y is essentially supported on the eigenvectors of B corresponding to small eigenvalues. In the following lemmata, we quantify the structure of y in terms of $\|A\|$ and the spectral decomposition of B. Similar results were implicitly used in [3].

Lemma 11.10. *Let A be a $r \times m$ matrix and B be a Hermitian $r \times r$ matrix. Let $\varepsilon, \tau > 0$. Let x and y be vectors with $\|x\| \leq \varepsilon\tau$ and $By = Ax$. Then $y = v + q$, where v and q are orthogonal, $\|q\| \leq \varepsilon$, and*

$$v \in \mathrm{Span}\{v_i(B) : |\lambda_i(B)| \leq \tau\|A\|\}.$$

In many cases, the norm of A will be too large for the above lemma to be useful. However, if we can write A as a sum of two parts, one with small norm and one with low rank, we can still obtain essentially the same conclusion using the following lemma.

Lemma 11.11. *Let A and J be $r \times m$ matrices with $\|A\| \leq \kappa$, and let B be a Hermitian $r \times r$ matrix. Let $\varepsilon, \tau > 0$. Let x and y be vectors with $\|x\| \leq \varepsilon\tau$, $\|y\| \leq 1$, and $By = (A+J)x$. Then*

(i) there exists a non-negative real number η (depending only on $B, \kappa, \tau, \varepsilon$) such that $B - \eta I$ is invertible,
(ii) the vector y can be decomposed as $y = v + q$, where $\|q\| \leq \varepsilon$,

$$v \in V := \mathrm{Span}\left\{\{v_i(B) : |\lambda_i(B)| \leq \tau\kappa\} \cup \mathrm{range}\left((B - \eta I)^{-1}J\right)\right\},$$

and

$$\dim(V) \leq |\{1 \leq i \leq r : |\lambda_i(B)| \leq \tau\kappa\}| + \mathrm{rank}(J). \tag{24}$$

Remark 11.12. One can similarly prove versions of Lemmas 11.10 and 11.11 when B is not Hermitian. In this case, one must rely on the singular value decomposition of B instead of the spectral theorem. In particular, the set V appearing in the statement of Lemma 11.11 will need to be defined in terms of the singular values and singular vectors of $B - \eta I$.

We prove Lemmas 11.10 and 11.11 in Appendix B.

12. Proofs of results concerning extremal coordinates

This section is devoted to the proofs of Corollary 4.4 and Theorem 4.7.

12.1. Proof of Corollary 4.4

The desired bound will follow from [86, Theorem 6.1] and a simple truncation argument. Let W be the Wigner matrix from Corollary 4.4 with sub-exponential atom variable ξ. Then, there exist $\alpha, \beta > 0$ such that

$$\mathbb{P}(|\xi| > t) \leq \beta \exp(-t^\alpha/\beta) \tag{25}$$

for all $t > 0$. In addition, since ξ is symmetric, it follows that ξ has mean zero. Let $C_1 > 0$ be given, and take $C > 0$ to be a large constant to be chosen later. Define

$$\tilde{\xi} := \xi \mathbf{1}_{\{|\xi| \leq C \log^{1/\alpha} n\}},$$

where $\mathbf{1}_E$ is the indicator function of the event E. Since ξ is symmetric it follows that $\tilde{\xi}$ has mean zero. Similarly, we define the matrix $\tilde{W} = (\tilde{w}_{ij})_{i,j=1}^n$ by

$$\tilde{w}_{ij} := w_{ij} \mathbf{1}_{\{|w_{ij}| \leq C \log^{1/\alpha} n\}}.$$

It follows that $\tilde{W}$ is an $n \times n$ Wigner matrix with atom variable $\tilde{\xi}$. Moreover, $\tilde{\xi}$ has mean zero and is $(C \log^{1/\alpha} n)$-bounded. Thus, [86, Theorem 6.1] can be applied to $\tilde{W}$,[2] and we obtain the desired conclusion for $\tilde{W}$. It remains to show that the same conclusion also holds for W. From (25), we have

$$\mathbb{P}(W \neq \tilde{W}) \leq \sum_{i \leq j} \mathbb{P}(w_{ij} \neq \tilde{w}_{ij})$$

$$\leq n^2 \mathbb{P}(|\xi| > C \log^{1/\alpha} n)$$

$$\leq \beta n^2 \exp\left(-\frac{C^\alpha}{\beta} \log n\right)$$

$$\leq \beta n^{-C_1}$$

by taking C sufficiently large. Therefore, on the event where $\tilde{W} = W$, we obtain the desired conclusion, and the proof is complete.

12.2. Proof of Theorem 4.7

We will need the following result from [54].

Theorem 12.1 *([54]). Let W_n be the $n \times n$ Wigner matrix from Theorem 4.7. Let*

---

[2] Technically, one also has to normalize the entries of $\tilde{W}$ to have unit variance as required by [86, Theorem 6.1]. However, this corresponds to multiplying $\tilde{W}$ by a positive scalar and the unit eigenvectors are invariant under such a scaling.

$$W_n = \begin{bmatrix} W_{n-1} & X \\ X^{\mathrm{T}} & w_{nn} \end{bmatrix},$$

where W_{n-1} is the upper-left $(n-1) \times (n-1)$ minor of W_n, $X \in \mathbb{R}^{n-1}$, and $w_{nn} \in \mathbb{R}$. Then there exist constants $C, c, c_0 > 0$ such that, for any $n^{-c_0} < \alpha < c_0$ and $\delta \geq n^{-c_0/\alpha}$,

$$\sup_{1 \leq i \leq n-1} \mathbb{P}\left(\lambda_i(W_{n-1}) - \lambda_i(W_n) \leq \frac{\delta}{\sqrt{n}} \right) \leq C\frac{\delta}{\sqrt{\alpha}} + C\exp(-c\log^2 n)$$

and

$$\sup_{2 \leq i \leq n} \mathbb{P}\left(\lambda_i(W_n) - \lambda_{i-1}(W_{n-1}) \leq \frac{\delta}{\sqrt{n}} \right) \leq C\frac{\delta}{\sqrt{\alpha}} + C\exp(-c\log^2 n).$$

Moreover,

$$\inf_{1 \leq j \leq n-1} |v_j(W_{n-1})^{\mathrm{T}} X| > 0 \tag{26}$$

with probability at least $1 - C\exp(-c\log^2 n)$.

Remark 12.2. The bound in (26) implies that X is not orthogonal to $v_j(W_{n-1})$ for any $1 \leq j \leq n-1$. By [55, Lemma 3], this implies that none of the eigenvalues of W_n coincide with an eigenvalue of W_{n-1}. In other words, the interlacing described in (21) is strict.

Theorem 12.1 follows from [54, Theorems 4.1 and 4.3] and the arguments given in [54, Section 4]. We will also need the following technical lemma.

Lemma 12.3. Let W_n be the $n \times n$ Wigner matrix from Theorem 4.7. Let

$$W_n = \begin{bmatrix} W_{n-1} & X \\ X^{\mathrm{T}} & w_{nn} \end{bmatrix},$$

where W_{n-1} is the upper-left $(n-1) \times (n-1)$ minor of W_n, $X \in \mathbb{R}^{n-1}$, and $w_{nn} \in \mathbb{R}$. Let $\kappa > 0$. Then there exist constants $C, c, c_1 > 0$ such that, for any $1 \leq i \leq n$ and any $e^{-n^{\kappa}} < \delta < 1$,

$$\sum_{j=1}^{n-1} \frac{|v_j(W_{n-1})^{\mathrm{T}} X|^2}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|^2} \leq \frac{(\log n)^{c_1}}{25}\left[\frac{1}{m_i^2} + \frac{n}{\delta^2} \right] \tag{27}$$

with probability at least $1 - C\exp(-c\log^2 n)$, where

$$m_i := \min_{1 \leq j \leq n-1} |\lambda_j(W_{n-1}) - \lambda_i(W_n)|.$$

Proof. Fix $1 \leq i \leq n$ and $e^{-n^{\kappa}} < \delta < 1$. Define the event Ω to be the intersection of the events

$$\bigcap_{j=1}^{n-1} \left\{ 0 < |v_j(W_{n-1})^{\mathrm{T}}X|^2 \leq \frac{1}{1000} \log^2 n \right\} \tag{28}$$

and

$$\left\{ \|W_n\| \leq C_0 \sqrt{n} \right\} \tag{29}$$

for some constant C_0 to be chosen later.

We claim that, for C_0 sufficiently large,

$$\mathbb{P}(\Omega^c) \leq C \exp(-c \log^2 n) \tag{30}$$

for some constants $C, c > 0$. To see this, note that $|v_j(W_{n-1})^{\mathrm{T}}X|$ is the length of the projection of the vector X onto the one-dimensional subspace spanned by $v_j(W_{n-1})$. Hence, by Corollary 11.9, the union bound, and (26), it follows that the event in (28) holds with probability at least $1 - C \exp(-c \log^2 n)$. In addition, the event in (29) can be dealt with by taking C_0 sufficiently large and applying Lemma 11.3. Combining the estimates above yields the bound in (30).

It now suffices to show that, conditionally on Ω, the bound in (27) holds with probability at least $1 - C' \exp(-c' \log^2 n)$. As Remark 12.2 implies, on Ω, the eigenvalues of W_{n-1} strictly interlace with the eigenvalues of W_n, and hence the terms

$$\sum_{j=1}^{n-1} \frac{|v_j(W_{n-1})^{\mathrm{T}}X|^2}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|^2}$$

and m_i^{-2} are well-defined. Moreover, on Ω, we have

$$\sum_{j=1}^{n-1} \frac{|v_j(W_{n-1})^{\mathrm{T}}X|^2}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|^2} \leq \frac{\log^2 n}{1000} \sum_{j=1}^{n-1} \frac{1}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|^2}.$$

Hence, conditionally on Ω, it suffices to show that

$$\sum_{j=1}^{n-1} \frac{1}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|^2} \leq 10(\log n)^{c_1} \left[\frac{1}{m_i^2} + \frac{n}{\delta^2} \right]$$

with probability at least $1 - C' \exp(-c' \log^2 n)$.

Define the sets

$$T := \left\{ 1 \leq j \leq n - 1 : |\lambda_j(W_{n-1}) - \lambda_i(W_n)| < \frac{\delta}{\sqrt{n}} \right\}$$

and

$$T_l := \left\{ 1 \leq j \leq n-1 : 2^l \frac{\delta}{\sqrt{n}} \leq |\lambda_j(W_{n-1}) - \lambda_i(W_n)| < 2^{l+1} \frac{\delta}{\sqrt{n}} \right\}$$

for $l = 0, 1, \ldots, L$, where L is the smallest integer such that $2^L \frac{\delta}{\sqrt{n}} \geq 2C_0\sqrt{n}$. In particular, as $\delta > e^{-n^\kappa}$, we obtain $L = O(n^{-\kappa})$. On the event Ω, it follows that every index $1 \leq j \leq n-1$ is contained in either T or $\cup_{l=0}^{L} T_l$. By Theorem 11.5 and the union bound,[3] there exist $C', c', c_1 > 0$ such that

$$|T| \leq (\log n)^{c_1}$$

and

$$|T_l| \leq 4(2^l) + (\log n)^{c_1}, \quad l = 0, \ldots, L$$

with probability at least $1 - C' \exp(-c' \log^2 n)$. Hence, on this same event, we conclude that

$$\sum_{j=1}^{n-1} \frac{1}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|^2} \leq \sum_{j \in T} \frac{1}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|^2}$$
$$+ \sum_{l=0}^{L} \sum_{j \in T_l} \frac{1}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|^2}$$
$$\leq \frac{|T|}{m_i^2} + \frac{n}{\delta^2} \sum_{l=0}^{L} \frac{|T_l|}{2^{2l}}$$
$$\leq \frac{(\log n)^{c_1}}{m_i^2} + \frac{10n(\log n)^{c_1}}{\delta^2}$$

for n sufficiently large. The proof of the lemma is complete. □

With Theorem 12.1 and Lemma 12.3 in hand, we are now ready to prove Theorem 4.7.

Proof of Theorem 4.7. Let W_n be the $n \times n$ Wigner matrix with atom variables ξ, ζ. We will bound the jth coordinate of the unit eigenvector v_i in magnitude from below. By symmetry, it suffices to consider the case when $j = n$. For this reason, we decompose W_n as

---

[3] Technically, one cannot apply Theorem 11.5 directly to these intervals because the intervals are defined in terms of the eigenvalues. For instance, $|T|$ is the number of eigenvalues of W_{n-1} in the open interval of radius $\frac{\delta}{\sqrt{n}}$ centered at $\lambda_i(W_n)$. However, one can first take (say) n^{100} equispaced points in the interval $[-C_0\sqrt{n}, C_0\sqrt{n}]$ and apply the theorem to intervals of radius $\frac{\delta}{\sqrt{n}}$ centered at each of these points. One can then deduce the desired conclusion by approximating $|T|$ using these deterministic intervals. A similar approximation argument works for each T_l.

$$W_n = \begin{bmatrix} W_{n-1} & X \\ X^{\mathrm{T}} & w_{nn} \end{bmatrix},$$

where W_{n-1} is the upper-left $(n-1) \times (n-1)$ minor of W_n, $X \in \mathbb{R}^{n-1}$, and $w_{nn} \in \mathbb{R}$.

Let c_0 be as in Theorem 12.1, and assume $n^{-c_0} < \alpha < c_0$ and $1 > \delta \geq n^{-c_0/\alpha}$ as the claim is trivial when $\delta \geq 1$. Define the event Ω to be the intersection of the events

$$\left\{ \inf_{1 \leq j \leq n-1} |v_j(W_{n-1})^{\mathrm{T}} X| > 0 \right\} \bigcap \left\{ m_i > \frac{\delta}{\sqrt{n}} \right\}$$

and

$$\left\{ \sum_{j=1}^{n-1} \frac{|v_j(W_{n-1})^{\mathrm{T}} X|^2}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|^2} \leq \frac{(\log n)^{c_1}}{25} \left[\frac{1}{m_i^2} + \frac{n}{\delta^2} \right] \right\},$$

where

$$m_i := \min_{1 \leq j \leq n-1} |\lambda_j(W_{n-1}) - \lambda_i(W_n)|$$

and c_1 is the constant from Lemma 12.3. If $m_i \leq \frac{\delta}{\sqrt{n}}$, then, by Cauchy's interlacing inequalities (21), this implies that either

$$|\lambda_i(W_{n-1}) - \lambda_i(W_n)| \leq \frac{\delta}{\sqrt{n}}$$

or

$$|\lambda_{i-1}(W_{n-1}) - \lambda_i(W_n)| \leq \frac{\delta}{\sqrt{n}}.$$

(Here, the first possibility can only occur if $i < n$, and the second possibility can only occur if $i > 1$.) Therefore, it follows from Theorem 12.1 and Lemma 12.3 that, there exist $C', c' > 0$ such that Ω holds with probability at least $1 - C'\frac{\delta}{\sqrt{\alpha}} - C' \exp(-c' \log^2 n)$.

Let $c_2 := c_1/2$. Then

$$\mathbb{P}\left(|v_i(n)| \leq \frac{\delta}{\sqrt{n}(\log n)^{c_2}} \right) \leq \mathbb{P}(\Omega^c) + \mathbb{P}\left(\left\{ |v_i(n)| \leq \frac{\delta}{\sqrt{n}(\log n)^{c_2}} \right\} \bigcap \Omega \right).$$

Hence, to complete the proof, we will show that the event

$$\left\{ |v_i(n)| \leq \frac{\delta}{\sqrt{n}(\log n)^{c_2}} \right\} \bigcap \Omega \tag{31}$$

is empty.

Fix a realization in this event. As Remark 12.2 implies, the eigenvalues of W_{n-1} strictly interlace with the eigenvalues of W_n. Hence, we apply Lemma 11.2 and obtain

$$\frac{1}{1 + \sum_{j=1}^{n-1} \frac{|v_j(W_{n-1})^{\mathrm{T}} X|^2}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|^2}} = |v_i(n)|^2 \leq \frac{\delta^2}{n(\log n)^{2c_2}}.$$

This implies, for n sufficiently large, that

$$\frac{n(\log n)^{2c_2}}{2\delta^2} \leq \sum_{j=1}^{n-1} \frac{|v_j(W_{n-1})^{\mathrm{T}} X|^2}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|^2}.$$

On the other hand, by definition of Ω, we have

$$\sum_{j=1}^{n-1} \frac{|v_j(W_{n-1})^{\mathrm{T}} X|^2}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|^2} \leq \frac{2(\log n)^{c_1}}{25} \frac{n}{\delta^2},$$

and thus

$$\frac{n(\log n)^{2c_2}}{2\delta^2} \leq \frac{2(\log n)^{c_1}}{25} \frac{n}{\delta^2}.$$

This is a contradiction since $2c_2 = c_1$. We conclude that the event in (31) is empty, and the proof is complete. $\square$

Remark 12.4. In the special case that one considers only $v_1(W_n)$ or $v_n(W_n)$, a simpler argument is possible by applying Lemma 11.1. Indeed, in this case, one can exploit the fact that the terms $\lambda_j(M_{n-1}) - \lambda_i(M_n)$ appearing in the denominator of (22) are always of the same sign due to the eigenvalue interlacing inequalities (21). However, this argument does not appear to generalize to any other eigenvectors.

13. Proofs of no-gaps delocalization results

This section is devoted to the proofs of Theorem 5.1 and Corollary 5.4.

13.1. Proof for Theorem 5.1

We begin with a few reductions. Let $\varepsilon > 0$. By definition of convergence in probability, it suffices to show

$$\left| \max_{S \subset [n]: |S| = \lfloor \delta n \rfloor} \|v_{k_n}\|_S^2 + \int_0^\delta H(u)\, du \right| \leq \varepsilon$$

and

$$\left| \min_{S \subset [n]:|S|=\lfloor \delta n \rfloor} \|v_{k_n}\|_S^2 + \int\limits_{1-\delta}^{1} H(u)\,du \right| \le \varepsilon$$

with probability $1 - o(1)$.

Let Z be a standard normal distribution with cumulative distribution function $\Phi(x)$. Let $F(x)$ be the cumulative distribution function of Z^2, which has the χ^2-distribution with 1 degree of freedom. Hence,

$$F(x) = \mathbb{P}(Z^2 \le x) = 2\Phi(\sqrt{x}) - 1$$

for $x \ge 0$. By setting $x = \sqrt{F^{-1}(u)}$ in the following integrals, we observe that

$$- \int\limits_{1-\delta}^{1} H(u)\,du = \int\limits_{0}^{\delta} F^{-1}(u)\,du = 2 \int\limits_{0}^{\Phi^{-1}(\frac{1+\delta}{2})} x^2 \Phi'(x)\,dx$$

and

$$- \int\limits_{0}^{\delta} H(u)\,du = \int\limits_{1-\delta}^{1} F^{-1}(u)\,du = 2 \int\limits_{\Phi^{-1}(1-\frac{\delta}{2})}^{\infty} x^2 \Phi'(x)\,dx. \tag{32}$$

Thus, it suffices to show

$$\left| \max_{S \subset [n]:|S|=\lfloor \delta n \rfloor} \|v_{k_n}\|_S^2 - 2 \int\limits_{\Phi^{-1}(1-\frac{\delta}{2})}^{\infty} x^2 \Phi'(x)\,dx \right| \le \varepsilon \tag{33}$$

and

$$\left| \min_{S \subset [n]:|S|=\lfloor \delta n \rfloor} \|v_{k_n}\|_S^2 - 2 \int\limits_{0}^{\Phi^{-1}(\frac{1+\delta}{2})} x^2 \Phi'(x)\,dx \right| \le \varepsilon \tag{34}$$

with probability $1 - o(1)$, where Φ is the cumulative distribution function of the standard normal distribution.

We now turn our attention to proving (33) and (34). In fact, (33) follows from (34) by applying the identity

$$\max_{S \subset [n]:|S|=\lfloor \delta n \rfloor} \|v_{k_n}\|_S^2 + \min_{S \subset [n]:|S|=\lfloor \delta n \rfloor} \|v_{k_n}\|_{S^c}^2 = 1$$

and using the fact that $2 \int_0^\infty x^2 \Phi'(x)\,dx = 1$. (Alternatively, one can prove (33) by repeating the arguments below.) Thus, it remains to verify (34).

Let $\delta \in (0, 1)$. For notational convenience, let v denote the unit eigenvector under consideration, and let $v(j)$ denote its jth coordinate. Let $c > 0$ be a (small) constant to be chosen later. For each $k \in \mathbb{N}$, define

$$N(c, k) := \sum_{j=1}^{n} \mathbf{1}_{\{c(k-1) \leq \sqrt{n}|v(j)| < ck\}}$$

to be the number of coordinates of $\sqrt{n}v$ with magnitude in the interval $[c(k-1), ck)$.

Let Z be a standard normal random variable. Define

$$f(c, k) := n\mathbb{P}(c(k-1) \leq |Z| < ck).$$

Since Φ is the cumulative distribution function of Z, it follows that

$$f(c, k) = 2n\left(\Phi(ck) - \Phi(c(k-1))\right).$$

From Corollary 3.4, it follows that, for each $1 \leq j \leq n$,

$$\mathbb{P}(c(k-1) \leq \sqrt{n}|v(j)| < ck) = \mathbb{P}(c(k-1) \leq |Z| < ck)(1 + o(1)).$$

We now claim that this identity holds uniformly for all $1 \leq j \leq n$. This does not follow from the formulation of Corollary 3.4, but instead follows from the second part of [15, Corollary 1.3], which gives a uniform bound on the convergence of moments. Indeed, uniform rates of convergence follow by combining the uniform convergence of moments with the identity

$$\left| e^{itX} - \sum_{l=0}^{s-1} \frac{(it)^l X^l}{l!} \right| \leq \frac{|t|^s |X|^s}{s!}$$

and the inequality from [39, page 538]. Therefore, we obtain that

$$\mathbb{E}N(c, k) = (1 + o(1))f(c, k).$$

In addition, by a similar argument, it follows that

$$\mathrm{Var}(N(c, k)) = o(n^2).$$

By Chebyshev's inequality, we have that, for any $\varepsilon > 0$,

$$\mathbb{P}\left(|N(c, k) - \mathbb{E}N(c, k)| \geq \varepsilon f(c, k)\right) \leq \frac{\mathrm{Var}(N(c, k))}{\varepsilon^2 f(c, k)^2}.$$

We note that $f(c, k) \geq c'n$ for some constant $c' > 0$ depending on c and k. Hence, we conclude that

$$N(c, k) = (1 + o(1))f(c, k) \tag{35}$$

with probability $1 - o(1)$.

We will return to (35) in a moment. We now approximate the distribution of $\min_{S \subset [n]:|S|=\lfloor \delta n \rfloor} \|v\|_S^2$. We will first need to determine the value of c from above and another parameter k_0. Take $c > 0$ and $k_0 \in \mathbb{N}$ such that

$$2c \left[\Phi^{-1} \left(\frac{1+\delta}{2} \right) \right]^2 < \frac{\varepsilon}{2} \tag{36}$$

and

$$ck_0 = \Phi^{-1} \left(\frac{1+\delta}{2} \right). \tag{37}$$

Such choices are always possible by taking $c > 0$ sufficiently small such that (36) holds, and then (by possibly decreasing c if necessary) choosing $k_0 \in \mathbb{N}$ which satisfies (37).

There are several important implications of these choices. First,

$$2 \sum_{k=1}^{k_0} \left(\Phi(ck) - \Phi((c(k-1))) \right) = 2 \left(\Phi(ck_0) - \Phi(0) \right) = \delta. \tag{38}$$

Second, we have

$$\left| 2 \sum_{k=1}^{k_0} c^2(k-1)^2 \left(\Phi(ck) - \Phi(c(k-1)) \right) - 2 \int_0^{\Phi^{-1}\left(\frac{1+\delta}{2}\right)} x^2 \Phi'(x) \, dx \right|$$

$$= 2 \left| \sum_{k=1}^{k_0} \left[c^2(k-1)^2 \left(\Phi(ck) - \Phi(c(k-1)) \right) - \int_{c(k-1)}^{ck} x^2 \Phi'(x) \, dx \right] \right|$$

$$= 2 \left| \sum_{k=1}^{k_0} \left[(c^2(k-1)^2 - c^2k^2)\Phi(ck) + 2 \int_{c(k-1)}^{ck} x \Phi(x) \, dx \right] \right| \tag{39}$$

$$= 4 \left| \sum_{k=1}^{k_0} \int_{c(k-1)}^{ck} x \left(\Phi(x) - \Phi(ck) \right) \, dx \right|$$

$$\leq 2c \left[\Phi^{-1} \left(\frac{1+\delta}{2} \right) \right]^2$$

$$< \frac{\varepsilon}{2}$$

by integration by parts and (36). Here the first inequality follows from the mean value theorem and the bound $0 \leq \Phi'(x) \leq 1$. By a similar argument, we obtain

$$\left| 2 \sum_{k=1}^{k_0} c^2 k^2 \left(\Phi(ck) - \Phi(c(k-1)) \right) - 2 \int_0^{\Phi^{-1}\left(\frac{1+\delta}{2}\right)} x^2 \Phi'(x) \, dx \right| < \frac{\varepsilon}{2}. \tag{40}$$

From (35), we have that for any $1 \leq k \leq k_0 + 1$,

$$N(c,k) = (1 + o(1)) f(c,k) = 2n(1 + o(1)) \left(\Phi(ck) - \Phi(c(k-1)) \right) \tag{41}$$

with probability $1 - o(1)$. In view of (38), we have

$$\sum_{k=1}^{k_0} f(c,k) = \delta n.$$

Thus, by the union bound, we obtain that, with probability $1 - o(1)$,

$$\sum_{k=1}^{k_0} N(c,k) = (1 + o(1)) \delta n = \lfloor \delta n \rfloor + o(n).$$

We observe that

$$\sum_{k=1}^{k_0} c^2 (k-1)^2 N(c,k) \leq n \min_{S \subset [n] : |S| = \sum_{k=1}^{k_0} N(c,k)} \|v\|_S^2 \leq \sum_{k=1}^{k_0} c^2 k^2 N(c,k)$$

by definition of $N(c,k)$. Thus, with probability $1 - o(1)$, there exists a sequence $(\varepsilon_n)_{n \geq 1}$ with $\varepsilon_n \searrow 0$ such that

$$\sum_{k=1}^{k_0} c^2 (k-1)^2 N(c,k) - \varepsilon_n c^2 k_0^2 N(c,k_0)$$

$$\leq n \min_{S \subset [n] : |S| = \lfloor \delta n \rfloor} \|v\|_S^2 \leq \sum_{k=1}^{k_0} c^2 k^2 N(c,k) + \varepsilon_n c^2 (k_0+1)^2 N(c,k_0+1).$$

Applying (41), we find that

$$2 \sum_{k=1}^{k_0} c^2 (k-1)^2 (\Phi(ck) - \Phi(c(k-1)))(1 + o(1))$$

$$\leq \min_{S \subset [n] : |S| = \lfloor \delta n \rfloor} \|v\|_S^2 \leq 2 \sum_{k=1}^{k_0} c^2 k^2 (\Phi(ck) - \Phi(c(k-1)))(1 + o(1))$$

with probability $1 - o(1)$. Therefore, by (39) and (40), we conclude that

$$\left| \min_{S \subset [n]: |S| = \lfloor \delta n \rfloor} \|v\|_S^2 - 2 \int_0^{\Phi^{-1}\left(\frac{1+\delta}{2}\right)} x^2 \Phi'(x)\, dx \right| \leq \varepsilon$$

with probability $1 - o(1)$, and the proof is complete.

13.2. Proof of Corollary 5.4

Before proving Corollary 5.4, we will need to establish the following bound.

Theorem 13.1 *(Sub-gaussian entries: Lower bound). Let ξ, ζ be real sub-gaussian random variables with mean zero, and assume ξ has unit variance. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Then there exist constants $C, c > 0$ and $0 < \eta, \delta < 1$ (depending only on the sub-gaussian moments of ξ and ζ) such that*

$$\min_{1 \leq j \leq n} \min_{S \subset [n]: |S| \geq \delta n} \|v_j(W)\|_S \geq \eta$$

with probability at least $1 - C \exp(-cn)$.

Theorem 13.1 follows directly from Theorem 5.2. However, Theorem 5.2 is much stronger than anything we need here. Additionally, the proof of Theorem 5.2 is long and complicated, and we do not touch on it in this survey. Instead, we provide a separate proof of Theorem 13.1 using the uniform bounds below.

Theorem 13.2 *(Uniform upper bound). Let ξ, ζ be real random variables, and let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Take $1 \leq m < n$. Then, for any $0 < \eta < 1$ and $K > 0$,*

$$\mathbb{P}\left(\exists j \in [n] \text{ and } S \subset [n] \text{ with } |S| = m \text{ such that } \|v_j(W)\|_S^2 \geq \eta \right)$$

$$\leq \mathbb{P}\left(\|W\| > K\sqrt{n} \right) + n \binom{n}{m} \mathbb{P}\left(\frac{1}{n} \sigma_m^2(B) \leq \frac{4(1-\eta)K^2}{\eta} \right),$$

where B is a $(n-m) \times m$ matrix whose entries are iid copies of ξ.

Proof. We first observe that, by changing η to $1 - \eta$, it suffices to show

$$\mathbb{P}\left(\exists j \in [n] \text{ and } S \subset [n] \text{ with } |S| = m \text{ such that } \|v_j(W)\|_S^2 \geq 1 - \eta \right)$$

$$\leq \mathbb{P}\left(\|W\| > K\sqrt{n} \right) + n \binom{n}{m} \mathbb{P}\left(\frac{1}{n} \sigma_m^2(B) \leq \frac{4\eta K^2}{1-\eta} \right).$$

For notational convenience, let $\lambda_1 \leq \cdots \leq \lambda_n$ denote the eigenvalues of W with corresponding unit eigenvectors $v_1, \ldots v_n$. Define the event

$$\Omega_{n,m}(\eta) := \left\{ \exists j \in [n] \text{ and } S \subset [n] \text{ with } |S| = m \text{ such that } \|v_j\|_S^2 \geq 1 - \eta \right\}.$$

Then

$$\mathbb{P}(\Omega_{n,m}(\eta)) \leq \mathbb{P}\left(\Omega_{n,m}(\eta) \cap \{\|W\| \leq K\sqrt{n}\}\right) + \mathbb{P}\left(\|W\| > K\sqrt{n}\right). \tag{42}$$

By the union bound and symmetry, we have

$$\mathbb{P}\left(\Omega_{n,m}(\eta) \cap \{\|W\| \leq K\sqrt{n}\}\right)$$

$$\leq \sum_{j=1}^n \mathbb{P}\left(\exists S \subset [n] \text{ with } |S| = m \text{ such that } \|v_j\|_{S^c}^2 \leq \eta \text{ and } \|W\| \leq K\sqrt{n}\right)$$

$$\leq \binom{n}{m} \sum_{j=1}^n \mathbb{P}\left(\sum_{k=m+1}^n |v_j(k)|^2 \leq \eta \text{ and } \|W\| \leq K\sqrt{n}\right),$$

where $v_j(k)$ denotes the kth entry of the unit vector v_j.

Write

$$W = \begin{pmatrix} A & B^{\mathrm{T}} \\ B & D \end{pmatrix},$$

where A is a $m \times m$ matrix, B is a $(n-m) \times m$ matrix, and D is a $(n-m) \times (n-m)$ matrix. In particular, the entries of B are iid copies of ξ. Decompose

$$v_j = \begin{pmatrix} x_j \\ y_j \end{pmatrix},$$

where x_j is a m-vector and y_j is a $(n-m)$-vector. Then the eigenvalue equation $Wv_j = \lambda_j v_j$ implies that $Bx_j + Dy_j = \lambda_j y_j$. Therefore, on the event where $\|y_j\|^2 \leq \eta$ and $\|W\| \leq K\sqrt{n}$, we have

$$\|Bx_j\|^2 = \|\lambda_j y_j - Dy_j\|^2 \leq 2\|y_j\|^2 \left(|\lambda_j|^2 + \|D\|^2\right) \leq 4\eta K^2 n.$$

Here we used the fact that the spectral norm of a matrix is not less than that of any sub-matrix. Since $\|x_j\|^2 + \|y_j\|^2 = 1$, we find that

$$(1-\eta)\sigma_m^2(B) \leq \|Bx_j\|^2.$$

Combining the bounds above, we conclude that

$$\mathbb{P}\left(\Omega_{n,m}(\eta) \cap \{\|W\| \le K\sqrt{n}\}\right) \le \binom{n}{m} \sum_{j=1}^{n} \mathbb{P}\left(\|y_j\|^2 \le \eta \text{ and } \|W\| \le K\sqrt{n}\right)$$

$$\le \binom{n}{m} \sum_{j=1}^{n} \mathbb{P}\left((1-\eta)\sigma_m^2(B) \le 4\eta K^2 n\right)$$

$$\le n\binom{n}{m}\mathbb{P}\left(\frac{1}{n}\sigma_m^2(B) \le \frac{4\eta K^2}{1-\eta}\right).$$

The proof is now complete by combining the bound above with (42). □

Theorem 13.3 *(Uniform lower bound). Let ξ, ζ be real random variables, and let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Take $1 \le m < n$. Then, for any $0 < \eta < 1$ and $K > 0$,*

$$\mathbb{P}\left(\exists j \in [n] \text{ and } S \subset [n] \text{ with } |S| = m \text{ such that } \|v_j(W)\|_S^2 \le \eta\right)$$

$$\le \mathbb{P}\left(\|W\| > K\sqrt{n}\right) + n\binom{n}{m}\mathbb{P}\left(\frac{1}{n}\sigma_{n-m}^2(B) \le \frac{4\eta K^2}{1-\eta}\right),$$

where B is a $m \times (n-m)$ matrix whose entries are iid copies of ξ.

Proof. From the relation

$$\|v_j(W)\|_S^2 + \|v_j(W)\|_{S^c}^2 = 1,$$

we find that

$$\mathbb{P}\left(\exists j \in [n] \text{ and } S \subset [n] \text{ with } |S| = m \text{ such that } \|v_j(W)\|_S^2 \le \eta\right)$$

$$= \mathbb{P}\left(\exists j \in [n] \text{ and } S \subset [n] \text{ with } |S| = n - m \text{ such that } \|v_j(W)\|_S^2 \ge 1 - \eta\right).$$

Therefore, Theorem 13.3 follows immediately from Theorem 13.2. □

Before proving Theorem 13.1, we prove the following upper bound.

Theorem 13.4 *(Sub-gaussian entries: Upper bound). Let ξ, ζ be sub-gaussian random variables with mean zero, and assume ξ has unit variance. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Then there exist constants $C, c > 0$ and $0 < \eta, \delta < 1$ (depending only on the sub-gaussian moments of ξ and ζ) such that*

$$\max_{1 \le j \le n} \max_{S \subset [n]: |S| \le \delta n} \|v_j(W)\|_S \le \eta$$

with probability at least $1 - C\exp(-cn)$.

Proof. We observe that it suffices to show

$$\max_{1 \leq j \leq n} \max_{S \subset [n]: |S| = \lfloor \delta n \rfloor} \|v_j(W)\|_S \leq \eta$$

with probability at least $1 - Cn \exp(-cn)$. Moreover, in view of Theorem 13.2 and Lemma 11.3, it suffices to show there exists $0 < \eta, \delta < 1$ such that

$$n \binom{n}{\lfloor \delta n \rfloor} \mathbb{P} \left(\frac{1}{n} \sigma^2_{\lfloor \delta n \rfloor}(B) \leq \frac{4 C_0^2 (1 - \eta)}{\eta} \right) \leq Cn \exp(-cn),$$

where B is a $(n - \lfloor \delta n \rfloor) \times \lfloor \delta n \rfloor$ matrix whose entries are iid copies of ξ.

Fix $0 < c_0 < 1$. Let $c > 0$ be the corresponding constant from Theorem 11.7 which depends only on the sub-gaussian moment of ξ. Let $0 < \delta < 1/2$ be such that

$$\sqrt{n - \lfloor \delta n \rfloor} - \sqrt{\lfloor \delta n \rfloor - 1} \geq c_0 \sqrt{n},$$

$$n - 2 \lfloor \delta n \rfloor + 1 \geq c_0 n,$$

and

$$\delta \log \left(\frac{2e}{\delta} \right) \leq c_0 \frac{c}{2}. \tag{43}$$

We observe that such a choice is always possible since $\delta' \log(1/\delta') \to 0$ as $\delta' \downarrow 0$. Set $m := \lfloor \delta n \rfloor$.

By Theorem 11.7 and our choice of δ, there exists $c_1 > 0$ (depending on c_0, c, and the sub-gaussian moments of ξ) such that

$$\mathbb{P} \left(\sigma_m^2(B) \leq c_1 n \right) \leq 2 \exp(-c_0 c n).$$

Taking $0 < \eta < 1$ sufficiently close to 1 (so that $\frac{4 C_0^2 (1 - \eta)}{\eta} < c_1$), we obtain

$$n \binom{n}{m} \mathbb{P} \left(\frac{1}{n} \sigma_m^2(B) \leq \frac{4 C_0^2 (1 - \eta)}{\eta} \right) \leq 2n \binom{n}{m} \exp(-c_0 c n)$$

$$\leq 2n \left(\frac{ne}{m} \right)^m \exp(-c_0 c n)$$

$$\leq 2n \exp \left(m \log \left(\frac{ne}{m} \right) - c_0 c n \right).$$

The claim now follows since

$$m \log \left(\frac{ne}{m} \right) \leq \delta n \log \left(\frac{2e}{\delta} \right) \leq c_0 \frac{c}{2} n$$

by (43) for n sufficiently large. $\square$

Using Theorem 13.4, we immediately obtain the proof of Theorem 13.1.

Proof of Theorem 13.1. Theorem 13.1 follows immediately from Theorem 13.4 and the relation

$$\min_{\substack{1 \leq j \leq n}} \min_{\substack{S \subset [n] \\ |S| \geq m}} \|v_j(W)\|_S^2 + \max_{\substack{1 \leq j \leq n}} \max_{\substack{S \subset [n] \\ |S| \leq n-m}} \|v_j(W)\|_S^2 = 1.$$

We also note that one can prove Theorem 13.1 directly by applying Theorem 13.3. $\square$

We now present the proof of Corollary 5.4. Indeed, Corollary 5.4 follows immediately from Theorem 13.1, Theorem 13.4, and the two lemmas below.

Lemma 13.5 (*Uniform lower bound implies upper bound on the ℓ^p-norm*). *Let v be a unit vector in $\mathbb{R}^n$, and let $1 \leq m \leq n$. If*

$$\min_{S \subset [n]:|S|=m} \|v\|_S^2 \geq \eta,$$

then, for any $1 \leq p \leq 2$,

$$\|v\|_{\ell^p}^p \leq \left\lceil \frac{n}{n-m} \right\rceil \frac{1-\eta}{\eta^{1-p/2}} m^{1-p/2}.$$

Proof. Let $S \subset [n]$ be the subset which minimizes $\|v\|_S^2$ given the constraint $|S| = m$. In particular, S contains the m-smallest coordinates of $v = (v_i)_{i=1}^n$ in absolute value. Let l be such that

$$|v_l| = \max\{|v_j| : j \in S\}.$$

Then

$$|S||v_l|^2 \geq \sum_{j \in S} |v_j|^2 \geq \eta$$

by assumption. Hence, we obtain

$$|v_l|^2 \geq \frac{\eta}{|S|}.$$

Therefore, we conclude that, for $1 \leq p \leq 2$,

$$1 - \eta \geq \sum_{j \in S^c} |v_j|^2 \geq |v_l|^{2-p} \sum_{j \in S^c} |v_j|^p \geq \left(\frac{\eta}{|S|} \right)^{1-p/2} \sum_{j \in S^c} |v_j|^p.$$

Since $|S| = m$, and S was the minimizer, we have

$$\max_{T \subset [n]:|T|=n-m} \sum_{j \in T} |v_j|^p \le \frac{1-\eta}{\eta^{1-p/2}} m^{1-p/2}.$$

Let $T_1, \ldots, T_{k_0}$ be a partition of $[n]$ such that $|T_k| \le n - m$ for $1 \le k \le k_0$, and $k_0 \le \left\lceil \frac{n}{n-m} \right\rceil$. We then find that

$$\sum_{j=1}^{n} |v_j|^p = \sum_{k=1}^{k_0} \sum_{j \in T_k} |v_j|^p \le \left\lceil \frac{n}{n-m} \right\rceil \frac{1-\eta}{\eta^{1-p/2}} m^{1-p/2},$$

and the proof is complete. □

Lemma 13.6 (*Uniform upper bound implies lower bound on the ℓ^p-norm*). *Let v be a unit vector in $\mathbb{R}^n$, and let $1 \le m \le n$. If*

$$\max_{S \subset [n]:|S|=m} \|v\|_S^2 \le \eta,$$

then, for any $1 \le p \le 2$,

$$\|v\|_{\ell^p}^p \ge \frac{1-\eta}{\eta^{1-p/2}} m^{1-p/2}.$$

Proof. Let $S \subset [n]$ be the subset which maximizes $\|v\|_S^2$ given the constraint $|S| = m$. In particular, S contains the m-largest coordinates of $v = (v_i)_{i=1}^n$ in absolute value. Let l be such that

$$|v_l| = \min\{|v_j| : j \in S\}.$$

Then

$$|S||v_l|^2 \le \sum_{j \in S} |v_j|^2 \le \eta$$

by assumption. Hence, we obtain

$$|v_l|^2 \le \frac{\eta}{|S|}.$$

Therefore, we conclude that, for $1 \le p \le 2$,

$$1 - \eta \le \sum_{j \in S^c} |v_j|^2 \le |v_l|^{2-p} \sum_{j \in S^c} |v_j|^p \le \left(\frac{\eta}{|S|} \right)^{1-p/2} \sum_{j=1}^{n} |v_j|^p.$$

Since $|S| = m$, the proof is complete. □

14. Proofs for random matrices with non-zero mean

This section is devoted to the proofs of Theorems 6.4, 6.6, 6.8 and 6.10.

14.1. Proof of Theorem 6.4

The proof of Theorem 6.4 is similar to the proof of Theorem 4.7. We begin with a result from [54].

Theorem 14.1 *([54]). Let A_n be the adjacency matrix of $G(n, p)$ for some fixed value of $p \in (0, 1)$. Let*

$$A_n = \begin{bmatrix} A_{n-1} & X \\ X^{\mathrm{T}} & 0 \end{bmatrix},$$

where A_{n-1} is the upper-left $(n - 1) \times (n - 1)$ minor of A_n and $X \in \{0, 1\}^{n-1}$. Then, for any $\alpha > 0$, there exists a constant C (depending on p, α) such that, for any $\delta > n^{-\alpha}$,

$$\sup_{1 \le i \le n-1} \mathbb{P}\left(\lambda_i(A_{n-1}) - \lambda_i(A_n) \le \frac{\delta}{\sqrt{n}} \right) \le C n^{o(1)} \delta + o(1)$$

and

$$\sup_{2 \le i \le n} \mathbb{P}\left(\lambda_i(A_n) - \lambda_{i-1}(A_{n-1}) \le \frac{\delta}{\sqrt{n}} \right) \le C n^{o(1)} \delta + o(1).$$

Moreover,

$$\inf_{1 \le j \le n-1} |v_j(A_{n-1})^{\mathrm{T}} X| > 0$$

with probability $1 - o(1)$. Here, the rate of convergence to zero implicit in the $o(1)$ terms depends on p and α.

Theorem 14.1 follows from [54, Theorem 2.7] and the arguments given in [54, Section 7] and [81]. We will also need the following version of Theorem 11.5 generalized to the adjacency matrix $A_n(p)$.

Lemma 14.2. *Let $A_n(p)$ be the adjacency matrix of $G(n, p)$ for some fixed value of $p \in (0, 1)$. Let N_I denote the number of eigenvalues of $\frac{1}{\sqrt{np(1-p)}} A_n(p)$ in the interval I. Then there exist constants $C, c, c' > 0$ (depending on p) such that, for any interval $I \subset \mathbb{R}$,*

$$\mathbb{P}\left(N_I \ge \frac{(1.1)}{\pi} n|I| + (\log n)^{c'} \right) \le C \exp(-c \log^2 n),$$

where $|I|$ denotes the length of I.

Proof. We observe that $A_n(p)$ has the same distribution as $W_n + p(\mathbf{J}_n - I_n)$, where W_n is the Wigner matrix whose diagonal entries are zero and whose upper-triangular entries are iid copies of the random variable defined in (2), $\mathbf{J}_n$ is the all-ones matrix, and I_n is the identity matrix. Thus, it suffices to prove the result for $W_n + p(\mathbf{J}_n - I_n)$. Accordingly, redefine N_I to be the number of eigenvalues of $\frac{1}{\sqrt{np(1-p)}}(W_n + p(\mathbf{J}_n - I_n))$ in the interval I.

For $\gamma \in \mathbb{R}$ and any interval I, we let $I + \gamma$ denote the interval I shifted to the right by γ units. Clearly, both I and $I + \gamma$ have the same length, i.e. $|I| = |I + \gamma|$. Set $\gamma := p/\sqrt{np(1-p)}$, and let $I \subset \mathbb{R}$ be any interval. It follows from Theorem 11.5 that the number of eigenvalues of $\frac{1}{\sqrt{np(1-p)}}W_n$ in the interval $I + \gamma$ is at most

$$(1.1)n \int_{I+\gamma} \rho_{\mathrm{sc}}(x)dx + (\log n)^{c'} \le \frac{1.1}{\pi}n|I| + (\log n)^{c'}$$

with probability at least $1 - C\exp(-c\log^2 n)$. The inequality above follows from bounding the semicircle density ρ_{sc} by $\frac{1}{\pi}$ (and the fact that the intervals I and $I + \gamma$ have the same length). The normalization factor $\sqrt{p(1-p)}$ ensures the entries of the matrix have unit variance as required by Theorem 11.5.

Since $\mathbf{J}_n$ is rank one, it follows from eigenvalue interlacing (see, for instance, [11, Exercise III.2.4]) that the number of eigenvalues of $\frac{1}{\sqrt{np(1-p)}}(W_n + p\mathbf{J}_n)$ in the interval $I + \gamma$ is at most

$$\frac{1.1}{\pi}n|I| + (\log n)^{c'} + 1 \le \frac{1.1}{\pi}n|I| + (\log n)^{c'+1}$$

for n sufficiently large. Subtracting γI_n from a matrix only shifts the eigenvalues of the matrix by γ. Hence, we conclude that

$$N_I \le \frac{1.1}{\pi}n|I| + (\log n)^{c'+1}$$

with probability at least $1 - C\exp(-c\log^2 n)$. $\square$

We will also need the following technical lemma.

Lemma 14.3. *Let A_n be the adjacency matrix of $G(n, p)$ for some fixed value of $p \in (0, 1)$. Let*

$$A_n = \begin{bmatrix} A_{n-1} & X \\ X^{\mathrm{T}} & 0 \end{bmatrix},$$

where A_{n-1} is the upper-left $(n-1) \times (n-1)$ minor of A_n and $X \in \{0, 1\}^{n-1}$. Let $\kappa > 0$. Then there exists a constant $c_1 > 0$ such that, for any $1 \le i \le n-1$ and any $e^{-n^\kappa} < \delta < 1$,

$$\sum_{j=1}^{n-1} \frac{|v_j(A_{n-1})^{\mathrm{T}} X|^2}{|\lambda_j(A_{n-1}) - \lambda_i(A_n)|^2} \leq \frac{(\log n)^{c_1}}{25} \left[\frac{1}{m_i^2} + \frac{n}{\delta^2} \right] \tag{44}$$

with probability $1 - o(1)$, where

$$m_i := \min_{1 \leq j \leq n-1} |\lambda_j(A_{n-1}) - \lambda_i(A_n)|.$$

Here, the rate of convergence to zero implicit in the $o(1)$ term depends on p and κ.

Proof. Fix $1 \leq i \leq n-1$ and $e^{-n^{\kappa}} < \delta < 1$. Define the event Ω to be the intersection of the events

$$\bigcap_{j=1}^{n-2} \left\{ |v_j(A_{n-1})^{\mathrm{T}} X|^2 \leq \frac{p(1-p)}{1000} \log^2 n \right\}, \tag{45}$$

$$\left\{ \inf_{1 \leq j \leq n-1} |v_j(A_{n-1})^{\mathrm{T}} X|^2 > 0 \right\}, \tag{46}$$

and

$$\{ \|A_n\| \leq 10np \}. \tag{47}$$

We claim that

$$\mathbb{P}(\Omega) = 1 - o(1). \tag{48}$$

Note that the event in (47) can be shown to hold with probability $1 - o(1)$ by combining the bounds in [40] with Hoeffding's inequality. Additionally, the event in (46) holds with probability $1 - o(1)$ by Theorem 14.1. We now consider the events in (45). For $1 \leq j \leq n-2$, by the Cauchy–Schwarz inequality, we obtain

$$
\begin{aligned}
|v_j(A_{n-1})^{\mathrm{T}} X| &\leq |v_j(A_{n-1})^{\mathrm{T}}(X - p\mathbf{1})| + |v_j(A_{n-1})^{\mathrm{T}}(p\mathbf{1} - pn^{1/2} v_{n-1}(A_{n-1}))| \\
&\leq |v_j(A_{n-1})^{\mathrm{T}}(X - p\mathbf{1})| + pn^{1/2} \|n^{-1/2}\mathbf{1} - v_{n-1}(A_{n-1})\| \\
&\leq |v_j(A_{n-1})^{\mathrm{T}}(X - p\mathbf{1})| + pn \|n^{-1/2}\mathbf{1} - v_{n-1}(A_{n-1})\|_{\ell^{\infty}}
\end{aligned}
$$

since $v_j(A_{n-1})$ is orthogonal to $v_{n-1}(A_{n-1})$. By Theorem 6.1, it follows that

$$pn \|n^{-1/2}\mathbf{1} - v_{n-1}(A_{n-1})\|_{\ell^{\infty}} \leq C_0 \sqrt{\log n}$$

with probability $1 - o(1)$. We now observe that the vector $X - p\mathbf{1}$ has mean zero and $|v_j(A_{n-1})^{\mathrm{T}}(X - p\mathbf{1})|$ is the length of the projection of the vector $X - p\mathbf{1}$ onto the one-dimensional subspace spanned by $v_j(A_{n-1})$. Hence, by Corollary 11.9 and the union bound, it follows that

$$\sup_{1 \leq j \leq n-2} |v_j(A_{n-1})^{\mathrm{T}}(X - p\mathbf{1})|^2 \leq \frac{p(1-p)}{1000^2} \log^2 n$$

with probability $1 - o(1)$. Combining the estimates above yields the bound in (48).

It now suffices to show that, conditionally on Ω, the bound in (44) holds with probability $1 - o(1)$. It follows (see Remark 12.2 for details) that, on Ω, the eigenvalues of A_{n-1} strictly interlace with the eigenvalues of A_n, and hence the terms

$$\sum_{j=1}^{n-1} \frac{|v_j(A_{n-1})^{\mathrm{T}}X|^2}{|\lambda_j(A_{n-1}) - \lambda_i(A_n)|^2}$$

and m_i^{-2} are well-defined.

It follows from the results in [40] that

$$|\lambda_{n-1}(A_{n-1}) - \lambda_i(A_n)| \geq \frac{1}{100} np$$

with probability $1 - o(1)$. Hence, we obtain

$$\frac{|v_{n-1}(A_{n-1})^{\mathrm{T}}X|^2}{|\lambda_{n-1}(A_{n-1}) - \lambda_i(A_n)|^2} \leq \frac{100^2 \|X\|^2}{n^2 p^2} \leq \frac{100^2}{np^2}.$$

Thus, on Ω, we have

$$\sum_{j=1}^{n-1} \frac{|v_j(A_{n-1})^{\mathrm{T}}X|^2}{|\lambda_j(A_{n-1}) - \lambda_i(A_n)|^2} \leq \frac{p(1-p)}{1000} \log^2 n \sum_{j=1}^{n-2} \frac{1}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|^2} + \frac{100^2}{np^2}$$

with probability $1 - o(1)$. Therefore, conditionally on Ω, it suffices to show that

$$\sum_{j=1}^{n-2} \frac{1}{|\lambda_j(A_{n-1}) - \lambda_i(A_n)|^2} \leq \frac{10}{p(1-p)} (\log n)^{c_1} \left[\frac{1}{m_i^2} + \frac{n}{\delta^2} \right]$$

with probability $1 - o(1)$.

Define the sets

$$T := \left\{ 1 \leq j \leq n-2 : \frac{|\lambda_j(A_{n-1}) - \lambda_i(A_n)|}{\sqrt{p(1-p)}} < \frac{\delta}{\sqrt{n}} \right\}$$

and

$$T_l := \left\{ 1 \leq j \leq n-2 : 2^l \frac{\delta}{\sqrt{n}} \leq \frac{|\lambda_j(A_{n-1}) - \lambda_i(A_n)|}{\sqrt{p(1-p)}} < 2^{l+1} \frac{\delta}{\sqrt{n}} \right\}$$

for $l = 0, 1, \ldots, L$, where L is the smallest integer such that $2^L \sqrt{p(1-p)} \frac{\delta}{\sqrt{n}} \geq 20np$. In particular, as $\delta > e^{-n^\kappa}$, we obtain $L = O(n^{-\kappa})$. On the event Ω, it follows that every

index $1 \leq j \leq n - 2$ is contained in either T or $\cup_{l=0}^{L} T_l$. By Lemma 14.2 and the union bound, there exist $C, c, c_1 > 0$ such that

$$|T| \leq (\log n)^{c_1}$$

and

$$|T_l| \leq 4(2^l) + (\log n)^{c_1}, \quad l = 0, \ldots, L$$

with probability at least $1 - C \exp(-c \log^2 n)$. Hence, on this same event, we conclude that

$$\sum_{j=1}^{n-2} \frac{p(1-p)}{|\lambda_j(A_{n-1}) - \lambda_i(A_n)|^2} \leq \sum_{j \in T} \frac{p(1-p)}{|\lambda_j(A_{n-1}) - \lambda_i(A_n)|^2}$$

$$+ \sum_{l=0}^{L} \sum_{j \in T_l} \frac{p(1-p)}{|\lambda_j(A_{n-1}) - \lambda_i(A_n)|^2}$$

$$\leq \frac{|T|}{m_i^2} + \frac{n}{\delta^2} \sum_{l=0}^{L} \frac{|T_l|}{2^{2l}}$$

$$\leq \frac{(\log n)^{c_1}}{m_i^2} + \frac{10 n (\log n)^{c_1}}{\delta^2}$$

for n sufficiently large. The proof of the lemma is complete. $\square$

With Theorem 14.1 and Lemma 14.3 in hand, we are now ready to prove Theorem 6.4.

Proof of Theorem 6.4. Let A_n be the adjacency matrix of $G(n, p)$. We will bound the jth coordinate of the unit eigenvector v_i in magnitude from below. By symmetry, it suffices to consider the case when $j = n$. For this reason, we decompose A_n as

$$A_n = \begin{bmatrix} A_{n-1} & X \\ X^{\mathrm{T}} & 0 \end{bmatrix},$$

where A_{n-1} is the upper-left $(n-1) \times (n-1)$ minor of A_n and $X \in \{0, 1\}^{n-1}$.

Let $1 \leq i \leq n - 1$. Let $\alpha > 0$, and assume $1 > \delta > n^{-\alpha}$ as the claim is trivial when $\delta \geq 1$. Define the event Ω to be the intersection of the events

$$\left\{ \inf_{1 \leq j \leq n-1} |v_j(A_{n-1})^{\mathrm{T}} X| > 0 \right\} \bigcap \left\{ m_i > \frac{\delta}{\sqrt{n}} \right\}$$

and

$$\left\{ \sum_{j=1}^{n-1} \frac{|v_j(A_{n-1})^{\mathrm{T}} X|^2}{|\lambda_j(A_{n-1}) - \lambda_i(A_n)|^2} \leq \frac{(\log n)^{c_1}}{25} \left[\frac{1}{m_i^2} + \frac{n}{\delta^2} \right] \right\},$$

where

$$m_i := \min_{1 \le j \le n-1} |\lambda_j(A_{n-1}) - \lambda_i(A_n)|$$

and c_1 is the constant from Lemma 14.3. If $m_i \le \frac{\delta}{\sqrt{n}}$, then, by Cauchy's interlacing inequalities (21), this implies that either

$$|\lambda_i(A_{n-1}) - \lambda_i(A_n)| \le \frac{\delta}{\sqrt{n}}$$

or

$$|\lambda_{i-1}(A_{n-1}) - \lambda_i(A_n)| \le \frac{\delta}{\sqrt{n}}.$$

(Here, the second possibility can only occur if $i > 1$.) Therefore, it follows from Theorem 14.1 and Lemma 14.3 that, there exists $C > 0$ such that Ω holds with probability at least $1 - C n^{o(1)} \delta - o(1)$.

Let $c_2 := c_1/2$. Then

$$\mathbb{P}\left(|v_i(n)| \le \frac{\delta}{\sqrt{n}(\log n)^{c_2}} \right) \le \mathbb{P}(\Omega^c) + \mathbb{P}\left(\left\{ |v_i(n)| \le \frac{\delta}{\sqrt{n}(\log n)^{c_2}} \right\} \bigcap \Omega \right).$$

Hence, to complete the proof, we will show that the event

$$\left\{ |v_i(n)| \le \frac{\delta}{\sqrt{n}(\log n)^{c_2}} \right\} \bigcap \Omega \tag{49}$$

is empty.

Fix a realization in this event. As discussed above (see Remark 12.2), on Ω, the eigenvalues of A_{n-1} strictly interlace with the eigenvalues of A_n. Hence, we apply Lemma 11.2 and obtain

$$\frac{1}{1 + \sum_{j=1}^{n-1} \frac{|v_j(A_{n-1})^{\mathrm{T}} X|^2}{|\lambda_j(A_{n-1}) - \lambda_i(A_n)|^2}} = |v_i(n)|^2 \le \frac{\delta^2}{n(\log n)^{2c_2}}.$$

This implies, for n sufficiently large, that

$$\frac{n(\log n)^{2c_2}}{2\delta^2} \le \sum_{j=1}^{n-1} \frac{|v_j(A_{n-1})^{\mathrm{T}} X|^2}{|\lambda_j(A_{n-1}) - \lambda_i(A_n)|^2}.$$

On the other hand, by definition of Ω, we have

$$\sum_{j=1}^{n-1} \frac{|v_j(A_{n-1})^{\mathrm{T}} X|^2}{|\lambda_j(A_{n-1}) - \lambda_i(A_n)|^2} \le \frac{2(\log n)^{c_1}}{25} \frac{n}{\delta^2},$$

and thus

$$\frac{n(\log n)^{2c_2}}{2\delta^2} \leq \frac{2(\log n)^{c_1}}{25} \frac{n}{\delta^2}.$$

This is a contradiction since $2c_2 = c_1$. We conclude that the event in (49) is empty, and the proof is complete. $\square$

14.2. Proof of Theorem 6.6

To begin, we introduce ε-nets as a convenient way to discretize a compact set.

Definition 14.4 (*ε-net*). Let (X, d) be a metric space, and let $\varepsilon > 0$. A subset $\mathcal{N}$ of X is called an ε-net of X if every point $x \in X$ can be approximated to within ε by some point $y \in \mathcal{N}$. That is, for every $x \in X$ there exists $y \in \mathcal{N}$ so that $d(x, y) \leq \varepsilon$.

The following estimate for the maximum size of an ε-net of a sphere is well-known.

Lemma 14.5. *A unit sphere in d dimensions admits an ε-net of size at most*

$$\left(1 + \frac{2}{\varepsilon}\right)^d.$$

Proof. Let S be the unit sphere in question. Let $\mathcal{N}$ be a maximal ε-separated subset of S. That is, $\|x - y\| \geq \varepsilon$ for all distinct $x, y \in \mathcal{N}$ and no subset of S containing $\mathcal{N}$ has this property. Such a set can always be constructed by starting with an arbitrary point in S and at each step selecting a point that is at least ε distance away from those already selected. Since S is compact, this procedure will terminate after a finite number of steps.

We now claim that $\mathcal{N}$ is an ε-net. Suppose to the contrary. Then there would exist $x \in S$ that is at least ε from all points in $\mathcal{N}$. In other words, $\mathcal{N} \cup \{x\}$ would still be an ε-separated subset of S. This contradicts the maximal assumption above.

We now proceed by a volume argument. At each point of $\mathcal{N}$ we place a ball of radius $\varepsilon/2$. By the triangle inequality, it is easy to verify that all such balls are disjoint and lie in the ball of radius $1 + \varepsilon/2$ centered at the origin. Comparing the volumes give

$$|\mathcal{N}| \leq \frac{(1 + \varepsilon/2)^d}{(\varepsilon/2)^d} = \left(1 + \frac{2}{\varepsilon}\right)^d. \quad \square$$

We will need the following lemmata in order to prove Theorem 6.6.

Lemma 14.6. *Let ξ, ζ be real sub-gaussian random variables with mean zero, and assume ξ has unit variance. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Let $\tau \geq \tau_0 > 0$ and $K > 1$. Then there exist constants $C, c, c' > 0$ (depending only on τ_0, K, and the sub-gaussian moments of ξ and ζ) such that*

$$\sup_{\lambda \in [-K\sqrt{n}, K\sqrt{n}]} \left|\{1 \leq i \leq n : |\lambda_i(W) - \lambda| \leq \tau\sqrt{n}\}\right| \leq 4\tau n + 2(\log n)^{c' \log \log n}$$

with probability at least

$$1 - C \exp\left(-c(\log n)^{c \log \log n}\right).$$

Proof. The proof relies on Theorem 11.4. In particular, for any interval $I \subset \mathbb{R}$, Theorem 11.4 implies that

$$\mathbb{P}\left(\left|N_I - n \int_I \rho_{\mathrm{sc}}(x)dx\right| \geq (\log n)^{c' \log \log n}\right) \leq C \exp\left(-c(\log n)^{c \log \log n}\right), \quad (50)$$

where N_I denotes the number of eigenvalues of $\frac{1}{\sqrt{n}}W$ in I. Here $C, c, c' > 0$ depend only on the sub-gaussian moments of ξ and ζ.

For each $\lambda \in \mathbb{R}$, let I_λ be the interval

$$I_\lambda := \left[\frac{\lambda}{\sqrt{n}} - \tau, \frac{\lambda}{\sqrt{n}} + \tau\right].$$

Thus, using the notation from above, the problem reduces to showing

$$\sup_{\lambda \in [-K\sqrt{n}, K\sqrt{n}]} N_{I_\lambda} \leq 4\tau n + 2(\log n)^{c' \log \log n}$$

with sufficiently high probability.

Let $\mathcal{N}$ be a n^{-1}-net of the interval $[-100K\sqrt{n}, 100K\sqrt{n}]$. Then $|\mathcal{N}| \leq C'n^{3/2}$, where $C' > 0$ depends only on K. In addition, a simple net argument reveals that

$$\sup_{\lambda \in [-K\sqrt{n}, K\sqrt{n}]} N_{I_\lambda} \leq 2 \sup_{\lambda \in \mathcal{N}} N_{I_\lambda}$$

for all $n \geq n_0$, where n_0 depends only on τ_0. Thus, we have

$$\mathbb{P}\left(\sup_{\lambda \in [-K\sqrt{n}, K\sqrt{n}]} N_{I_\lambda} > 4\tau n + 2(\log n)^{c' \log \log n}\right)$$

$$\leq \mathbb{P}\left(\sup_{\lambda \in \mathcal{N}} N_{I_\lambda} > 2\tau n + (\log n)^{c' \log \log n}\right)$$

$$\leq \sum_{\lambda \in \mathcal{N}} \mathbb{P} \left(\left| N_{I_\lambda} - n \int_{I_\lambda} \rho_{\mathrm{sc}}(x) dx \right| > (\log n)^{c' \log \log n} \right)$$

by the union bound. The claim now follows by applying (50). $\quad\square$

Remark 14.7. If J is a real symmetric matrix with rank k, then, by [11, Theorem III.2.1], it follows that

$$\left| \left\{ 1 \leq i \leq n : |\lambda_i(W + J) - \lambda| \leq \tau\sqrt{n} \right\} \right| \leq \left| \left\{ 1 \leq i \leq n : |\lambda_i(W) - \lambda| \leq \tau\sqrt{n} \right\} \right| + k.$$

Lemma 14.8. *Let ξ, ζ be real sub-gaussian random variables with mean zero, and assume ξ has unit variance. Let W be an $n \times n$ Wigner matrix with atom variables ξ, ζ. Let k be a non-negative integer. Let J be a $n \times n$ deterministic real symmetric matrix with rank at most k. In addition, let H be a subspace of $\mathbb{R}^n$, which may depend only on W, that satisfies $\dim(H) \leq k$ almost surely. Let $\tau_1 \geq \tau \geq \tau_0 > 0$. For each $\lambda \in \mathbb{R}$, define the subspaces*

$$V_\lambda := \mathrm{Span} \left\{ v_i(W + J) : |\lambda_i(W + J) - \lambda| \leq \tau\sqrt{n} \right\}$$

and

$$H_\lambda := \mathrm{Span} \left\{ V_\lambda \cup H \right\}.$$

Then there exist constants $C, c, c' > 0$ (depending only on τ_0, τ_1, and the sub-gaussian moments of ξ and ζ) such that, with probability at least

$$1 - C \exp \left(-c(\log n)^{c \log \log n} \right),$$

the following holds. For every $\varepsilon > 0$, there exists a subset $\mathcal{N}$ of the unit sphere in $\mathbb{R}^n$ (depending only on W, J, and ε) such that

(i) *for every $\lambda \in \mathbb{R}$ and every unit vector $w \in H_\lambda$, there exists $w' \in \mathcal{N}$ such that $|w - w'| \leq \varepsilon$,*

(ii) $|\mathcal{N}| \leq n^2 \left(1 + \frac{2}{\varepsilon} \right)^{4\tau n + 2(\log n)^{c' \log \log n} + 2k}.$

Proof. By Lemma 11.3, $\|W\| \leq C_0\sqrt{n}$ with probability at least $1 - C_0 \exp(-c_0 n)$. On this event, there exists some constant $K > 1$ (depending only on C_0 and τ_1) such that if $|\lambda| \geq K\sqrt{n}$, then

$$\left| \left\{ 1 \leq i \leq n : |\lambda_i(W) - \lambda| \leq \tau\sqrt{n} \right\} \right| = 0.$$

Thus, by applying Lemma 14.6, we conclude that

$$\sup_{\lambda \in \mathbb{R}} \left| \left\{ 1 \leq i \leq n : |\lambda_i(W) - \lambda| \leq \tau \sqrt{n} \right\} \right| \leq 4\tau n + 2(\log n)^{c' \log \log n}$$

with probability at least

$$1 - C \exp \left(-c(\log n)^{c \log \log n} \right),$$

where $C, c, c' > 0$ depend only on τ_0, τ_1, and the sub-gaussian moments of ξ and ζ. In view of Remark 14.7, we have

$$\sup_{\lambda \in \mathbb{R}} \dim(V_\lambda) \leq 4\tau n + 2(\log n)^{c' \log \log n} + k \tag{51}$$

on the same event. For the remainder of the proof, we fix a realization in which $\dim(H) \leq k$ and the bound in (51) holds.

Observe that the collection $\{V_\lambda\}_{\lambda \in \mathbb{R}}$ contains at most n^2 distinct subspaces. This follows since each V_λ has the form

$$V_\lambda = \mathrm{Span}\left\{ v_i(W + J) : a_\lambda \leq \lambda_i(W + J) \leq b_\lambda \right\}$$

for some $a_\lambda, b_\lambda \in \mathbb{R}$. Let $V_{\lambda_1}, \ldots, V_{\lambda_N}$ be the distinct subspaces, where $N \leq n^2$.

From (51), we have

$$\sup_{1 \leq j \leq N} \dim(H_{\lambda_j}) \leq 4\tau n + 2(\log n)^{c' \log \log n} + 2k.$$

For each $1 \leq j \leq N$, let $\mathcal{N}_j$ be an ε-net of the unit sphere in H_{λ_j}. In particular, by Lemma 14.5, we can choose $\mathcal{N}_j$ such that

$$|\mathcal{N}_j| \leq \left(1 + \frac{2}{\varepsilon} \right)^{4\tau n + 2(\log n)^{c' \log \log n} + 2k}.$$

Set $\mathcal{N} := \cup_{j=1}^{N} \mathcal{N}_j$. Then, by construction,

$$|\mathcal{N}| \leq N \left(1 + \frac{2}{\varepsilon} \right)^{4\tau n + 2(\log n)^{c' \log \log n} + 2k} \leq n^2 \left(1 + \frac{2}{\varepsilon} \right)^{4\tau n + 2(\log n)^{c' \log \log n} + 2k}.$$

It remains to show that $\mathcal{N}$ satisfies property (i). Let w be a unit vector in H_λ for some $\lambda \in \mathbb{R}$. Then $w \in H_{\lambda_j}$ for some $1 \leq j \leq N$. Thus, there exists $w' \in \mathcal{N}_j \subseteq \mathcal{N}$ such that $|w - w'| \leq \varepsilon$, and the proof is complete. $\square$

We now prove Theorem 6.6.

Proof of Theorem 6.6. From the identity

$$\|v_j(W+J)\|_S^2 + \|v_j(W+J)\|_{S^c}^2 = 1,$$

it suffices to prove the lower bound for $\|v_j(W+J)\|_S$. Let $m := \lfloor \delta n \rfloor$ and $r := n - m$. By symmetry, it suffices to consider the case when $S := \{1, \ldots, m\}$.

We decompose $W + J$ as

$$M := W + J = \begin{pmatrix} A & B \\ B^{\mathrm{T}} & D \end{pmatrix} + \begin{pmatrix} J_A & 0 \\ 0 & J_D \end{pmatrix},$$

where A and J_A are $m \times m$ matrices, D and J_D are $r \times r$ matrices, and B is a $m \times r$ matrix. In particular, A is an m-dimensional Wigner matrix, D is an r-dimensional Wigner matrix, and B^{T} is an $r \times m$ matrix whose entries are iid copies of ξ. In addition, J_A and J_D are diagonal matrices both having rank at most k.

Fix $1 \leq j \leq n$ such that $\lambda_j(M) \in [\lambda_1(W), \lambda_n(W)]$, and write

$$v_j(M) = \begin{pmatrix} x_j \\ y_j \end{pmatrix},$$

where x_j is an m-vector and y_j is an r-vector. From the eigenvalue equation $Mv_j(M) = \lambda_j(M)v_j(M)$, we obtain

$$Ax_j + By_j + J_A x_j = \lambda_j(M)x_j \tag{52}$$

$$B^{\mathrm{T}}x_j + Dy_j + J_D y_j = \lambda_j(M)y_j. \tag{53}$$

In order to reach a contradiction, assume $\|x_j\| \leq \tau \varepsilon$ for some constants $0 < \tau, \varepsilon < 1$ (depending only on δ, k, and the sub-gaussian moments of ξ and ζ) to be chosen later. Define the event

$$\Omega_{n,1} := \{\|W\| \leq C_0 \sqrt{n}\}.$$

In particular, on the event $\Omega_{n,1}$, $|\lambda_j(M)| \leq C_0 \sqrt{n}$.

From (53), we find

$$(D + J_D - \lambda_j(M)I)y_j = -B^{\mathrm{T}}x_j.$$

Thus, on the event $\Omega_{n,1}$, Lemma 11.10 implies that $y_j = v_j + q_j$, where v_j, q_j are orthogonal, $\|q_j\| \leq \varepsilon$, and

$$v_j \in V_j := \mathrm{Span}\{v_i(D + J_D) : |\lambda_i(D + J_D) - \lambda_j(M)| \leq C_0 \tau \sqrt{n}\}.$$

Let $0 < c_0 < 1$ be a constant (depending only on δ, k, and the sub-gaussian moments of ξ and ζ) to be chosen later. By Lemma 14.8, there exists a subset $\mathcal{N}$ of the unit sphere in $\mathbb{R}^r$ such that

(i) for every unit vector $v \in V_j$ there exists $v' \in \mathcal{N}$ such that $\|v - v'\| \le c_0$,

(ii) we have

$$|\mathcal{N}| \le n^2 \left(1 + \frac{2}{c_0}\right)^{12\tau(\sqrt{n}/\sqrt{r})n + 2(\log n)^{c' \log\log n} + 2k},$$

(iii) $\mathcal{N}$ depends only on D, J_D, and c_0.

We now condition on the sub-matrix D such that properties (i) and (ii) hold. By independence, this conditioning does not effect the matrices A and B. Since $\mathcal{N}$ only depends on D, J_D, and c_0, we now treat $\mathcal{N}$ as a deterministic set.

Let $H := \mathrm{range}(J_A)^\perp$, and let P_H denote the orthogonal projection onto H. Then, from (52), we have

$$
\begin{aligned}
2C_0\sqrt{n}\|x_j\| \ge \|(A - \lambda_j(M))x_j\| = \|By_j + J_A x_j\| &\ge \|P_H B y_j\| \\
&\ge \|P_H B v_j\| - C_0\varepsilon\sqrt{n} \\
&\ge \|v_j\| \inf_{\substack{v \in V_j \\ \|v\|=1}} \|P_H B v\| - C_0\varepsilon\sqrt{n}
\end{aligned}
$$

on the event $\Omega_{n,1}$. As $\|q_j\|^2 + \|v_j\|^2 = \|y_j\|^2 = 1 - \|x_j\|^2$, we have

$$\|v_j\|^2 \ge 1 - \varepsilon^2\tau^2 - \varepsilon^2,$$

and hence, we conclude that

$$2C_0\sqrt{n}\|x_j\| \ge \sqrt{1 - \varepsilon^2\tau^2 - \varepsilon^2} \inf_{\substack{v \in V_j \\ \|v\|=1}} \|P_H B v\| - C_0\varepsilon\sqrt{n}. \tag{54}$$

We now obtain a lower bound for

$$\inf_{\substack{v \in V_j \\ \|v\|=1}} \|P_H B v\|.$$

Indeed, since J_A has rank at most k, $\dim(H) \ge n - k$. Thus, by taking c_0 sufficiently small, we find that

$$
\begin{aligned}
\mathbb{P}\left(\inf_{\substack{v \in V_j \\ \|v\|=1}} \|P_H B v\| \le \frac{1}{100}\sqrt{m-k} \right) & \\
\le \mathbb{P}\left(\inf_{v \in \mathcal{N}} \|P_H B v\| \le \frac{1}{100}\sqrt{m-k} + c_0\|B\| \right) & \\
\le \mathbb{P}\left(\left\{ \inf_{v \in \mathcal{N}} \|P_H B v\| \le \frac{1}{100}\sqrt{m-k} + c_0\|B\| \right\} \cap \Omega_{n,1} \right) + \mathbb{P}\left(\Omega_{n,1}^c\right)
\end{aligned}
$$

$$\leq \mathbb{P}\left(\inf_{v \in \mathcal{N}} \|P_H B v\| \leq \frac{1}{100}\sqrt{m-k} + C_0 c_0 \sqrt{n}\right) + \mathbb{P}\left(\Omega_{n,1}^c\right)$$

$$\leq \mathbb{P}\left(\inf_{v \in \mathcal{N}} \|P_H B v\| \leq \frac{1}{2}\sqrt{m-k}\right) + \mathbb{P}\left(\Omega_{n,1}^c\right)$$

$$\leq \sum_{v \in \mathcal{N}} \mathbb{P}\left(\|P_H B v\| \leq \frac{1}{2}\sqrt{m-k}\right) + \mathbb{P}\left(\Omega_{n,1}^c\right).$$

Therefore, in view of Lemma 11.8, we obtain

$$\mathbb{P}\left(\inf_{\substack{v \in V_j \\ \|v\|=1}} \|P_H B v\| \leq \frac{1}{100}\sqrt{m-k}\right)$$

$$\leq 2n^2 \left(1 + \frac{2}{c_0}\right)^{12\tau(\sqrt{n}/\sqrt{r})n + 2(\log n)^{c' \log\log n} + 2k} \exp(-c(m-k)) + \mathbb{P}\left(\Omega_{n,1}^c\right),$$

where $c > 0$ depends only on the sub-gaussian moment of ξ. Therefore, by taking τ sufficiently small and applying Lemma 11.3, we have

$$\inf_{\substack{v \in V_j \\ \|v\|=1}} \|P_H B v\| \geq \frac{1}{100}\sqrt{m-k}$$

with probability at least

$$1 - C_2 \exp\left(-c_2 (\log n)^{c_2 \log\log n}\right),$$

where $C_2, c_2 > 0$ depend only on δ, k, and the sub-gaussian moments of ξ and ζ.

On this event, (54) implies that

$$\varepsilon \tau \geq \|x_j\| \geq \frac{\sqrt{1 - \varepsilon^2 \tau^2 - \varepsilon^2}}{200 C_0} \sqrt{\frac{m-k}{n}} - \frac{\varepsilon}{2},$$

a contradiction for ε sufficiently small. Therefore, on the same event, we conclude that $\|x_j\| \geq \varepsilon \tau$. $\square$

14.3. Proof of Theorem 6.8

Unsurprisingly, the proof of Theorem 6.8 is very similar to the proof of Theorem 6.6.

Proof of Theorem 6.8. Let $m := \lfloor \delta n \rfloor$ and $r := n - m$. By symmetry, it suffices to consider the case when $S := \{1, \ldots, m\}$.

We decompose $W + J$ as

$$M := W + J = \begin{pmatrix} A & B \\ B^{\mathrm{T}} & D \end{pmatrix} + \begin{pmatrix} J_A & J_B \\ J_B^{\mathrm{T}} & J_D \end{pmatrix},$$

where A and J_A are $m \times m$ matrices, B and J_B are $m \times r$ matrices, and D and J_D are $r \times r$ matrices. In particular, A is an m-dimensional Wigner matrix, D is an r-dimensional Wigner matrix, and B^{T} is an $r \times m$ matrix whose entries are iid copies of ξ.

Since the rank of any sub-matrix is not more than the rank of the original matrix (see, for example, [43, Section 0.4.5]), it follows that J_A, J_B, and J_D all have rank at most k.

Fix $\varepsilon_1 n \le j \le (1 - \varepsilon_1) n$. Then, by [11, Theorem III.2.1], for n sufficiently large (in terms of k and ε_1),

$$\lambda_{j+k}(W) \le \lambda_j(M) \le \lambda_{j-k}(W). \tag{55}$$

Express $v_j(M)$ as

$$v_j(M) = \begin{pmatrix} x_j \\ y_j \end{pmatrix},$$

where x_j is an m-vector and y_j is an r-vector. From the eigenvalue equation $M v_j(M) = \lambda_j(M) v_j(M)$, we obtain

$$A x_j + B y_j + J_A x_j + J_B y_j = \lambda_j(M) x_j \tag{56}$$
$$B^{\mathrm{T}} x_j + D y_j + J_B^{\mathrm{T}} x_j + J_D y_j = \lambda_j(M) y_j. \tag{57}$$

In order to reach a contradiction, assume $\frac{1}{n^{1-\varepsilon_0}} \le \|x_j\| \le \tau \varepsilon$ for some constants $0 < \tau, \varepsilon < 1$ (depending only on δ, k, ε_0, ε_1, and the sub-gaussian moments of ξ and ζ) to be chosen later. Define the event

$$\Omega_{n,1} := \{\|W\| \le C_0 \sqrt{n}\}.$$

In particular, in view of (55), on the event $\Omega_{n,1}$, $|\lambda_j(M)| \le C_0 \sqrt{n}$.

Let γ_j denote the classical location of the jth eigenvalue of a Wigner matrix. That is, γ_j is defined by

$$n \int_{-\infty}^{\gamma_j} \rho_{\mathrm{sc}}(x) dx = j,$$

where ρ_{sc} is defined in (23). Let $\alpha > 0$ be a small parameter (depending on ε_0) to be chosen later. From (55) and [48, Theorem 3.6], we conclude that the event

$$\Omega_{n,2} := \left\{ |\lambda_j(M) - \sqrt{n}\gamma_j| \le \frac{n^\alpha}{n^{1/2}} \right\}$$

holds with probability at least $1 - C \exp\left(-c(\log n)^{c \log \log n}\right)$, where $C, c > 0$ depend on α, k, and the sub-gaussian moments of ξ and ζ.

From (57), we obtain

$$(D + J_D - \lambda_j(M)I)y_j = -(B^{\mathrm{T}} + J_B^{\mathrm{T}})x_j,$$

which we rewrite as

$$(D + J_D - \sqrt{n}\gamma_j I)y_j = -(B^{\mathrm{T}} + (\sqrt{n}\gamma_j - \lambda_j(M))Q_j + J_B^{\mathrm{T}})x_j,$$

where Q_J is the rank one matrix given by

$$Q_j := \frac{y_j x_j^{\mathrm{T}}}{\|x_j\|^2}.$$

In particular, $Q_j x_j = y_j$. Since $\|x_j\| \geq \frac{1}{n^{1-\varepsilon_0}}$, we have

$$\|(\sqrt{n}\gamma_j - \lambda_j(M))Q_j\| \leq \frac{n^\alpha}{n^{1/2}}\frac{1}{\|x_j\|} \leq n^{1/2-\varepsilon_0/2}$$

on the event $\Omega_{n,2}$ by taking α sufficient small. Thus, on the event $\Omega_{n,1} \cap \Omega_{n,2}$, we have

$$\|B^{\mathrm{T}} + (\sqrt{n}\gamma_j - \lambda_j(M))Q_j\| \leq C_0\sqrt{n} + n^{1/2-\varepsilon_0/2}.$$

Thus, on the same event, we apply Lemma 11.11 and obtain the following.

(i) There exists $\eta > 0$ such that $D + J_D - (\sqrt{n}\gamma_j + \eta)I$ is invertible.
(ii) $y_j = v_j + q_j$, where $\|q_j\| \leq \varepsilon$, $v_j \in \mathrm{Span}\{V_j \cup H_j\}$,

$$V_j := \mathrm{Span}\left\{v_i(D + J_D) : |\lambda_i(D + J_D) - \sqrt{n}\gamma_j| \leq C_0\tau\sqrt{n} + \tau n^{1/2-\varepsilon_0/2}\right\}$$

and

$$H_j := \mathrm{range}\left((D + J_D - (\sqrt{n}\gamma_j + \eta)I)^{-1}J_B^{\mathrm{T}}\right).$$

In particular, V_j and H_j depend only on D, and H_j has dimension at most k.

We pause a moment to note the following. We introduced the classical location γ_j so that the subspace H_j depends only on D and not on the entire matrix W. If we had not introduced γ_j, then H_j would depend on $\lambda_j(M)$ (and hence on W). Since H_j only depends on D, Lemma 14.8 is applicable, and the net $\mathcal{N}$, introduced below, also depends only on D.

Let $c_0 > 0$ be a constant (depending on δ, k, and the sub-gaussian moments of ξ and ζ) to be chosen later. Then, by Lemma 14.8, there exists a subset $\mathcal{N}$ of the unit ball in $\mathbb{R}^r$ such that

(i) for every unit vector $v \in \mathrm{Span}\{V_j \cup H_j\}$, there exists $v' \in \mathcal{N}$ such that $\|v - v'\| \leq c_0$,

(ii) we have

$$|\mathcal{N}| \leq n^2 \left(1 + \frac{2}{c_0}\right)^{12\tau\frac{\sqrt{n}}{\sqrt{r}}n + 4\tau\frac{\sqrt{n}}{\sqrt{r}n^{\varepsilon_0/2}}n + 2(\log n)^{c' \log\log n} + 2k},$$

(iii) $\mathcal{N}$ depends only on D, J_D, J_B^{T}, and c_0.

We now condition on the sub-matrix D such that the first two properties hold. By independence, this conditioning does not effect the matrices A and B. We now treat $\mathcal{N}$ as a deterministic set.

Returning to (56), on the event $\Omega_{n,1}$, we have

$$
\begin{aligned}
2C_0\sqrt{n}\|x_j\| \geq \|(A - \lambda_j(M)I)x_j\| &= \|By_j + J_B y_j + J_A x_j\| \\
&= \|By_j + J'v_j(M)\| \\
&\geq \|P_H B y_j\| \\
&\geq \|P_H B v_j\| - C_0\varepsilon\sqrt{n},
\end{aligned}
$$

where $J' = (J_A \;\; J_B)$, $H = (\operatorname{range} J')^{\perp}$, and P_H is the orthogonal projection onto H. (Note that H and H_j are different subspaces.) Thus, we conclude that

$$2C_0\sqrt{n}\|x_j\| \geq \left(\sqrt{1 - \varepsilon^2\tau^2} - \varepsilon\right) \inf_{\substack{v \in \operatorname{Span}\{V_j \cup H_j\} \\ \|v\|=1}} \|P_H B v\| - C_0\varepsilon\sqrt{n}. \tag{58}$$

We now obtain a lower bound for

$$\inf_{\substack{v \in \operatorname{Span}\{V_j \cup H_j\} \\ \|v\|=1}} \|P_H B v\|$$

as in the proof of Theorem 6.6. Indeed, for c_0 sufficiently small, we find

$$
\begin{aligned}
&\mathbb{P}\left(\inf_{\substack{v \in \operatorname{Span}\{V_j \cup H_j\} \\ \|v\|=1}} \|P_H B v\| \leq \frac{1}{100}\sqrt{m-k}\right) \\
&\leq 2n^2 \left(1 + \frac{2}{c_0}\right)^{12\tau\frac{\sqrt{n}}{\sqrt{r}}n + 4\tau\frac{\sqrt{n}}{\sqrt{r}n^{\varepsilon_0/2}}n + 2(\log n)^{c' \log\log n} + 2k} \exp(-c(m-k)) + \mathbb{P}\left(\Omega_{n,1}^c\right).
\end{aligned}
$$

Therefore, from (58) and by taking τ sufficiently small, we find that

$$\varepsilon\tau \geq \|x_j\| \geq \frac{\sqrt{1 - \varepsilon^2\tau^2} - \varepsilon}{200 C_0}\sqrt{\frac{m-k}{n}} - \frac{\varepsilon}{2}$$

with probability at least $1 - C_2 \exp\left(-c_2 (\log n)^{c_2 \log \log n}\right)$. This is a contradiction for ε sufficiently small. Thus, we conclude that, on the same event, either $\|x_j\| \geq \varepsilon\tau$ or $\|x_j\| \leq \frac{1}{n^{1-\varepsilon_0}}$. $\square$

14.4. Proof of Theorem 6.10

In order to prove Theorem 6.10, we will apply Theorem 3.3.

Without loss of generality, we assume $\theta > 0$. If $\theta < 0$, one can consider $-W - J$ instead of $W + J$. Indeed, $-W$ is a Wigner matrix with atom variable $-\xi, -\zeta$. In addition, $-W - J$ and $W + J$ have the same eigenvectors while the eigenvalues only differ by sign.

Recall that $M := W + J$. Beginning with the eigenvalue equation

$$Mv_j(M) = \lambda_j(M)v_j(M),$$

we multiply on the left by $v_j(W)^{\mathrm{T}}$ to obtain

$$\lambda_j(W)v_j(W)^{\mathrm{T}}v_j(M) + v_j(W)^{\mathrm{T}}Jv_j(M) = \lambda_j(M)v_j(W)^{\mathrm{T}}v_j(M).$$

Thus, we find that

$$|v_j(W)^{\mathrm{T}}Jv_j(M)| \leq |\lambda_j(M) - \lambda_j(W)|.$$

As J has rank 1, eigenvalue interlacing (see, for instance, [11, Exercise III.2.4]) implies that

$$|v_j(W)^{\mathrm{T}}Jv_j(M)| \leq \lambda_{j+1}(W) - \lambda_j(W).$$

Since $J = \theta uu^{\mathrm{T}}$, we conclude that

$$|v_j(W) \cdot u||v_j(M) \cdot u| \leq \frac{\lambda_{j+1}(W) - \lambda_j(W)}{\theta}.$$

We will bound $\lambda_{j+1}(W) - \lambda_j(W)$ above using the rigidity of eigenvalues estimate from [35]. Indeed, by [35, Theorem 2.2], it follows that

$$\lambda_{j+1}(W) - \lambda_j(W) \leq \frac{C(\log n)^{c \log \log n}}{\sqrt{n}}$$

with probability $1 - o(1)$ for some constants $C, c > 0$. Thus, we obtain that

$$\sqrt{n}|v_j(W) \cdot u||v_j(M) \cdot u| \leq \frac{C(\log n)^{c \log \log n}}{\theta}$$

with probability $1 - o(1)$.

From Theorem 3.3, we have

$$\sqrt{n}|v_j(W) \cdot u| \geq \frac{1}{\log n}$$

with probability $1 - o(1)$. Thus, we conclude that

$$|v_j(M) \cdot u| \leq \frac{C(\log n)^{(c+1)\log\log n}}{\theta}$$

with probability $1 - o(1)$, and the proof is complete.

Acknowledgments

The first author thanks Professor Roman Vershynin for clarifications. The third author would like to thank Professor Tiefeng Jiang for many useful discussions and constant encouragement. The authors also thank the anonymous referee for valuable comments.

Appendix A. Proof of Lemma 10.3

For each $n \geq 1$, let V_n and W_n be independent χ^2-distribution random variables with α_n and β_n degrees of freedom respectively. Then

$$\frac{V_n}{V_n + W_n} \sim \text{Beta}\left(\frac{\alpha_n}{2}, \frac{\beta_n}{2}\right).$$

This can be verified by computing the distribution of the ratio directly; see [72] for details.

Thus, in order to prove Lemma 10.3, it suffices to show that

$$\sqrt{\frac{(\alpha_n + \beta_n)^3}{2\alpha_n\beta_n}}\left(\frac{V_n}{V_n + W_n} - \frac{\alpha_n}{\alpha_n + \beta_n}\right) \longrightarrow N(0,1) \tag{59}$$

in distribution as $n \to \infty$.

We decompose the left-hand side of (59) as

$$\sqrt{\frac{(\alpha_n + \beta_n)^3}{2\alpha_n\beta_n}}\left(\frac{V_n}{V_n + W_n} - \frac{\alpha_n}{\alpha_n + \beta_n}\right)$$

$$= \sqrt{\frac{\alpha_n + \beta_n}{2\alpha_n\beta_n}}\left(\frac{\beta_n V_n - \alpha_n W_n}{\alpha_n + \beta_n}\right)\frac{\alpha_n + \beta_n}{V_n + W_n}$$

$$= \sqrt{\frac{\alpha_n + \beta_n}{2\alpha_n\beta_n}}\left(\frac{\beta_n\sqrt{\alpha_n}\left(\frac{V_n - \alpha_n}{\sqrt{\alpha_n}}\right) - \alpha_n\sqrt{\beta_n}\left(\frac{W_n - \beta_n}{\sqrt{\beta_n}}\right)}{\alpha_n + \beta_n}\right)\frac{\alpha_n + \beta_n}{V_n + W_n}$$

$$= \left[\sqrt{\frac{\beta_n}{\alpha_n + \beta_n}} \left(\frac{V_n - \alpha_n}{\sqrt{2\alpha_n}} \right) - \sqrt{\frac{\alpha_n}{\alpha_n + \beta_n}} \left(\frac{W_n - \beta_n}{\sqrt{2\beta_n}} \right) \right] \frac{\alpha_n + \beta_n}{V_n + W_n}.$$

By definition of the χ^2-distribution, V_n has the same distribution as the sum of α_n independent squared standard normal random variables. Similarly, W_n has the same distribution as the sum of β_n independent squared standard normal random variables. Thus, by the central limit theorem,

$$\frac{V_n - \alpha_n}{\sqrt{2\alpha_n}} \longrightarrow N(0,1) \quad \text{and} \quad \frac{W_n - \beta_n}{\sqrt{2\beta_n}} \longrightarrow N(0,1)$$

in distribution as $n \to \infty$. Moreover, since V_n and W_n are independent, we conclude that

$$\sqrt{\frac{\beta_n}{\alpha_n + \beta_n}} \left(\frac{V_n - \alpha_n}{\sqrt{2\alpha_n}} \right) - \sqrt{\frac{\alpha_n}{\alpha_n + \beta_n}} \left(\frac{W_n - \beta_n}{\sqrt{2\beta_n}} \right) \longrightarrow N(0,1)$$

in distribution as $n \to \infty$. Here we used Slutsky's theorem (see Theorem 11.4 in [42, Chapter 5]) since both

$$\sqrt{\frac{\beta_n}{\alpha_n + \beta_n}} \quad \text{and} \quad \sqrt{\frac{\alpha_n}{\alpha_n + \beta_n}}$$

converge to limits in $[0,1]$ by supposition.

Finally, by the law of large numbers, we observe that

$$\frac{\alpha_n + \beta_n}{V_n + W_n} \longrightarrow 1$$

almost surely as $n \to \infty$. Thus, by another application of Slutsky's theorem, we conclude that

$$\left[\sqrt{\frac{\beta_n}{\alpha_n + \beta_n}} \left(\frac{V_n - \alpha_n}{\sqrt{2\alpha_n}} \right) - \sqrt{\frac{\alpha_n}{\alpha_n + \beta_n}} \left(\frac{W_n - \beta_n}{\sqrt{2\beta_n}} \right) \right] \frac{\alpha_n + \beta_n}{V_n + W_n} \longrightarrow N(0,1)$$

in distribution as $n \to \infty$, and the proof is complete.

Appendix B. Proof of Lemmas 11.10 and 11.11

We now present the proof of Lemmas 11.10 and 11.11.

Proof of Lemma 11.10. By the spectral theorem, the eigenvectors $v_1(B), \ldots, v_r(B)$ of B form an orthonormal basis. Write $y = \sum_{i=1}^r \alpha_i v_i(B)$. Define

$$q := \sum_{i:|\lambda_i(B)|>\tau\|A\|} \alpha_i v_i(B),$$

and set $v := y - q$. Clearly, v and q are orthogonal. Moreover,

$$v \in \mathrm{Span}\{v_i(B) : |\lambda_i(B)| \leq \tau\|A\|\}$$

by construction. It remains to show $\|q\| \leq \varepsilon$.

We now utilize the equation $By = Ax$. We first consider the case when $\|A\| = 0$. In this case, $By = 0$; in other words, y is in the null space of B. Thus, the vector q is zero, and the claim follows.

Assume $\|A\| > 0$. Since $\|Ax\| \leq \|A\|\|x\| \leq \varepsilon\tau\|A\|$, we have $\|By\| \leq \varepsilon\tau\|A\|$. Thus, by the spectral theorem, we obtain

$$\|By\|^2 = \sum_{i=1}^{r} \lambda_i^2(B)|\alpha_i|^2 \leq \varepsilon^2\tau^2\|A\|^2.$$

By definition of q, this implies

$$\tau^2\|A\|^2\|q\|^2 = \tau^2\|A\|^2 \sum_{i:|\lambda_i(B)|>\tau\|A\|} |\alpha_i|^2 \leq \sum_{i:|\lambda_i(B)|>\tau\|A\|} \lambda_i^2(B)|\alpha_i|^2 \leq \varepsilon^2\tau^2\|A\|^2.$$

Therefore, we conclude that

$$\|q\|^2 \leq \frac{\varepsilon^2\tau^2\|A\|^2}{\tau^2\|A\|^2} = \varepsilon^2,$$

and the proof is complete. □

Proof of Lemma 11.11. Let $\{\eta_k\}_{k=1}^{\infty}$ be a non-increasing sequence of non-negative real numbers such that $\lim_{k\to\infty} \eta_k = 0$ and $B - \eta_k I$ is invertible for all $k \geq 1$. It is always possible to find such a sequence since B has at most r distinct eigenvalues.

Without loss of generality, we assume $0 < \varepsilon < 1$. Indeed, if $\varepsilon \geq 1$, part (ii) of the lemma follows by taking $q := y$ and $v := 0$.

Define the value

$$L := \frac{\kappa\varepsilon\tau + \eta_k}{\varepsilon} \tag{60}$$

and the subspace

$$W := \mathrm{Span}\{v_i(B) : |\lambda_i(B)| \leq \tau\kappa\}.$$

We make two simple observations related to the subspace W.

- If $|\lambda_i(B)| \leq \tau\kappa$, then, by the triangle inequality,

$$|\lambda_i(B) - \eta_k| \leq \tau\kappa + \eta_k \leq L.$$

Here the last inequality follows from the fact that $0 < \varepsilon < 1$.

- If $|\lambda_i(B) - \eta_k| \leq L$, then

$$|\lambda_i(B)| \leq L + \eta_k = \kappa\tau + \eta_k\left(1 + \frac{1}{\varepsilon}\right).$$

From the two observations above, we conclude that, for all k sufficiently large (i.e. η_k sufficiently small),

$$\{1 \leq i \leq r : |\lambda_i(B)| \leq \tau\kappa\} = \{1 \leq i \leq r : |\lambda_i(B) - \eta_k| \leq L\}.$$

Fix k sufficiently large so that the above equality holds. Then

$$W = \mathrm{Span}\{v_i(B) : |\lambda_i(B) - \eta_k| \leq L\}. \tag{61}$$

Let

$$V := \mathrm{Span}\left\{W \cup \mathrm{range}\left((B - \eta_k I)^{-1}J\right)\right\}.$$

The bound in (24) follows from the fact that the rank of $(B - \eta_k I)^{-1}J$ is no larger than the rank of J (see [43, Section 0.4]).

Since B is Hermitian, we have

$$\lambda_i(B - \eta_k I) = \lambda_i(B) - \eta_k, \quad v_i(B - \eta_k I) = v_i(B)$$

for all $1 \leq i \leq r$. In particular, by the spectral theorem, we have

$$(B - \eta_k I)^{-1} = \sum_{i=1}^{r} \frac{1}{\lambda_i(B) - \eta_k} v_i(B)v_i(B)^*. \tag{62}$$

We rewrite $By = (A + J)x$ as

$$(B - \eta_k I)y = Ax - \eta_k y + Jx,$$

and hence obtain

$$y = (B - \eta_k I)^{-1}(Ax - \eta_k y) + (B - \eta_k I)^{-1}Jx.$$

Decompose

$$Ax - \eta_k y = \sum_{i=1}^{r} \alpha_i v_i(B).$$

Then, by (62), we have

$$(B - \eta_k I)^{-1}(Ax - \eta_k y) = \sum_{i=1}^{r} \frac{1}{\lambda_i(B) - \eta_k} \alpha_i v_i(B) = q + w,$$

where

$$q := \sum_{i:|\lambda_i(B)-\eta_k|>L} \frac{1}{\lambda_i(B) - \eta_k} \alpha_i v_i(B)$$

and

$$w := \sum_{i:|\lambda_i(B)-\eta_k|\leq L} \frac{1}{\lambda_i(B) - \eta_k} \alpha_i v_i(B).$$

Set $v := w + (B - \eta_k I)^{-1} Jx$. From (61), we find that $w \in W$, and hence $v \in V$. It remains to show $\|q\| \leq \varepsilon$.

By definition of q, we obtain

$$\|q\|^2 = \sum_{i:|\lambda_i(B)-\eta_k|>L} \frac{1}{|\lambda_i(B) - \eta_k|^2} |\alpha_i|^2 \leq \frac{1}{L^2} \|Ax - \eta_k y\|^2.$$

By supposition, we have $\|Ax - \eta_k y\| \leq \kappa\varepsilon\tau + \eta_k$. Thus, in view of (60), we conclude that

$$\|q\|^2 \leq \frac{(\kappa\varepsilon\tau + \eta_k)^2}{L^2} = \varepsilon^2,$$

and the proof of the lemma is complete. $\square$

References

[1] N. Anantharaman, E. Le Masson, Quantum ergodicity on large regular graphs, Duke Math. J. 164 (4) (2015) 723–765.
[2] G.W. Anderson, A. Guionnet, O. Zeitouni, An Introduction to Random Matrices, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010.
[3] S. Arora, A. Bhaskara, Eigenvectors of random graphs: delocalization and nodal domains, preprint, available at http://www.cs.princeton.edu/~bhaskara/files/deloc.pdf.
[4] A. Auffinger, G. Ben Arous, S. Péché, Poisson convergence for the largest eigenvalues of heavy tailed random matrices, Ann. Inst. Henri Poincaré Probab. Stat. 45 (3) (2009) 589–610.
[5] Z. Bai, J.W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, second edition, Springer Series in Statistics, Springer, New York, 2010.
[6] J. Baik, G. Ben Arous, S. Péché, Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices, Ann. Probab. 33 (5) (2005) 1643–1697.
[7] R. Bauerschmidt, A. Knowles, H.-T. Yau, Local semicircle law for random regular graphs, preprint, arXiv:1503.08702, 2015.
[8] F. Benaych-Georges, A. Guionnet, Central limit theorem for eigenvectors of heavy tailed matrices, Electron. J. Probab. 19 (54) (2014) 1–27.
[9] F. Benaych-Georges, R. Rao Nadakuditi, Eigenvalues and eigenvectors of finite, low rank perturbation of large random matrices, Adv. Math. 227 (1) (2009) 494–521.
[10] F. Benaych-Georges, S. Péché, Localization and delocalization for heavy tailed band matrices, Ann. Inst. Henri Poincaré Probab. Stat. 50 (4) (2014) 1385–1403.

[11] R. Bhatia, Matrix Analysis, Graduate Texts in Mathematics, vol. 169, Springer-Verlag, New York, 1997.
[12] A. Bloemendal, L. Erdos, A. Knowles, H.-T. Yau, J. Yin, Isotropic local laws for sample covariance and generalized Wigner matrices, Electron. J. Probab. 19 (33) (2014) 1–53.
[13] C. Bordenave, A. Guionnet, Localization and delocalization of eigenvectors for heavy-tailed random matrices, Probab. Theory Related Fields 157 (3) (2013) 885–953.
[14] C. Bordenave, A. Guionnet, Delocalization at small energy for heavy-tailed random matrices, preprint, arXiv:1603.08845, 2016.
[15] P. Bourgade, H.-T. Yau, The eigenvector moment flow and local quantum unique ergodicity, arXiv:1312.1301, 2013.
[16] S. Brooks, E. Lindenstrauss, Non-localization of eigenfunctions on large regular graphs, Israel J. Math. 193 (1) (2013) 1–14.
[17] C. Cacciapuoti, A. Maltsev, B. Schlein, Local Marchenko–Pastur law at the hard edge of sample covariance matrices, preprint, arXiv:1206.1730, 2012.
[18] P. Cizeau, J.P. Bouchaud, Theory of Lévy matrices, Phys. Rev. E 50 (1994).
[19] S. Csörgő, E. Haeusler, D.M. Mason, The asymptotic distribution of extreme sums, Ann. Probab. 19 (2) (1991) 783–811.
[20] L. de Haan, A. Ferreira, Extreme Value Theory: An Introduction, Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.
[21] Y. Dekel, J.R. Lee, N. Linial, Eigenvectors of random graphs: nodal domains, Random Structures Algorithms 39 (1) (2011) 39–58.
[22] I. Dumitriu, S. Pal, Sparse regular random graphs: spectral density and eigenvectors, Ann. Probab. 40 (5) (2012) 2197–2235.
[23] L. Erdős, Universality of Wigner random matrices: a survey of recent results, Uspekhi Mat. Nauk 66 (3(399)) (2011) 67–198.
[24] L. Erdös, A. Knowles, Quantum diffusion and eigenfunction delocalization in a random band matrix model, Comm. Math. Phys. 303 (2) (2011) 509–554.
[25] L. Erdös, A. Knowles, Quantum diffusion and delocalization for band matrices with general distribution, Ann. Henri Poincaré 12 (7) (2011) 1227–1319.
[26] L. Erdős, H.-T. Yau, Universality of local spectral statistics of random matrices, Bull. Amer. Math. Soc. (N.S.) 49 (3) (2012) 377–414.
[27] L. Erdős, B. Schlein, H.-T. Yau, Local semicircle law and complete delocalization for Wigner random matrices, Comm. Math. Phys. 287 (2) (2009) 641–655.
[28] L. Erdős, B. Schlein, H.-T. Yau, Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices, Ann. Probab. 37 (3) (2009) 815–852.
[29] L. Erdős, J. Ramírez, B. Schlein, T. Tao, V. Vu, H.-T. Yau, Bulk universality for Wigner Hermitian matrices with subexponential decay, Math. Res. Lett. 17 (4) (2010) 667–674.
[30] L. Erdős, B. Schlein, H.-T. Yau, Wegner estimate and level repulsion for Wigner random matrices, Int. Math. Res. Not. IMRN (3) (2010) 436–479.
[31] L. Erdős, B. Schlein, H.-T. Yau, Universality of random matrices and local relaxation flow, Invent. Math. 185 (1) (2011) 75–119.
[32] L. Erdős, H.-T. Yau, J. Yin, Universality for generalized Wigner matrices with Bernoulli distribution, J. Comb. 2 (1) (2011) 15–81.
[33] L. Erdős, B. Schlein, H.-T. Yau, J. Yin, The local relaxation flow approach to universality of the local statistics for random matrices, Ann. Inst. Henri Poincaré Probab. Stat. 48 (1) (2012) 1–46.
[34] L. Erdős, H.-T. Yau, J. Yin, Bulk universality for generalized Wigner matrices, Probab. Theory Related Fields 154 (1–2) (2012) 341–407.
[35] L. Erdős, H.-T. Yau, J. Yin, Rigidity of eigenvalues of generalized Wigner matrices, Adv. Math. 229 (3) (2012) 1435–1515.
[36] L. Erdős, A. Knowles, H.-T. Yau, Averaging fluctuations in resolvents of random band matrices, Ann. Henri Poincaré 14 (8) (2013) 1837–1926.
[37] L. Erdős, A. Knowles, H.-T. Yau, J. Yin, Spectral statistics of Erdős–Rényi graphs I: local semicircle law, Ann. Probab. 41 (3B) (2013) 2279–2375.
[38] L. Erdős, A. Knowles, H.-T. Yau, J. Yin, Delocalization and diffusion profile for random band matrices, Comm. Math. Phys. 323 (1) (2013) 367–416.
[39] W. Feller, An Introduction to Probability Theory and Its Applications. Vol. II, second edition, John Wiley & Sons, Inc., New York–London–Sydney, 1971.
[40] Z. Füredi, J. Komlós, The eigenvalues of random symmetric matrices, Combinatorica 1 (3) (1981) 233–241.

[41] L. Gallardo, Au sujet du contenu probabiliste d'un lemme d'Henri Poincaré, in: Saint-Flour Probability Summer Schools, Saint-Flour, 1979/1980, Ann. Sci. Univ. Clermont-Ferrand II Math. 19 (1981) 185–190.

[42] A. Gut, Probability: A Graduate Course, Springer Texts in Statistics, Springer, New York, 2005.

[43] R.A. Horn, C.R. Johnson, Matrix Analysis, second edition, Cambridge University Press, Cambridge, 2013.

[44] T. Jiang, How many entries of a typical orthogonal matrix can be approximated by independent normals?, Ann. Probab. 34 (4) (2006) 1497–1529.

[45] A. Knowles, J. Yin, Eigenvector distribution of Wigner matrices, Probab. Theory Related Fields 155 (3–4) (2013) 543–582.

[46] A. Knowles, J. Yin, The outliers of a deformed Wigner matrix, Ann. Probab. 42 (5) (2014) 1980–2031.

[47] B. Laurent, P. Massart, Adaptive estimation of a quadratic functional by model selection, Ann. Statist. 28 (5) (2000) 1302–1338.

[48] J.O. Lee, J. Yin, A necessary and sufficient condition for edge universality of Wigner matrices, Duke Math. J. 163 (1) (2014) 117–173.

[49] J. Matoušek, Lectures on Discrete Geometry, Graduate Texts in Mathematics, vol. 212, Springer-Verlag, New York, 2002.

[50] F. McSherry, Spectral partitioning of random graphs, in: 42nd IEEE Symposium on Foundations of Computer Science, Las Vegas, NV, 2001, IEEE Computer Soc., Los Alamitos, CA, 2001, pp. 529–537.

[51] P. Mitra, Entrywise bounds for eigenvectors of random graphs, Electron. J. Combin. 16 (1) (2009), Research Paper 131, 18 pp.

[52] M.E. Newman, Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E 74 (3) (2006) 036104.

[53] M.E. Newman, Modularity and community structure in networks, Proc. Nat. Acad. Sci. 103 (23) (2006) 8577–8582.

[54] H. Nguyen, T. Tao, V. Vu, Random matrices: tail bounds for gaps between eigenvalues, preprint, arXiv:1504.00396, 2015.

[55] S. O'Rourke, B. Touri, Controllability of random systems: universality and minimal controllability, preprint, arXiv:1506.03125, 2015.

[56] S. O'Rourke, V. Vu, K. Wang, Random perturbation of low rank matrices: improving classical bounds, preprint, arXiv:1311.2657, 2016.

[57] L. Page, S. Brin, R. Motwani, T. Winograd, The pagerank citation ranking: bringing order to the web, Technical report, Stanford InfoLab, 1999.

[58] N.S. Pillai, J. Yin, Universality of covariance matrices, Ann. Appl. Probab. 24 (3) (2014) 935–1001.

[59] A. Pizzo, D. Renfrew, A. Soshnikov, On finite rank deformations of Wigner matrices, Ann. Inst. Henri Poincaré Probab. Stat. 49 (1) (2013) 64–94.

[60] A. Pothen, H.D. Simon, K.-P. Liou, Partitioning sparse matrices with eigenvectors of graphs, in: Sparse Matrices, Gleneden Beach, OR, 1989, SIAM J. Matrix Anal. Appl. 11 (3) (1990) 430–452.

[61] D. Renfrew, A. Soshnikov, On finite rank deformations of Wigner matrices II: delocalized perturbations, Random Matrices Theory Appl. 2 (1) (2013).

[62] M. Rudelson, R. Vershynin, Smallest singular value of a random rectangular matrix, Comm. Pure Appl. Math. 62 (12) (2009) 1707–1739.

[63] M. Rudelson, R. Vershynin, Non-asymptotic theory of random matrices: extreme singular values, in: Proceedings of the International Congress of Mathematicians. Volume III, Hindustan Book Agency, New Delhi, 2010, pp. 1576–1602.

[64] M. Rudelson, R. Vershynin, Delocalization of eigenvectors of random matrices with independent entries, Duke Math. J. 164 (13) (2015) 2507–2538.

[65] M. Rudelson, R. Vershynin, Hanson–Wright inequality and sub-Gaussian concentration, Electron. Commun. Probab. 18 (82) (2013) 1–9.

[66] M. Rudelson, R. Vershynin, No-gaps delocalization for general random matrices, preprint, arXiv:1506.04012, 2015.

[67] J. Schenker, Eigenvector localization for random band matrices with power law band width, Comm. Math. Phys. 290 (3) (2009) 1065–1097.

[68] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 888–905.

[69] A. Soshnikov, Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails, Electron. Commun. Probab. 9 (2004) 82–91.

[70] A. Soshnikov, Poisson statistics for the largest eigenvalues in random matrix ensembles, in: Mathematical Physics of Quantum Mechanics, in: Lecture Notes in Phys., vol. 690, Springer, Berlin, 2006, pp. 351–364.

[71] T. Spencer, Random banded and sparse matrices, in: The Oxford Handbook of Random Matrix Theory, Oxford University Press, 2011, pp. 471–488.

[72] M.D. Springer, The Algebra of Random Variables, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York–Chichester–Brisbane, 1979.

[73] R. Takahashi, Normalizing constants of a distribution which belongs to the domain of attraction of the Gumbel distribution, Statist. Probab. Lett. 5 (3) (1987) 197–200.

[74] T. Tao, Topics in Random Matrix Theory, vol. 132, American Mathematical Soc., 2012.

[75] T. Tao, V. Vu, Random matrices: universality of local eigenvalue statistics up to the edge, Comm. Math. Phys. 298 (2) (2010) 549–572.

[76] T. Tao, V. Vu, Random matrices: universality of local eigenvalue statistics, Acta Math. 206 (1) (2011) 127–204.

[77] T. Tao, V. Vu, The Wigner–Dyson–Mehta bulk universality conjecture for Wigner matrices, Electron. J. Probab. 16 (77) (2011) 2104–2121.

[78] T. Tao, V. Vu, Random covariance matrices: universality of local statistics of eigenvalues, Ann. Probab. 40 (3) (2012) 1285–1315.

[79] T. Tao, V. Vu, Random matrices: universal properties of eigenvectors, Random Matrices Theory Appl. 1 (01) (2012) 1150001.

[80] T. Tao, V. Vu, Random matrices: the universality phenomenon for Wigner ensembles, in: Modern Aspects of Random Matrix Theory, in: Proc. Sympos. Appl. Math., vol. 72, Amer. Math. Soc., Providence, RI, 2014, pp. 121–172.

[81] T. Tao, V. Vu, Random matrices have simple spectrum, available at arXiv:1412.1438.

[82] L.V. Tran, V.H. Vu, K. Wang, Sparse random graphs: eigenvalues and eigenvectors, Random Structures Algorithms 42 (1) (2013) 110–134.

[83] R. Vershynin, Spectral norm of products of random and deterministic matrices, Probab. Theory Related Fields 150 (3–4) (2011) 471–509.

[84] R. Vershynin, Introduction to the non-asymptotic analysis of random matrices, in: Compressed Sensing, Cambridge Univ. Press, Cambridge, 2012, pp. 210–268.

[85] U. von Luxburg, A tutorial on spectral clustering, Stat. Comput. 17 (4) (2007) 395–416.

[86] V. Vu, K. Wang, Random weighted projections, random quadratic forms and random eigenvectors, Random Structures Algorithms 47 (4) (2015) 792–821.

[87] K. Wang, Random covariance matrices: universality of local statistics of eigenvalues up to the edge, Random Matrices Theory Appl. 1 (01) (2012) 1150005.