

Noise sensitivity of the top eigenvector of a Wigner matrix ^{*}

Charles Bordenave[†] Gábor Lugosi ^{‡§¶} Nikita Zhivotovskiy^{||}

March 12, 2019

Abstract

We investigate the noise sensitivity of the top eigenvector of a Wigner matrix in the following sense. Let v be the top eigenvector of an $N \times N$ Wigner matrix. Suppose that k randomly chosen entries of the matrix are resampled, resulting in another realization of the Wigner matrix with top eigenvector $v^{[k]}$. We prove that, with high probability, when $k \ll N^{5/3-o(1)}$, then v and $v^{[k]}$ are almost collinear and when $k \gg N^{5/3}$, then $v^{[k]}$ is almost orthogonal to v .

1 Introduction

In this paper we study the *noise sensitivity* of top eigenvectors of Wigner matrices. For a positive integer N , let $X = (X_{i,j})$ be a symmetric $N \times N$ matrix such that, for $i \leq j$, the $X_{i,j}$ are independent real random variables, such that for some constant

^{*}Gábor Lugosi was supported by the Spanish Ministry of Economy and Competitiveness, Grant MTM2015-67304-P and FEDER, EU; “High-dimensional problems in structured probabilistic models - Ayudas Fundación BBVA a Equipos de Investigación Científica 2017”; and Google Focused Award “Algorithms and Learning for AI”. Charles Bordenave was supported by the research grants ANR-14-CE25-0014 and ANR-16-CE40-0024-01. Nikita Zhivotovskiy was supported by RSF grant No. 18-11-00132.

[†]Institut de Mathématiques de Marseille, CNRS & Aix-Marseille University, Marseille, France.

[‡]Department of Economics and Business, Pompeu Fabra University, Barcelona, Spain, gabor.lugosi@upf.edu

[§]ICREA, Pg. Llus Companys 23, 08010 Barcelona, Spain

[¶]Barcelona Graduate School of Economics

^{||}This work was prepared while Nikita Zhivotovskiy was a postdoctoral fellow at the department of Mathematics, Technion I.I.T. and researcher at National University Higher School of Economics. Now visiting researcher at Google.

$\delta > 0$ and for all $i \leq j$, $\mathbb{E}X_{i,j} = 0$ and $\mathbb{E}\exp(|X_{i,j}|^\delta) \leq 1/\delta$. To guarantee that X is a symmetric matrix, we set $X_{i,j} = X_{j,i}$. Finally, we assume that the off-diagonal entries have the unit variance: for all $i \neq j$, $\mathbb{E}X_{ij}^2 = 1$ and for all i , $\mathbb{E}X_{ii}^2 = \sigma_0^2$, for some $\sigma_0 \geq 0$. Throughout this text, we call such matrix X a Wigner matrix. In this paper we are concerned with large matrices and the main results are asymptotic, concerning $N \rightarrow \infty$. The distribution of the entries $X_{i,j}$ may change with N though we suppress this dependence in the notation. However, the values of σ_0 and δ are assumed to be the same for all N .

Let $\lambda = \sup_{w \in S^{N-1}} \langle w, Xw \rangle$ be the top eigenvalue of X and let v denote the corresponding eigenvector. In this paper we study the noise sensitivity of v . In particular, we are interested in the behavior of the top eigenvector $v^{[k]}$ of the symmetric matrix $X^{[k]}$ obtained by resampling k random entries of X . The main finding of the paper is that, with high probability, when $k \leq N^{5/3-o(1)}$, then v and $v^{[k]}$ are almost collinear and when $k \gg N^{5/3}$, then $v^{[k]}$ is almost orthogonal to v .

Related work and proof technique

Noise sensitivity is an important notion in probability that has been extensively studied since the pioneering work of Benjamini, Kalai, and Schramm [2]. Noise sensitivity has mostly been studied in the context of Boolean functions and it has been shown to have deep connections with threshold phenomena, measure concentration, and isoperimetric inequalities, see Talagrand [22], Friedgut and Kalai [11], Kahn, Kalai, and Linial [16], Bourgain, Kahn, Kalai, Katznelson, and Linial [4] for some of the key early work and Garban [12], Garban and Steif [14], Kalai and Safra [17], O’Donnell [20] for surveys. The key techniques for studying noise sensitivity typically use elements of harmonic analysis, in particular, hypercontractivity ([22], [16]) but also the “randomized algorithm” approach of Schramm and Steif [21] and other techniques, see Garban, Pete, and Schramm [13].

Our approach is inspired by Chatterjee’s work [7] who shows that, for functions of independent standard Gaussian random variables, the notion of noise sensitivity (or “chaos” as Chatterjee calls it) is deeply related to the notion of “superconcentration”.

In fact, a result in a similar spirit to ours for the *Gaussian Unitary Ensemble* was proved by Chatterjee [7, Section 3.6]. However, instead of resampling random entries of the matrix, the perturbations considered in [7] are different. In Chatterjee’s model, *every* entry of the matrix X is perturbed by replacing X by $Y = e^{-t}X + \sqrt{1 - e^{-2t}}X'$ where X' is an independent copy of X and $t > 0$. It is proved in [7] that the top eigenvectors of X and Y are approximately orthogonal (in the sense that the expectation of their inner product goes to zero as $N \rightarrow \infty$) as soon as $t \gg N^{-1/3}$.

Chatterjee uses this example to illustrate how “superconcentration” implies “chaos”. His techniques crucially depend on the Gaussian assumption as in that case explicit formulas may be exploited. Our techniques are similar in the sense that our starting point is also “superconcentration” (i.e., the fact that the variance of the largest eigenvalue of a Wigner matrix is small). However, outside of the Gaussian realm, the notions of superconcentration and chaos are murkier. Starting from a general formula for the variance of a function of independent random variables, due to Chatterjee [5], we establish a monotonicity lemma that allows us to make the connection between the variance of the top eigenvalue and the inner product of interest. Then we use the fact that the top eigenvector has a small variance (i.e., in a sense, it is “superconcentrated”). The monotonicity lemma may be of independent interest and it may have further uses when one tries to prove that “superconcentration implies chaos” for functions of independent—not necessarily Gaussian—random variables.

Result

To formally describe the setup, let X be a symmetric $N \times N$ Wigner matrix as defined above. For a positive integer $k \leq \binom{N}{2} + N = N(N+1)/2$, let the random matrix $X^{[k]}$ be defined as follows. Let $S_k = \{(i_1, j_1), \dots, (i_k, j_k)\}$ be a set of k pairs chosen uniformly at random (without replacement) from the set of all ordered pairs (i, j) of indices with $1 \leq i \leq j \leq N$. The entries of $X^{[k]}$ below the diagonal are

$$X_{i,j}^{[k]} = \begin{cases} X'_{i,j} & \text{if } (i, j) \in S_k \\ X_{i,j} & \text{otherwise,} \end{cases}$$

where $(X'_{i,j})_{1 \leq i \leq j \leq N}$ are independent random variables, independent of X and $X'_{i,j}$ has the same distribution as $X_{i,j}$, for all $i \leq j$. In words, $X^{[k]}$ is obtained from X by resampling k random entries of the matrix below the diagonal and also the corresponding terms above and including the diagonal. Clearly, $X^{[k]}$ has the same distribution as X . Denote unit eigenvectors corresponding to the largest eigenvalues of X and $X^{[k]}$ by v and $v^{[k]}$, respectively.

Our main results are the following.

Theorem 1. *Assume that X is a Wigner matrix as above. If $k/N^{5/3} \rightarrow \infty$, then*

$$\mathbb{E} \left| \langle v, v^{[k]} \rangle \right| = o(1).$$

Conversely, our second result asserts that when $k \leq N^{5/3-o(1)}$ then v and $v^{[k]}$ are almost aligned.

Theorem 2. *Assume that X is a Wigner matrix as above. There exists a constant $c > 0$ such that, with $\varepsilon_N = (\log N)^{-c \log \log N}$,*

$$\mathbb{E} \max_{1 \leq k \leq \varepsilon_N N^{5/3}} \min_{s \in \{-1, 1\}} \|v - sv^{[k]}\|_2 = o(1).$$

The proof of Theorem 2 actually establishes that $\max_k \min_s \sqrt{N} \|v - sv^{[k]}\|_\infty$ goes to 0 in probability.

The following heuristic argument may provide an intuition of why the threshold is at $k = N^{5/3+o(1)}$. Since the seminal work of Erdős, Schlein, and Yau [10], it is well known that unit eigenvectors of random matrices are delocalized in the sense that $\|v\|_\infty = N^{-1/2+o(1)}$ with high probability. From this, we might infer from the derivative of a simple eigenvalue as the function of the matrix entries that

$$\lambda^{[1]} - \lambda \simeq (1 + \mathbb{1}(i_1 \neq j_1)) v_{i_1} X_{i_1, j_1} v_{j_1} \simeq \frac{X_{i_1, j_1}}{N^{1+o(1)}},$$

where v_i is the i -th component of v . Assuming that v_i is nearly independent of any matrix entry X_{ij} , since X_{ij} is centered with unit variance, we would get from the central limit theorem that

$$\lambda^{[k]} - \lambda = \sum_{t=0}^{k-1} (\lambda^{[t+1]} - \lambda^{[t]}) \simeq \frac{\sqrt{k}}{N^{1+o(1)}}.$$

On the other hand, the known behavior of random matrices at the edge of the spectrum implies that the second largest eigenvalue of X is at distance of order $N^{-1/6}$ from λ . The above heuristic should thus break down when $\sqrt{k}/N^{1+o(1)}$ is of order $N^{-1/6}$. It gives the threshold at $k = N^{5/3+o(1)}$.

Remark. We expect that the arguments of Theorem 1 for the noise sensitivity of the top eigenvalue may be modified to prove analogous results for the eigenvector corresponding to the j -th largest eigenvalue, $1 \leq j \leq N$. However, the threshold is expected to occur at values different from $N^{5/3}$. In particular, a simple heuristic argument suggests that for the j -th eigenvector the threshold occurs around $N^{5/3+o(1)} \min(j, N-j+1)^{-2/3}$. However, to keep the presentation transparent, in this paper we focus on the top eigenvalue.

Interestingly, the proof that the top eigenvalue is very sensitive to resampling more than $\Theta(N^{5/3})$ entries involves proving that it is insensitive to resampling just a single entry. As a consequence the proofs of Theorems 1 and 2 share common techniques.

The rest of the paper is dedicated to proving Theorems 1 and 2. In Section 2 we introduce a general tool for proving noise sensitivity that generalizes Chatterjee's ideas based of "superconcentration" to functions of independent, not

necessarily standard normal random variables. In Section 3 we summarize some of the tools from random matrix theory that are crucial for our arguments. In Sections 4 and 5 we give the proofs of Theorems 1 and 2.

2 Variance and noise sensitivity

The first building block in the proof of Theorem 1 is a formula for the variance of an arbitrary function of independent random variables, due to Chatterjee [5]. For any positive integer i , denote $[i] = \{1, \dots, i\}$.

Lemma 1. [5] *Let X_1, \dots, X_n be independent random variables taking values in some set \mathcal{X} and let $f : \mathcal{X}^n \rightarrow \mathbb{R}$ be a measurable function. Denote $X = (X_1, \dots, X_n)$. Let $X' = (X'_1, \dots, X'_n)$ be an independent copy of X . Under the notation*

$$X^{(i)} = (X_1, \dots, X_{i-1}, X'_i, X_{i+1}, \dots, X_n) \quad \text{and} \quad X^{[i]} = (X'_1, \dots, X'_i, X_{i+1}, \dots, X_n)$$

and, in particular, $X^{[0]} = X$ and $X^{[n]} = X'$, we have

$$\text{Var}(f(X)) = \frac{1}{2} \sum_{i=1}^n \mathbb{E} \left[(f(X) - f(X^{(i)})) (f(X^{[i-1]}) - f(X^{[i]})) \right].$$

In general, for $A \subseteq [n]$ let X^A denote the random vector, obtained from X by replacing the components indexed by A by corresponding components of X' .

In the variance formula above, the order of the variables does not matter and the formula remains valid after permuting the indices $1, \dots, n$ arbitrarily. In particular, one may take the variables in random order. Thus, if $\sigma = (\sigma(1), \dots, \sigma(n))$ is a random permutation sampled uniformly from the symmetric group S_n and $\sigma([i])$ denotes $\{\sigma(1), \dots, \sigma(i)\}$, then

$$\text{Var}(f(X)) = \frac{1}{2} \sum_{i=1}^n \mathbb{E} \left[(f(X) - f(X^{\sigma([i])})) (f(X^{\sigma([i-1])}) - f(X^{\sigma([i])})) \right]. \quad (2.1)$$

Note that on the right-hand side of (2.1) the expectation is taken with respect to both X, X' , and the random permutation σ .

One would intuitively expect that the terms on the right-hand side of (2.1) decrease with i , as the differences $f(X) - f(X^{\sigma([i])})$ and $f(X^{\sigma([i-1])}) - f(X^{\sigma([i])})$ become less correlated as more randomly chosen components get resampled. This is indeed the case and this fact is one of our main tools in proving noise sensitivity. We believe that the following lemma can be useful in diverse situations. The proof is given in Section 4.1 below.

Lemma 2. Consider the setup of Lemma 1 and the notation above. In addition, assume that X_1, \dots, X_n are identically distributed. For $i \in [n]$, denote

$$B_i = \mathbb{E} \left[\left(f(X) - f(X^{\sigma(i)}) \right) \left(f(X^{\sigma([i-1])}) - f(X^{\sigma([i])}) \right) \right],$$

where the expectation is taken with respect to components of vectors and random permutations. Then $B_i \geq B_{i+1}$ for all $i = 1, \dots, n-1$ and $B_n \geq 0$. In particular, for any $k \in [n]$,

$$B_k \leq \frac{2\text{Var}(f(X))}{k}.$$

We also introduce a modification of Lemma 2 that will be more convenient for our purposes. To do so, we introduce the following notation. Let j have uniform distribution on $[n]$. Let $X^{(j) \circ \sigma([i-1])}$ denote the vector obtained from $X^{\sigma([i-1])}$ by replacing its j -th component by an independent copy of the random variable X_j , denoted by X_j'' . Observe that j may belong to $\sigma([i-1])$ and in this case X_j'' is independent of X_j' appearing in $X^{\sigma([i-1])}$. With this notation in mind we may prove the following version of Lemma 2.

Lemma 3. Using the notation of Lemma 2, assuming that j is chosen uniformly at random from the set $[n]$ and independently of other random variables involved, we have for any $k \in [n]$,

$$B'_k \leq \frac{2\text{Var}(f(X))}{k} \left(\frac{n+1}{n+k-1} \right),$$

where for any $i \in [n]$,

$$B'_i = \mathbb{E} \left[\left(f(X) - f(X^{(j)}) \right) \left(f(X^{\sigma([i-1])}) - f(X^{(j) \circ \sigma([i-1])}) \right) \right].$$

3 Random matrix results

In the proof of Theorem 1 we apply Lemma 3 with f being the top eigenvalue of a Wigner matrix. The usefulness of this bound crucially hinges on the fact that the variance of the top eigenvalue is small, that is, in a sense, the top eigenvalue is “superconcentrated”. This fact is quantified in this section.

Our first lemma on the variance of λ is obtained as a combination of a result of Ledoux and Rider [19] on Gaussian ensembles and the universality of fluctuations for Wigner matrices as stated in Erdős, Yau and Yin [9].

Lemma 4. Assume that X is a Wigner matrix as in Theorem 1. Let λ denote the largest eigenvalue of X . Then,

$$\text{Var}(\lambda) \leq (c + o(1))N^{-1/3},$$

where $c > 0$ is an absolute constant.

We also need the following delocalization result of the top eigenvector of a Wigner matrix which can be found in Tao and Vu [23, Proposition 1.12].

Lemma 5. [23]. *Assume that X is a Wigner matrix as in Theorem 1. For any real $c_0 > 0$, there exists a constant $C > 0$, such that, with probability at least $1 - CN^{-c_0}$, any eigenvector w of X with $\|w\|_2 = 1$ satisfies*

$$\|w\|_\infty \leq \frac{(\log N)^C}{\sqrt{N}}.$$

Our final lemma is a perturbation inequality in ℓ^∞ -norm of the top eigenvector of a Wigner matrix when a single entry is re-sampled. The proof uses precise estimates on the eigenvalue spacings in Wigner matrices proved in Tao and Vu [23] and Erdős, Yau, and Yin [9].

Lemma 6. *Let X be a Wigner matrix as in Theorem 1 and X' an independent copy of X . For any (i, j) with $1 \leq i, j \leq N$. Denote by $Y^{(ij)}$ the symmetric matrix obtained from X by replacing the entry X_{ij} by X'_{ij} and X_{ji} by X'_{ji} . For any $0 < \alpha < 1/10$, there exists $\kappa > 0$ such that, for all N large enough, with probability at least $1 - N^{-\kappa}$,*

$$\max_{1 \leq i, j \leq N} \inf_{s \in \{-1, 1\}} \|sv - u^{(ij)}\|_\infty \leq N^{-\frac{1}{2} - \alpha},$$

where v and $u^{(ij)}$ are any unit eigenvectors corresponding to the largest eigenvalues of X and $Y^{(ij)}$.

4 Proof of Theorem 1

Now we are ready for the proof of the main results of the paper.

We start by fixing some notation. Let λ denote the largest eigenvalue of the Wigner matrix X of Theorem 1 and let $v \in S^{N-1}$ be a corresponding normalized eigenvector. Let $k \in \left[\binom{N}{2} + N \right]$ to be specified later and let $X^{[k]}$ be the random symmetric matrix obtained by resampling k random entries below the diagonal and including the diagonal, as defined in the introduction. We denote by $S_k \subset \left[\binom{N}{2} + N \right]$ the set of random positions of the k resampled entries. Let $\lambda^{[k]}$ denote the top eigenvalue of $X^{[k]}$ and $v^{[k]}$ a corresponding normalized eigenvector.

For $1 \leq i \leq j \leq N$, we denote by $Y^{(ij)}$ the symmetric matrix obtained from X by replacing the entry X_{ij} by X''_{ij} where X'' is an independent copy of X . We obtain $Y^{[k]}_{(ij)}$ from $X^{[k]}$ by the same operation. We denote by $(\mu_{(ij)}, u_{(ij)})$, and $(\mu^{[k]}_{(ij)}, u^{[k]}_{(ij)})$ the top eigenvalue/eigenvector pairs of $Y_{(ij)}$ and $Y^{[k]}_{(ij)}$, respectively. Let (s, t) be a pair

of indices with $1 \leq t \leq s \leq N$ chosen uniformly at random from $\binom{N}{2} + N$. For ease of notation, we set $Y = Y_{(st)}$, $\mu = \mu_{(st)}$ and $u = u_{(st)}$. We define similarly $Y^{[k]} = Y_{(st)}^{[k]}$, $\mu^{[k]} = \mu_{(st)}^{[k]}$ and $u^{[k]} = u_{(st)}^{[k]}$.

By applying Lemma 3 to the function of $n = \binom{N}{2} + N$ independent random variables $f((X_{i,j})_{1 \leq i \leq j \leq N}) = \lambda$, we obtain that, for any $k \in \binom{N}{2} + N$,

$$\frac{2\text{Var}(\lambda)}{k} \cdot \frac{\binom{N}{2} + N + 1}{\binom{N}{2} + N + k - 1} \geq \mathbb{E}[(\lambda - \mu)(\lambda^{[k]} - \mu^{[k]})]. \quad (4.1)$$

Now we are ready to prove Theorem 1. Using the notation of the previous section we have

$$\mathbb{E}[(\lambda - \mu)(\lambda^{[k]} - \mu^{[k]})] = \mathbb{E}[(\langle v, Xv \rangle - \langle u, Yu \rangle)(\langle v^{[k]}, X^{[k]}v^{[k]} \rangle - \langle u^{[k]}, Y^{[k]}u^{[k]} \rangle)].$$

Observe that $Z = X - Y = X^{[k]} - Y^{[k]}$ and the elements of $X - Y$ are all zeros except at most two that correspond to resampled values. Moreover, using the fact that v maximizes $\langle v, Xv \rangle$ and u maximizes $\langle u, Yu \rangle$ we have

$$\langle u, (X - Y)u \rangle \leq \langle v, Xv \rangle - \langle u, Yu \rangle \leq \langle v, (X - Y)v \rangle$$

If the element $X_{t,s}$ of X was resampled to get Y , we have, for any vector x ,

$$\langle x, (X - Y)x \rangle = Z_{t,s}x_t x_s$$

and $Z_{t,s} = (X_{t,s} - X'_{t,s})(1 + \mathbb{1}(t \neq s))$. Therefore, it is straightforward to see that

$$\begin{aligned} & (\langle v, Xv \rangle - \langle u, Yu \rangle)(\langle v^{[k]}, X^{[k]}v^{[k]} \rangle - \langle u^{[k]}, Y^{[k]}u^{[k]} \rangle) \\ & \geq Z_{t,s}^2 \min\{v_t v_s v_t^{[k]} v_s^{[k]}, u_t u_s v_t^{[k]} v_s^{[k]}, v_t v_s u_t^{[k]} u_s^{[k]}, u_t u_s u_t^{[k]} u_s^{[k]}\}. \end{aligned}$$

Fix $0 < \alpha < 1/10$ and let C be as in Lemma 5 for $c_0 = 10$. We define $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2$ to be the union of following two events:

- \mathcal{E}_1 : $\max_{i,j} \|v - u_{(ij)}\|_\infty \leq \frac{1}{N^{\frac{1}{2}+\alpha}}$ and $\max_{i,j} \|v^{[k]} - u_{(ij)}^{[k]}\|_\infty \leq \frac{1}{N^{\frac{1}{2}+\alpha}}$.
- \mathcal{E}_2 : $\|y\|_\infty \leq \frac{(\log N)^C}{\sqrt{N}}$ for all $y \in \{v, u_{(ij)}, v^{[k]}, u_{(ij)}^{[k]} : 1 \leq i, j \leq N\}$.

By Lemmas 5, 6, and the union bound, we have, for all N large enough, $\mathbb{P}(\mathcal{E}_2^c) \leq N^{-6}$ and for some $\kappa > 0$, $\mathbb{P}(\mathcal{E}^c) \leq N^{-\kappa}$ (provided that we choose properly the \pm -phase for the eigenvectors u and $u^{[k]}$). Observe that when \mathcal{E} holds for

$$x, y \in \{v_t v_s v_t^{[k]} v_s^{[k]}, u_t u_s v_t^{[k]} v_s^{[k]}, v_t v_s u_t^{[k]} u_s^{[k]}, u_t u_s u_t^{[k]} u_s^{[k]}\}, \quad (4.2)$$

we have, for all N large enough,

$$|x - y| \leq \frac{4(\log N)^{3C}}{N^{2+\alpha}}.$$

We show this, for brevity, only for $v_t v_s v_t^{[k]} v_s^{[k]}$ and $u_t u_s u_t^{[k]} u_s^{[k]}$. Denoting $\delta_t = v_t - u_t$ and $\delta_t^{[k]} = v_t^{[k]} - u_t^{[k]}$, we write

$$u_t u_s u_t^{[k]} u_s^{[k]} = (v_t - \delta_t)(v_s - \delta_s)(v_t^{[k]} - \delta_t^{[k]})(v_s^{[k]} - \delta_s^{[k]}),$$

then open the brackets and use that, on \mathcal{E} ,

$$\max\{|\delta_t|, |\delta_s|, |\delta_t^{[k]}|, |\delta_s^{[k]}|\} \leq N^{-\frac{1}{2}-\alpha} \quad \text{and} \quad \max\{|v_t|, |v_s|, |v_t^{[k]}|, |v_s^{[k]}|\} \leq (\log N)^C / \sqrt{N}.$$

Let $\mathbb{E}_{s,t}$ denote the conditional expectation given (X, X', S_k) (that is we integrate only on the random pair (s, t)). Note that the event \mathcal{E} is measurable with respect to (X, X', S_k) . If \mathcal{E} holds, we thus have

$$\begin{aligned} & \mathbb{E}_{s,t} Z_{t,s}^2 \min \left\{ v_t v_s v_t^{[k]} v_s^{[k]}, u_t u_s v_t^{[k]} v_s^{[k]}, v_t v_s u_t^{[k]} u_s^{[k]}, u_t u_s u_t^{[k]} u_s^{[k]} \right\} \\ & \geq \mathbb{E}_{s,t} \left(Z_{t,s}^2 v_t v_s v_t^{[k]} v_s^{[k]} \right) - \frac{4(\log N)^{3C}}{N^{2+\alpha}} \mathbb{E}_{s,t} Z_{t,s}^2. \end{aligned}$$

If $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2$ does not hold, we have using that all the vectors are of unit norm (and therefore $\max\{|v_t|, |v_s|, |v_t^{[k]}|, |v_s^{[k]}|\} \leq 1$),

$$\begin{aligned} & \mathbb{E}_{s,t} Z_{t,s}^2 \min \left\{ v_t v_s v_t^{[k]} v_s^{[k]}, u_t u_s v_t^{[k]} v_s^{[k]}, v_t v_s u_t^{[k]} u_s^{[k]}, u_t u_s u_t^{[k]} u_s^{[k]} \right\} \\ & \geq -\frac{(\log N)^{4C}}{N^2} \mathbb{1}(\mathcal{E}^c) \mathbb{E}_{s,t} Z_{t,s}^2 - \mathbb{1}(\mathcal{E}_2^c) \mathbb{E}_{s,t} Z_{t,s}^2. \end{aligned}$$

Note that $\mathbb{E} Z_{t,s}^4 \leq c_1^2$ for some constant c_1 depending on δ . Altogether, recalling (4.1), by the Cauchy-Schwarz inequality and using $\frac{\binom{N}{2} + N + 1}{\binom{N}{2} + N + k - 1} \leq 1$,

$$\mathbb{E} Z_{t,s}^2 v_t v_s v_t^{[k]} v_s^{[k]} \leq \frac{2\text{Var}(\lambda)}{k} + 4c_1 \frac{(\log N)^3}{N^{2+\alpha}} + c_1 \frac{(\log N)^{4C}}{N^2} \sqrt{\mathbb{P}(\mathcal{E}^c)} + c_1 \sqrt{\mathbb{P}(\mathcal{E}_2^c)}.$$

Observe that v depends only on X as well as $v^{[k]}$ depends only on $X^{[k]}$ and both do not depend on the random choice of (s, t) . Therefore,

$$\mathbb{E}_{s,t} Z_{t,s}^2 v_t v_s v_t^{[k]} v_s^{[k]} = \frac{1}{\binom{N}{2} + N} \sum_{1 \leq i < j \leq N} Z_{i,j}^2 v_i v_j v_i^{[k]} v_j^{[k]}. \quad (4.3)$$

Now, using (4.3), we get

$$\mathbb{E} \left(\sum_{1 \leq i, j \leq N} \tilde{Z}_{i,j}^2 v_i v_j v_i^{[k]} v_j^{[k]} \right) \leq 2N^2 \frac{\text{Var}(\lambda)}{k} + \varepsilon_N, \quad (4.4)$$

where $\tilde{Z}_{i,j} = Z_{i,j}/2$ if $i \neq j$, $\tilde{Z}_{i,i} = Z_{i,i}$ and

$$\varepsilon_N = 4c_1 \frac{(\log N)^3}{N^\alpha} + c_1 (\log N)^{4C} \sqrt{\mathbb{P}(\mathcal{E}^c)} + c_1 N^2 \sqrt{\mathbb{P}(\mathcal{E}_2^c)}.$$

Note that for $i \neq j$, $\mathbb{E} \tilde{Z}_{i,j}^2 = 2$ and $\mathbb{E} \tilde{Z}_{i,i}^2 = \sigma_0^2$. We would like to replace in (4.4), the variables $\tilde{Z}_{i,j}^2$ by 2. We have

$$\mathbb{E} \left(\sum_{i=1}^N v_i v_i v_i^{[k]} v_i^{[k]} \right) \leq \frac{(\log N)^{4C}}{N} + N \mathbb{P}(\mathcal{E}_2^c).$$

Setting $W_{i,j} = \mathbb{E} \tilde{Z}_{i,j}^2 - \tilde{Z}_{i,j}^2$, we deduce that

$$2\mathbb{E} \left(\sum_{1 \leq i, j \leq N} v_i v_j v_i^{[k]} v_j^{[k]} \right) \leq 2N^2 \frac{\text{Var}(\lambda)}{k} + \mathbb{E} \left(\sum_{1 \leq i, j \leq N} W_{i,j} v_i v_j v_i^{[k]} v_j^{[k]} \right) + \varepsilon'_N, \quad (4.5)$$

where

$$\varepsilon'_N = \varepsilon_N + |2 - \sigma_0^2| \frac{(\log N)^{4C}}{N} + N|2 - \sigma_0^2| \mathbb{P}(\mathcal{E}_2^c).$$

From Lemmas 5 and 6, when $k/N^{5/3} \rightarrow \infty$, ε'_N goes to 0 with N . We would also like get rid of the second term on the right-hand side of (4.5): the variables $W_{i,j}$ are centered but they are not independent of v and $v^{[k]}$. We may circumvent this difficulty by a new application of Lemma 6. Let X'' be an independent copy of X , independent of X' . For $1 \leq i, j \leq N$, let $w_{(ij)}$ and $w_{(ij)}^{[k]}$ be the top eigenvectors of the symmetric matrices obtained from X and $X^{[k]}$ by replacing the entry (i, j) by $X''_{i,j}$. By construction, $w_{(ij)}^{[k]}$ is independent of $W_{i,j}$. Let C be as in Lemma 5 for $c_0 = 10$. We define $\mathcal{E}' = \mathcal{E}'_1 \cup \mathcal{E}'_2$ as the union of following two events:

- \mathcal{E}'_1 : $\max_{i,j} \|v - w_{(ij)}\|_\infty \leq \frac{1}{N^{\frac{1}{2}+\alpha}}$ and $\max_{i,j} \|v^{[k]} - w_{(ij)}^{[k]}\|_\infty \leq \frac{1}{N^{\frac{1}{2}+\alpha}}$.
- \mathcal{E}'_2 : $\|y\|_\infty \leq \frac{(\log N)^C}{\sqrt{N}}$ for all $y \in \{v, w_{(ij)}, v^{[k]}, w_{(ij)}^{[k]} : 1 \leq i, j \leq N\}$.

By Lemmas 5, 6, and the union bound, we have, for all N large enough, $\mathbb{P}(\mathcal{E}'^c) \leq N^{-6}$ and $\mathbb{P}(\mathcal{E}'^c) \leq N^{-\kappa}$ (provided that we choose properly the \pm -phase for the eigenvectors w and $w^{[k]}$). The argument below (4.2) implies that, if \mathcal{E}' holds, for all $1 \leq i, j \leq N$,

$$|v_i v_j v_i^{[k]} v_j^{[k]} - (w_{(ij)})_i (w_{(ij)})_j (w_{(ij)}^{[k]})_i (w_{(ij)}^{[k]})_j| \leq \frac{4(\log N)^{3C}}{N^{2+\alpha}}.$$

Therefore, since $\mathbb{E}W_{i,j}(w_{(ij)})_i (w_{(ij)})_j (w_{(ij)}^{[k]})_i (w_{(ij)}^{[k]})_j = 0$, we get

$$\mathbb{E} \left(\sum_{1 \leq i, j \leq N} W_{i,j} v_i v_j v_i^{[k]} v_j^{[k]} \right) \leq c_2 \frac{4(\log N)^{3C}}{N^\alpha} + c_2 (\log N)^{4C} \sqrt{\mathbb{P}(\mathcal{E}'^c)} + c_2 N^2 \sqrt{\mathbb{P}(\mathcal{E}'^c)}.$$

where $c_2^2 = \max_{i,j} \mathbb{E}|W_{i,j}|^2 \leq 4c_1^2$. So finally from (4.5), we get

$$2\mathbb{E} \left(\sum_{1 \leq i, j \leq N} v_i v_j v_i^{[k]} v_j^{[k]} \right) \leq 2N^2 \frac{\text{Var}(\lambda)}{k} + \varepsilon_N'', \quad (4.6)$$

where $\varepsilon_N'' \rightarrow 0$ with N .

Now, combining Jensen's inequality and (4.6),

$$\left(\mathbb{E} \left| \langle v, v^{[k]} \rangle \right| \right)^2 \leq \mathbb{E} \left(\sum_{i=1}^N v_i v_i^{[k]} \right)^2 \leq \mathbb{E} \left(\sum_{1 \leq i, j \leq N} v_i v_j v_i^{[k]} v_j^{[k]} \right) \leq N^2 \frac{\text{Var}(\lambda)}{k} + \frac{\varepsilon_N''}{2}.$$

From Lemma 4, the claim follows.

4.1 Proof of Lemma 2 and Lemma 3

We start with the following technical lemma.

Lemma 7. *Let $f : \mathcal{X}^n \rightarrow \mathbb{R}$ be a measurable function and let $\sigma \in S_n$ be any fixed permutation. Fix $i \in [n-1]$ and $j \in [n]$ such that $j \notin \sigma([i])$. Let X_1, \dots, X_n be independent random variables taking values in \mathcal{X} . Then*

$$\begin{aligned} A_i &= \mathbb{E} \left[\left(f(X) - f(X^{\sigma(i)}) \right) \left(f(X^{\sigma([i-1])}) - f(X^{\sigma([i])}) \right) \right] \\ &\geq \mathbb{E} \left[\left(f(X) - f(X^{\sigma(i)}) \right) \left(f(X^{\sigma([i-1]) \cup j}) - f(X^{\sigma([i]) \cup j}) \right) \right] \\ &\geq 0. \end{aligned}$$

Proof. Without loss of generality, we may consider one particular permutation σ , defined as follows: set $\sigma(k) = k$ for $k \notin \{1, i\}$, $\sigma(i) = 1$, $\sigma(1) = i$, and we may also assume that $j = i + 1$. The proof is identical for any other σ and j . In our case,

$$A_i = \mathbb{E}\left[\left(f(X) - f(X^{(1)})\right)\left(f(X^{[i]\setminus\{1\}}) - f(X^{[i]})\right)\right].$$

Moreover, we have

$$\mathbb{E}\left[\left(f(X) - f(X^{(\sigma(i))})\right)\left(f(X^{\sigma([i-1]\cup j)} - f(X^{\sigma([i]\cup j)})\right)\right] = A_{i+1}.$$

We introduce a simplifying notation. Denote $B = (X_2, \dots, X_i)$, $B' = (X'_2, \dots, X'_i)$ and $C = (X_{i+2}, \dots, X_n)$. Therefore, we may rewrite

$$A_i = \mathbb{E}\left[\left(f(X_1, B, X_{i+1}, C) - f(X'_1, B, X_{i+1}, C)\right)\left(f(X_1, B', X_{i+1}, C) - f(X'_1, B', X_{i+1}, C)\right)\right]$$

and

$$A_{i+1} = \mathbb{E}\left[\left(f(X_1, B, X_{i+1}, C) - f(X'_1, B, X_{i+1}, C)\right)\left(f(X_1, B', X'_{i+1}, C) - f(X'_1, B', X'_{i+1}, C)\right)\right].$$

Denote $h(X_1, X'_1, X_{i+1}, C) = \mathbb{E}\left[\left(f(X_1, B, X_{i+1}, C) - f(X'_1, B, X_{i+1}, C)\right) \mid X_1, X'_1, X_{i+1}, C\right]$. Using the independence of B, B' and their independence of the remaining random variables, we have

$$A_i = \mathbb{E}h(X_1, X'_1, X_{i+1}, C)^2.$$

At the same time, using the same notation for h we have, by the Cauchy-Shwarz inequality and the fact that X_{i+1} and X'_{i+1} have the same distribution,

$$\begin{aligned} A_{i+1} &= \mathbb{E}h(X_1, X'_1, X_{i+1}, C)h(X_1, X'_1, X'_{i+1}, C) \\ &= \mathbb{E}\left[\mathbb{E}[h(X_1, X'_1, X_{i+1}, C)h(X_1, X'_1, X'_{i+1}, C) \mid X_1, X'_1, C]\right] \\ &\leq \mathbb{E}h(X_1, X'_1, X_{i+1}, C)^2 \\ &= A_i. \end{aligned}$$

Now to prove that $A_i \geq 0$, it is sufficient to show that $A_n \geq 0$. Denoting $g(X_1) = \mathbb{E}[f(X) \mid X_1]$, we have

$$\begin{aligned} A_n &= \mathbb{E}\left[\left(f(X) - f(X^{(1)})\right)\left(f(X^{[n]\setminus\{1\}}) - f(X^{[n]})\right)\right] \\ &= \mathbb{E}\left(f(X)f(X^{[n]\setminus\{1\}}) - f(X)f(X^{[n]}) - f(X^{(1)})f(X^{[n]\setminus\{1\}}) + f(X^{(1)})f(X^{[n]})\right) \\ &= 2\mathbb{E}f(X)f(X^{[n]\setminus\{1\}}) - 2(\mathbb{E}f(X))^2 \\ &= 2\mathbb{E}\left[\mathbb{E}[f(X)f(X^{[n]\setminus\{1\}}) \mid X_1]\right] - 2(\mathbb{E}f(X))^2 \\ &= 2\mathbb{E}[g(X_1)^2] - 2(\mathbb{E}f(X))^2 \\ &\geq 0, \end{aligned}$$

where we used Jensen's inequality and that $\mathbb{E}g(X_1) = \mathbb{E}f(X)$. ■

We proceed with the proof of Lemma 2.

Proof. In this proof by writing $i + 1$ we mean $i + 1 \pmod n$. For each permutation $\sigma \in S_n$ and fixed $i \in [n]$ we construct a corresponding permutation σ' by defining $\sigma'(i) = \sigma(i + 1)$, $\sigma'(i + 1) = \sigma(i)$ and $\sigma'(k) = \sigma(k)$ for $k \neq \{i, i + 1\}$.

It is straightforward to see that for any fixed i there is a one-to-one correspondence between $\sigma \in S_n$ and σ' . By observing that $\sigma'([i]) = \sigma([i - 1]) \cup \sigma(i + 1)$ and $\sigma'([i + 1]) = \sigma([i + 1])$ we have, conditionally on σ ,

$$\begin{aligned} & \mathbb{E}\left[\left(f(X) - f(X^{\sigma(i)})\right)\left(f(X^{\sigma([i-1])}) - f(X^{\sigma([i])})\right)\right] \\ &= \mathbb{E}\left[\left(f(X) - f(X^{\sigma'(i+1)})\right)\left(f(X^{\sigma([i-1])}) - f(X^{\sigma([i])})\right)\right] \\ &\geq \mathbb{E}\left[\left(f(X) - f(X^{\sigma'(i+1)})\right)\left(f(X^{\sigma'([i])}) - f(X^{\sigma'([i+1])})\right)\right], \end{aligned}$$

where in the last step we used Lemma 7. Using the one to one correspondence between all σ and σ' , we have

$$\begin{aligned} B_i &= \mathbb{E}_\sigma \mathbb{E}\left[\left(f(X) - f(X^{\sigma(i)})\right)\left(f(X^{\sigma([i-1])}) - f(X^{\sigma([i])})\right)\right] \\ &= \frac{1}{n!} \sum_\sigma \mathbb{E}\left[\left(f(X) - f(X^{\sigma(i)})\right)\left(f(X^{\sigma([i-1])}) - f(X^{\sigma([i])})\right)\right] \\ &\geq \frac{1}{n!} \sum_{\sigma'} \mathbb{E}\left[\left(f(X) - f(X^{\sigma'(i+1)})\right)\left(f(X^{\sigma'([i])}) - f(X^{\sigma'([i+1])})\right)\right] \\ &= \mathbb{E}_\sigma \mathbb{E}\left[\left(f(X) - f(X^{\sigma(i+1)})\right)\left(f(X^{\sigma([i])}) - f(X^{\sigma([i+1])})\right)\right] \\ &= B_{i+1}. \end{aligned}$$

The proof that $B_n \geq 0$ follows from Lemma 7 as well. ■

Finally, we prove Lemma 3.

Proof. To prove this Lemma we show a precise recursive formula for B'_i . We have,

$$\begin{aligned} B'_i &= \mathbb{E}\left[\left(f(X) - f(X^{(j)})\right)\left(f(X^{\sigma([i-1])}) - f(X^{(j) \circ \sigma([i-1])})\right)\right] \\ &= \mathbb{E}_\sigma \left(\mathbb{E}\left[\left(f(X) - f(X^{(j)})\right)\left(f(X^{\sigma([i-1] \setminus \{j\})}) - f(X^{(j) \circ \sigma([i-1] \setminus \{j\})})\right)\right] \mathbb{P}(j \in \sigma(i-1)) \right. \\ &\quad \left. + \mathbb{E}_\sigma \left(\mathbb{E}\left[\left(f(X) - f(X^{(j)})\right)\left(f(X^{\sigma([i-1] \setminus \{j\})}) - f(X^{(j) \circ \sigma([i-1] \setminus \{j\})})\right)\right] \mathbb{P}(j \notin \sigma(i-1)) \right) \right). \end{aligned}$$

We have $\mathbb{P}(j \in \sigma(i-1)) = \frac{i-1}{n}$ and the second summand is equal to $B_i \frac{n-i+1}{n}$. We proceed with the first summand. For $i \geq 1$, we have

$$\begin{aligned} & \mathbb{E}_\sigma \mathbb{E}\left[\left(f(X) - f(X^{(j)})\right)\left(f(X^{\sigma([i-1] \setminus \{j\})}) - f(X^{(j) \circ \sigma([i-1] \setminus \{j\})})\right)\right] \mathbb{P}(j \in \sigma(i-1)) \\ &= \mathbb{E}_\sigma \mathbb{E}\left[\left(f(X) - f(X^{(j)})\right)\left(f(X^{\sigma([i-1] \setminus \{j\})}) - f(X^{(j) \circ \sigma([i-1] \setminus \{j\})})\right)\right] \mathbb{P}(j \in \sigma(i-1)) \\ &\quad + \mathbb{E}_\sigma \mathbb{E}\left[\left(f(X) - f(X^{(j)})\right)\left(f(X^{\sigma([i-1] \setminus \{j\})}) - f(X^{\sigma([i-1] \setminus \{j\})})\right)\right] \mathbb{P}(j \in \sigma(i-1)) \\ &= B_{i-1} - B'_i. \end{aligned}$$

This implies

$$B'_i = \frac{i-1}{n}(B_{i-1} - B'_i) + \frac{n-i+1}{n}B_i,$$

which, using Lemma 2, leads to

$$B'_i = \frac{n}{n+i-1} \left(\frac{i-1}{n}B_{i-1} + \frac{n-i+1}{n}B_i \right) \leq \frac{2\text{Var}(f(X))}{i} \left(\frac{n+1}{n+i-1} \right).$$

The claim follows. ■

4.2 Proof of Lemma 4

We start with a special case. Let us say that a Wigner matrix as in Theorem 1 is *standard* if for all i , $\mathbb{E}X_{ii}^2 = 2$. In this case, the variance of the entries of X is equal to the variance of the entries of a random matrix Y sampled from the Gaussian Orthogonal Ensemble (GOE). If μ is the largest eigenvalue of Y , it follows from [19, Corollary 3] that for some absolute constant $c > 0$,

$$\text{Var}(\mu) \leq cN^{-1/3}.$$

On the other hand, it follows from [9, Theorem 2.4] (see also [18, Theorem 1.6] for a statement which can be used directly) that,

$$N^{1/3} |\text{Var}(\mu) - \text{Var}(\lambda)| = o(1).$$

We obtain the first claim of the lemma for standard Wigner matrices. To conclude the proof of the lemma for Wigner matrices, it suffices to prove that for any Wigner matrix X , for some $\kappa \geq 1/3$, we have for all N large enough,

$$\mathbb{E}|\lambda - \lambda_0|^2 \leq N^{-\kappa}. \tag{4.7}$$

where λ_0 is the largest eigenvalue of a matrix X_0 obtained from X by setting to 0 all diagonal entries. We will prove it for any $\kappa < 1/2$ (an improvement of the forthcoming Lemma 11 would give (4.7) for any $\kappa < 1$). The proof requires some care since the operator norm of $X - X_0$ may be much larger than 1 and the rank of $X - X_0$ could be N .

There is an easy inequality which is half of (4.7). Let v_0 be a unit eigenvector of X_0 with eigenvalue λ_0 . We have

$$\lambda \geq \langle v_0, Xv_0 \rangle = \langle v_0, X_0v_0 \rangle + \langle v_0, (X - X_0)v_0 \rangle = \lambda_0 + \sum_{i=1}^N (v_0)_i^2 X_{ii},$$

where $(v_0)_i$ is the i -th coordinate of v_0 . We observe that v_0 is independent of (X_{ii}) . With $(x)_+^2 = \max(x, 0)^2$, we deduce that

$$\mathbb{E}(\lambda_0 - \lambda)_+^2 \leq \mathbb{E} \sum_{i=1}^N (v_0)_i^4 \mathbb{E} X_{ii}^2 \leq \mathbb{E} \|v_0\|_\infty^2 \sigma_0^2.$$

We write, $\mathbb{E} \|v_0\|_\infty^2 \leq (\log N)^{2C}/N + \mathbb{P}(\|v_0\|_\infty \geq (\log N)^C/\sqrt{N})$. From Lemma 5 applied to $c_0 = 2$, we deduce that for some constant $C > 0$,

$$\mathbb{E}(\lambda_0 - \lambda)_+^2 \leq \frac{(\log N)^C}{N}.$$

It implies the easy half of (4.7) for any $\kappa < 1$.

The proof of the converse inequality is more involved. For ease of notation, we introduce the number for $N \geq 3$,

$$L = L_N = (\log N)^{\log \log N}. \quad (4.8)$$

We say that a sequence of events (A_N) holds *with overwhelming probability* if for any $C > 0$, there exists a constant $c > 0$ such that $\mathbb{P}(A_N) \geq 1 - cN^{-C}$. We repeatedly use the fact that a polynomial union of events of overwhelming probability is an event of overwhelming probability. We start with a small deviation lemma which can be found, for example, in [8, Appendix B].

Lemma 8. *Assume that (Z_i) $1 \leq i \leq N$ are independent centered complex variables such that for some $\delta > 0$, for all i , $\mathbb{E} \exp |Z_i|^\delta \leq 1/\delta$. Then, for any $(x_i) \in \mathbb{C}^N$ with overwhelming probability,*

$$\left| \sum_{i=1}^N x_i Z_i \right| \leq L \|x\|_2.$$

For $z = E + i\eta$ with $\eta > 0$ and $E \in \mathbb{R}$, we introduce the resolvent matrices

$$R(z) = (X - zI)^{-1} \text{ and } R_0(z) = (X_0 - zI)^{-1},$$

where I denotes the identity matrix. The following lemma asserts that the resolvent can be used to estimate the largest eigenvalue of X and X_0 .

Lemma 9. *Let X be a Wigner matrix as in Theorem 1 and let $\lambda_1 \geq \dots \geq \lambda_N$ be its eigenvalues. For any $1 \leq k \leq N$, there exists an integer $1 \leq i \leq N$ such that for all E and $\eta > 0$*

$$\frac{1}{2} \max(\eta, |\lambda_k - E|)^{-2} \leq N \eta^{-1} \operatorname{Im} R(E + i\eta)_{ii}.$$

Moreover, let $1 \leq k \leq L$. There exists $c_0 > 0$ such that with overwhelming probability, we have $|\lambda_k - 2\sqrt{N}| < L^{c_0}N^{-1/6}$ and for all integers $1 \leq i \leq N$, and all E such that $|E - 2\sqrt{N}| < L^{c_0}N^{-1/6}$,

$$N\eta^{-1} \operatorname{Im} R(E + i\eta)_{ii} \leq L^{c_0} \min_{1 \leq j \leq N} (\lambda_j - E)^{-2}.$$

Proof. From the spectral theorem, we have

$$\operatorname{Im} R_{ii} = \sum_{p=1}^N \frac{\eta(v_p)_i^2}{(\lambda_p - E)^2 + \eta^2},$$

where (v_1, \dots, v_N) is an orthonormal basis of eigenvectors of X and $(v_p)_i$ is the i -th coordinate of v_p . In particular,

$$N\eta^{-1} \operatorname{Im} R_{ii} \geq \frac{N(v_1)_i^2}{(\lambda_k - E)^2 + \eta^2} \geq \frac{N(v_1)_i^2}{2 \max(\eta, |\lambda_k - E|)^2}.$$

From the pigeonhole principle, for some i , $(v_1)_i^2 \geq 1/N$ and the first statement of the lemma follows.

Fix an integer $1 \leq k \leq L$. From [9, Theorem 2.2] and Lemma 5, for some constants $c_0, C_0 > 0$, we have, with overwhelming probability, that the following event \mathcal{E} holds: $|\lambda_k - 2\sqrt{N}| \leq L^{c_0}N^{-1/6}$, for all integers $1 \leq p \leq N$,

$$\lambda_p \leq 2\sqrt{N} - 2C_0p^{2/3}N^{-1/6} + L^c p^{-1/3}N^{-1/6},$$

and $\|v_p\|_\infty^2 \leq L/N$. We set $q = \lfloor CL^{3c_0} \rfloor$ for some C . Let E be such that $|E - 2\sqrt{N}| \leq L^{c_0}N^{-1/6}$. On the event \mathcal{E} , if C is large enough, we have, for all $p > q$, $E - \lambda_p \geq C_0p^{2/3}N^{-1/6}$ and

$$\sum_{p=q+1}^N \frac{N(v_p)_i^2}{(\lambda_p - E)^2 + \eta^2} \leq \sum_{p=q+1}^N \frac{L}{(\lambda_p - E)^2} \leq \frac{1}{C_0^2} \sum_{p=q+1}^N \frac{LN^{1/3}}{p^{4/3}} \leq c_1 LN^{1/3} q^{-1/3}.$$

On the other hand, on the same event \mathcal{E} , we have

$$\sum_{p=1}^q \frac{N(v_p)_i^2}{(\lambda_p - E)^2 + \eta^2} \leq \sum_{p=1}^q \frac{N(v_p)_i^2}{\min_{1 \leq j \leq N} (\lambda_j - E)^2} \leq \frac{Lq}{\min_{1 \leq j \leq N} (\lambda_j - E)^2}.$$

It remains to adjust the value of the constant c_0 to conclude the proof. ■

The next step in the proof of (4.7) is a comparison between the resolvent of X and X_0 for z close to $2\sqrt{N}$. The following result is a corollary of [9, Theorem 2.1 (ii)].

Lemma 10. *Let X be a Wigner matrix as in Theorem 1. There exists $c > 0$ such that, with overwhelming probability, the following event holds: for all $z = E + i\eta$ such that $|2\sqrt{N} - E| \leq \sqrt{N}$ and $N^{-1/2}L^c \leq \eta \leq N^{1/2}$, all $i \neq j$, we have*

$$|R(z)_{ij}| \leq \Delta \quad \text{and} \quad |R(z)_{ii}| \leq cN^{-1/2},$$

where $\Delta = L^c(|E - 2\sqrt{N}| + \eta)^{1/4}N^{-7/8}\eta^{-1/2} + L^cN^{-2}\eta^{-1}$.

Proof. Let $Y = X/\sqrt{N}$ and for $z \in \mathbb{C}$, $\text{Im}(z) > 0$, $G(z) = (Y - zI)^{-1}$. We have $R(z) = N^{-1/2}G(zN^{-1/2})$. Theorem 2.1 (ii) in [9] asserts that with overwhelming probability for all $w = a + ib$ such that $|a| \leq 5$ and $N^{-1}L^c \leq b \leq 1$, all $i \neq j$, we have

$$|G(w)_{ij}| \leq \delta \quad \text{and} \quad |G(w)_{ii} - m(w)| \leq \delta,$$

where $\delta = L^c\sqrt{\text{Im}(m(w))/(Nb)} + L^c(Nb)^{-1}$ and $m(w)$ is the Cauchy-Stieltjes transform of the semi-circular law (for its precise definition see [9]). Then [9, Lemma 3.4] implies that, for some $C > 0$, for all $w = a + ib$, $|a| \leq 5$ and $0 \leq b \leq 1$, we have $|m(w)| \leq C$ and $|\text{Im}(m(w))| \leq C\sqrt{|a-2|+b}$. We apply the above result for $a = E/\sqrt{N}$ and $b = \eta/\sqrt{N}$. We obtain the claimed statement for $R(z) = N^{-1/2}G(zN^{-1/2})$. \blacksquare

We use Lemma 10 to estimate the difference between $R(z)$ and $R_0(z)$.

Lemma 11. *Let X be a Wigner matrix as in Theorem 1, let X_0 be obtained from X by setting to 0 all diagonal entries, and let c_0 be as in Lemma 9. With overwhelming probability, the following event holds: for all $z = E + i\eta$ such that $|2\sqrt{N} - E| \leq L^{c_0}N^{-1/6}$ and $\eta = N^{-1/4}$, all i ,*

$$|R_0(z)_{ii} - R(z)_{ii}| \leq \frac{1}{4N\eta}.$$

Proof. The resolvent identity states that if $A - zI$ and $B - zI$ are invertible matrices then

$$(A - zI)^{-1} = (B - zI)^{-1} + (B - zI)^{-1}(B - A)(A - zI)^{-1}. \quad (4.9)$$

Applying twice this identity, it implies that

$$R = R_0 + R_0(X_0 - X)R_0 + R_0(X_0 - X)R_0(X_0 - X)R$$

(where we omit to write the parameter z for ease of notation). For any integer $1 \leq i \leq N$, we thus have

$$R_{ii} - (R_0)_{ii} = - \sum_j (R_0)_{ij}X_{jj}(R_0)_{ji} + \sum_{j,k} (R_0)_{ij}X_{jj}(R_0)_{jk}X_{kk}R_{ki} = -I(z) + J(z).$$

Note that X_{jj} is independent of R_0 . By Lemma 8 and Lemma 10 we find that, with overwhelming probability,

$$|I(z)| \leq L\Delta^2\sqrt{N} + cLN^{-1}.$$

For a given $z = E + i\eta$ such that $|E - 2\sqrt{N}| \leq L^{c_0} N^{-1/6}$ and $\eta = N^{-1/4}$, it is straightforward to check that, for some $c > 0$, $\Delta = L^c N^{-19/24}$ and $|I(z)| \leq L^c N^{-13/12} = o(1/(N\eta))$.

Similarly, we have

$$|J(z)| \leq \sum_k |X_{kk}| |R_{ki}| |G_k| \quad \text{with} \quad G_k = \sum_j (R_0)_{ij} X_{jj} (R_0)_{jk}.$$

For a given z , by Lemma 8 and Lemma 10, we have with overwhelming probability, for all k , $|G_k| \leq L^c N^{-13/12}$ and $|J(z)| \leq L(\Delta N + cN^{-1/2}) L^c N^{-13/12} = o(1/(N\eta))$.

For a given z , let \mathcal{E}_z be the event that $\max_{1 \leq i \leq N} |R(z)_{ii} - R_0(z)_{ii}| \leq (8N\eta)^{-1}$ and \mathcal{E}'_z the event that $\max_{1 \leq i \leq N} |R(z)_{ii} - R_0(z)_{ii}| \leq (4N\eta)^{-1}$. We have proved so far that for a given $z = E + i\eta$ such that $|E - 2\sqrt{N}| \leq L^{c_0} N^{-1/6}$ and $\eta = N^{-1/4}$, with overwhelming probability, \mathcal{E}_z holds. By a net argument, it implies that with overwhelming probability, the events \mathcal{E}'_z hold jointly for all $z = E + i\eta$ with $|E - 2\sqrt{N}| \leq L^{c_0} N^{-1/6}$ and $\eta = N^{-1/4}$. Indeed, from the resolvent identity (4.9), we have $|R_{ij}(E + i\eta) - R_{ij}(E' + i\eta)| \leq \eta^{-2} |E - E'|$. It follows that if $|E - E'| \leq \eta^2 (8N\eta)^{-1} \leq N^{-1}$ then $|R(z)_{ii} - R_0(z)_{ii}| \leq (8N\eta)^{-1}$. Let \mathcal{N} be a finite subset of the interval $K = \{E : |E - 2\sqrt{N}| \leq L^{c_0} N^{-1/6}\}$ such that for all $E \in K$, $\min_{E' \in \mathcal{N}} |E - E'| \leq N^{-1}$. We may assume that \mathcal{N} has at most N elements. From what precedes we have the inclusion, with $\eta = N^{-1/4}$,

$$\bigcap_{z=E+i\eta; E \in \mathcal{N}} \mathcal{E}_z \subseteq \bigcap_{z=E+i\eta; E \in K} \mathcal{E}'_z.$$

From the union bound, the right-hand side holds with overwhelming probability. It concludes the proof of the lemma. \blacksquare

Now we have all ingredients necessary to conclude the proof of (4.7). Let $\eta = N^{-1/4}$. We prove that for some $c > 0$, with overwhelming probability,

$$\lambda \leq \lambda_0 + L^c \eta.$$

By Lemma 9, with overwhelming probability, $|\lambda - 2\sqrt{N}| \leq L^{c_0} N^{-1/6}$ and for some j ,

$$N\eta^{-1} \operatorname{Im} R(\lambda + i\eta)_{jj} \geq \frac{1}{2} \eta^{-2}.$$

and if $\lambda > \lambda_0$,

$$N\eta^{-1} \operatorname{Im} R_0(\lambda + i\eta)_{jj} \leq L^{c_0} (\lambda - \lambda_0)^{-2}.$$

By Lemma 11, we deduce that with overwhelming probability, if $\lambda > \lambda_0$,

$$\frac{1}{4} \eta^{-2} \leq L^{c_0} (\lambda - \lambda_0)^{-2}.$$

Hence, $\lambda \leq \lambda_0 + 2L^{c_0/2} \eta$, concluding the proof of (4.7).

4.3 Proof of Lemma 6

Let $\lambda = \lambda_1 \geq \dots \geq \lambda_N$ be the eigenvalues of X . For any (i, j) , let $\lambda^{(ij)}$ be the largest eigenvalue of $Y^{(ij)}$. We start by proving that λ and $\lambda^{(ij)}$ are close compared to their fluctuations. We have

$$\lambda \geq \langle u^{(ij)}, Xu^{(ij)} \rangle = \lambda^{(ij)} + \langle u^{(ij)}, (X - Y^{(ij)})u^{(ij)} \rangle \geq \lambda^{(ij)} - 2(|X_{ij}| + |X'_{ij}|) \|u^{(ij)}\|_\infty^2.$$

Since X and $Y^{(ij)}$ have the same distribution, we deduce from Lemma 5 that, with overwhelming probability, $\|v\|_\infty \leq L/\sqrt{N}$, $\max_{ij} \|u^{(ij)}\|_\infty \leq L/\sqrt{N}$ and $\max_{ij} (|X_{ij}| + |X'_{ij}|) \leq L/2$, where L is defined in (4.8). On this event, we get

$$\lambda \geq \lambda^{(ij)} - \frac{L^3}{N}.$$

Reversing the role X and $Y^{(ij)}$ and using the union bound, we deduce that with overwhelming probability,

$$\max_{ij} |\lambda - \lambda^{(ij)}| \leq \frac{L^3}{N}.$$

It follows from [23, Theorem 1.14] that, for any $\rho > 0$, there exists $\kappa > 0$ such that, for all N large enough,

$$\mathbb{P}(\lambda_2 < \lambda - N^{-1/2-\rho}) \geq 1 - N^{-\kappa}.$$

Let (v_1, \dots, v_p) be an orthonormal basis of eigenvectors of X associated to the eigenvalues $(\lambda_1, \dots, \lambda_N)$ with $v_1 = v$. We set $\theta = 2/5 - 3\rho/5$ and $q = \lfloor N^\theta \rfloor$. For some constant $c > 0$ to be defined and $\rho \in (0, 1/16)$, we introduce the event \mathcal{E}_ρ such that

- $\lambda_2 < \lambda - N^{-1/2-\rho}$ and $\lambda_q \leq \lambda - cq^{2/3}N^{-1/6}$;
- $\max_{1 \leq p \leq q} \|v_p\|_\infty \leq L/\sqrt{N}$ and $\max_{ij} \|u^{(ij)}\|_\infty \leq L/\sqrt{N}$;
- $\max_{ij} (|X_{ij}| + |X'_{ij}|) \leq L/2$.

From what precedes, Lemma 5 and [9, Theorem 2.2], for some c small enough, for any $\rho > 0$ there exists $\kappa > 0$ such that for all N large enough, $\mathbb{P}(\mathcal{E}_\rho) \geq 1 - N^{-\kappa}$. Note also, that we have checked that if \mathcal{E}_ρ holds then $\max_{ij} |\lambda - \lambda^{(ij)}| \leq L^3/N$.

On the event \mathcal{E}_ρ , we now prove that v and $u^{(ij)}$ are close in ℓ^∞ -norm. For a fixed (i, j) , we write, $u^{(ij)} = \alpha v + \beta x + \gamma y$, where $\alpha^2 + \beta^2 + \gamma^2 = 1$ with β, γ non-negative real numbers, x is a unit vector in the vector space spanned by (v_2, \dots, v_q) , and y is a unit vector in the vector space spanned by (v_{q+1}, \dots, v_N) . Set $w = (X - Y^{(ij)})u^{(ij)} + (\lambda - \lambda^{(ij)})u^{(ij)}$. We have

$$\lambda u^{(ij)} = \alpha \lambda v + \beta Xx + \gamma Xy + w.$$

Taking the scalar product with y , we find

$$\lambda\gamma = \lambda\langle y, u^{(ij)} \rangle = \gamma\langle y, Xy \rangle + \langle y, w \rangle \leq (\lambda - cq^{2/3}N^{-1/6})\gamma + \langle y, w \rangle.$$

Hence,

$$\gamma \leq c^{-1}q^{-2/3}N^{1/6}\|w\|_2 \leq c^{-1}q^{-2/3}N^{1/6}\left(\frac{L^2}{\sqrt{N}} + \frac{L^4}{N^{3/2}}\right) \leq 2c^{-1}L^4N^{-2\theta/3-1/3}.$$

Similarly, taking the scalar product with x , we find

$$\beta \leq N^{1/2+\rho}\langle x, w \rangle \leq N^{1/2+\rho}\left(\left|\langle x, (X - Y^{(ij)})u^{(ij)} \rangle\right| + \frac{L^3}{N}\right).$$

Since $|\langle a, b \rangle| \leq \|a\|_\infty\|b\|_1 \leq m\|a\|_\infty\|b\|_\infty$ where m is the number of non-zeros entries of b , we have $\left|\langle x, (X - Y^{(ij)})u^{(ij)} \rangle\right| \leq \|x\|_\infty L^2/\sqrt{N}$. By construction, $x = \sum_{p=2}^q \gamma_p v_p$ where $\sum_p |\gamma_p|^2 = 1$. If \mathcal{E}_ρ holds, we deduce that

$$\|x\|_\infty \leq \sum_{p=2}^q |\gamma_p| \|v_p\|_\infty \leq \frac{L}{\sqrt{N}} \sum_{p=2}^q |\gamma_p| \leq \frac{L\sqrt{q}}{\sqrt{N}} \leq LN^{\theta/2-1/2}.$$

So finally,

$$\beta \leq 2L^3N^{-1/2+\theta/2+\rho},$$

We deduce that $|\alpha| = \sqrt{1 - \beta^2 - \gamma^2} \geq 1 - \beta - \gamma$ is positive for all N large enough. We set $s = \alpha/|\alpha|$. We find, since $\|y\|_\infty \leq \|y\|_2 \leq 1$,

$$\|sv - u^{(ij)}\|_\infty \leq (1 - |\alpha|)\|v\|_\infty + \beta\|x\|_\infty + \gamma \leq 2\beta\|x\|_\infty + 2\gamma.$$

For our choice of $\theta = 2/5 - 3\rho/5$, this last expression is $O(L^4N^{-3/5+8\rho/5})$. Since $\rho < 1/16$, we have $3/5 - 8\rho/5 > 1/2$. Hence, finally, if we set $\beta = 1/10 - 8\rho/5 > 0$, we get that $\|sv - u^{(ij)}\|_\infty = O(L^4N^{-1/2-\beta})$. This concludes the proof of the lemma.

5 Proof of Theorem 2

The proof of Theorem 2 relies on a careful perturbation argument. Recall that $S_k = \{(i_1, j_1), \dots, (i_k, j_k)\}$ is the set of k pairs chosen uniformly at random (without replacement) from the set of all ordered pairs (i, j) of indices with $1 \leq i \leq j \leq N$ which is used in the definition of $X^{[k]}$. We denote by λ and $\lambda^{[k]}$ the largest eigenvalues of X and $X^{[k]}$. Recall the definition of $L = L_N$ in (4.8) and the notion of overwhelming probability immediately below (4.8). The main technical lemma is the following:

Lemma 12. *Let X be a Wigner matrix as in Theorem 2 and let $\lambda = \lambda_1 \geq \dots \geq \lambda_N$ be its eigenvalues. For any $c > 0$ there exists a constant $c_2 > 0$ such that for all $\varepsilon > 0$, for all N large enough, with probability at least $1 - \varepsilon$,*

$$\max_{k \leq N^{5/3}L^{-c_2}} \max_{p \in \{1,2\}} |\lambda_p - \lambda_p^{[k]}| \leq N^{-1/6}L^{-c}.$$

We postpone the proof of Lemma 12 to the next subsection. We denote by $R(z) = (X - zI)^{-1}$ and $R^{[k]}(z) = (X^{[k]} - zI)^{-1}$ the resolvent of X and $X^{[k]}$. The proof of Lemma 12 is based on this comparison lemma on the resolvents.

Lemma 13. *Let X be a Wigner matrix as in Theorem 1. Let $c_0 > 0$ be as in Lemma 9 and let $c_1 > 0$. There exists $c_2 > 0$ such that, with overwhelming probability, the following event holds: for all $k \leq N^{5/3}L^{-c_2}$, for all $z = E + i\eta$ such that $|2\sqrt{N} - E| \leq L^{c_0}N^{-1/6}$ and $\eta = N^{-1/6}L^{-c_1}$,*

$$\max_{1 \leq i, j \leq N} N\eta |R^{[k]}(z)_{ij} - R(z)_{ij}| \leq \frac{1}{L^2}.$$

We postpone the proof of Lemma 12 to the next subsection. Our next lemma connects the resolvent with eigenvectors.

Lemma 14. *Let X be a Wigner matrix as in Theorem 1 and let $\varepsilon > 0$. There exist c_1, c_2 such that the following event holds for all N large enough with probability at least $1 - \varepsilon$: for all $k \leq N^{5/3}L^{-c_2}$, we have, with $z = \lambda + i\eta$, $\eta = N^{-1/6}L^{-c_1}$,*

$$\max_{1 \leq i, j \leq N} |N\eta \operatorname{Im} R(z)_{ij} - Nv_i v_j| \leq \frac{1}{L^2} \quad \text{and} \quad \max_{1 \leq i, j \leq N} |N\eta \operatorname{Im} R^{[k]}(z)_{ij} - Nv_i^{[k]} v_j^{[k]}| \leq \frac{1}{L^2}.$$

Proof. Let $\lambda = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ be the eigenvalues of X . Let (v_1, \dots, v_N) be an eigenvector basis of X . Recall that

$$N\eta \operatorname{Im} R(E + i\eta)_{ij} = \sum_{p=1}^N \eta^2 \frac{N(v_p)_i (v_p)_j}{(\lambda_p - E)^2 + \eta^2}.$$

As in the proof of Lemma 9, from [9, Theorem 2.2] and Lemma 5, for some constants $c_0, C > 0$, we have with overwhelming probability that the following event \mathcal{E} holds: $|\lambda - 2\sqrt{N}| \leq L^{c_0}N^{-1/6}$, for all integers $1 \leq p \leq N$, $\|v_p\|_\infty^2 \leq L/N$ and for all $q > p$ with $q = \lfloor L^{c_0} \rfloor$ and E such that $|E - 2\sqrt{N}| \leq L^{c_0}N^{-1/6}$ we have

$$\sum_{p=q+1}^N \frac{N(v_p)_i (v_p)_j}{(\lambda_p - E)^2 + \eta^2} \leq CLN^{1/3}q^{-1/3}.$$

On the other hand, let \mathcal{E}_δ be the event that $\lambda_2 \leq \lambda - \delta N^{-1/6}$. Fix $\varepsilon > 0$. From [3, Theorem 2.7] and, e.g., [1, Chapter 3], there exists $\delta > 0$ such that

$$\mathbb{P}(\mathcal{E}_\delta) \geq 1 - \varepsilon.$$

On the event $\mathcal{E} \cap \mathcal{E}_\delta$, if $|\lambda - E| \leq (\delta/2)N^{-1/6}$, we have

$$\sum_{p=2}^q \frac{N(v_p)_i(v_p)_j}{(\lambda_p - E)^2 + \eta^2} \leq \frac{4}{\delta^2} LqN^{1/3}.$$

Finally, if $|\lambda - E| \leq \eta/L^2$, on the event \mathcal{E} , we find easily, if v_i is i -th coordinate of v ,

$$\left| \eta^2 \frac{Nv_i v_j}{(\lambda - E)^2 + \eta^2} - Nv_i v_j \right| \leq \frac{1}{L^3}.$$

For some $c_1 > 0$, we thus find, that if $\eta = N^{-1/6}L^{-c_1}$ then on the event $\mathcal{E} \cap \mathcal{E}_\delta$, for all E such that $|\lambda - E| \leq \eta/L^2$ we have

$$\max_{i,j} |N\eta \operatorname{Im} R(E + i\eta)_{ij} - Nv_i v_j| \leq \frac{1}{L^2}.$$

We apply this last estimate R and $E = \lambda$. For each k , let $\mathcal{E}^{[k]}$ be the event corresponding to \mathcal{E} for $X^{[k]}$ instead of X . We apply the above estimate on the event $\mathcal{E}'_k = \mathcal{E}^{[k]} \cap \mathcal{E}_\delta \cap \{\max_{p=1,2} |\lambda_p - \lambda_p^{[k]}| \leq \eta/L^2\}$ to $R^{[k]}$ and $E = \lambda$. By Lemma 12 and the union bound $\bigcap_{k \leq N^{5/3}L^{-c_2}} \mathcal{E}'_k$ has probability at least $1 - 2\varepsilon$. It concludes the proof. ■

We may now conclude the proof of Theorem 2. Let c_1, c_2 be as in Lemma 14, $k \leq N^{5/3}L^{-c_2}$ and $\eta = N^{-1/6}L^{-c_1}$. Up to increasing the value of c_2 , we may also assume that the conclusion of Lemma 13 holds. By Lemma 5, Lemma 13 and Lemma 14, for any $\varepsilon > 0$, for all N large enough, with probability at least $1 - \varepsilon$, it holds that for some $c > 0$: $\sqrt{N}\|v\|_\infty \leq (\log N)^c$, $\sqrt{N}\|v^{[k]}\|_\infty \leq (\log N)^c$ and

$$\max_{i,j} |Nv_i v_j - Nv_i^{[k]} v_j^{[k]}| \leq \frac{3}{L^2}.$$

Applied to $i = j$, we get that for some $s_i \in \{-1, 1\}$,

$$\sqrt{N}|s_i v_i - v_i^{[k]}| \leq \frac{\sqrt{3}}{L}.$$

Notably, we find

$$(1 - s_i s_j)N|v_i v_j| \leq |Nv_i v_j - Nv_i^{[k]} v_j^{[k]}| + \frac{2\sqrt{3}}{L}(\log N)^c \leq \frac{4}{L}(\log N)^c.$$

Let $J = \{1 \leq i \leq N : \sqrt{N}|v_i| \geq L^{-1/3}\}$. It follows from the above inequality that for $i, j \in J$, $s_i = s_j$. Let s be this common value. We have for all $i \in J$,

$$\sqrt{N}|sv_i - v_i^{[k]}| \leq \frac{\sqrt{3}}{L}.$$

Moreover, for all $i \notin J$, by definition,

$$\sqrt{N}|sv_i - v_i^{[k]}| \leq \sqrt{N}|v_i| + \sqrt{N}|v_i^{[k]}| \leq L^{-1/3} + L^{-1/3} + \sqrt{3}L^{-1}.$$

It concludes the proof of Theorem 2.

5.1 Proof of Lemma 13

For technical convenience, we start by truncating our random variables (X_{ij}) . Set $\tilde{X}_{ij} = X_{ij}\mathbb{1}(|X_{ij}| \leq (\log N)^c)$ and $\tilde{X}'_{ij} = X'_{ij}\mathbb{1}(|X'_{ij}| \leq (\log N)^c)$ with $c = 2/\delta$. The matrix \tilde{X} has independent entries above the diagonal. Moreover, since $\mathbb{E}\exp(|X_{ij}|^\delta) \leq 1/\delta$, with overwhelming probability, $X = \tilde{X}$ and $X' = \tilde{X}'$. It is also straightforward to check that $\mathbb{E}|X_{ij}|^2\mathbb{1}(|X_{ij}| \geq (\log N)^c) = O(\exp(-(\log N)^2/2))$. It implies that $|\mathbb{E}\tilde{X}_{ij}| = O(\exp(-(\log N)^2/2))$ and $\text{Var}(\tilde{X}_{ij}) = 1 + O(\exp(-(\log N)^2/2))$ for $i \neq j$. We define the matrix \bar{X} with for $i \neq j$,

$$\bar{X}_{ij} = (\tilde{X}_{ij} - \mathbb{E}\tilde{X}_{ij})/\sqrt{\text{Var}(\tilde{X}_{ij})} \quad \text{and} \quad \bar{X}_{ii} = \tilde{X}_{ij} - \mathbb{E}\tilde{X}_{ij}.$$

The matrix \bar{X} is a Wigner matrix as in Theorem 2 with entries in $[-L/4, L/4]$. Moreover, from Gershgorin's circle theorem [15, Theorem 6.6.1], with overwhelming probability, the operator norm of $X - \bar{X}$ satisfies $\|X - \bar{X}\| = O(N \exp(-(\log N)^2/2))$. Observe that from the spectral theorem, for any Hermitian matrix A , $\|(A - z)^{-1}\| \leq |\text{Im}(z)|^{-1}$. In particular, from the resolvent identity (4.9), we get $\|(X - z)^{-1} - (\bar{X} - z)^{-1}\| = \|(X - z)^{-1}(X - \bar{X})(\bar{X} - z)^{-1}\| \leq \text{Im}(z)^{-2}\|X - \bar{X}\| = O(N^3 \exp(-(\log N)^2/2))$ if $\text{Im}(z) \geq N^{-1}$. The same truncation procedure applies for $X^{[k]}$. In the proof of Lemma 13, we may thus assume without loss of generality that the random variables X_{ij} have support in $[-L/4, L/4]$.

The proof of Lemma 13 is based on a martingale argument. Thanks to the resolvent identity (4.9), we will write $R_{ij}^{[k]}(z) - R_{ij}(z)$ as a sum of martingale differences up to small error terms. For $0 \leq t \leq k$, let \mathcal{F}_t be the σ -algebra generated by the random variable X , S_k and $(X'_{i_s, j_s})_{1 \leq s \leq t}$. For $1 \leq i, j \leq N$, we set

$$T_{ij} = \{t : \{i_t, j_t\} \cap \{i, j\} \neq \emptyset\}.$$

Note that T_{ij} is \mathcal{F}_0 -measurable. We have

$$\mathbb{E}|T_{ij}| = \frac{2k}{N+1}.$$

Besides, from [6, Proposition 1.1], for any $u > 0$,

$$\mathbb{P}\left(|T_{ij}| \geq \mathbb{E}|T_{ij}| + u\right) \leq \exp\left(-\frac{u^2}{4\mathbb{E}|T_{ij}| + 2u}\right).$$

If $k \leq N^{5/3}L^{-c_2}$, it follows that with overwhelming probability, the following event, say \mathcal{T} , holds: $\max_{ij}|T_{ij}| \leq 4k'/N$ where for ease of notation we have set

$$k' = \min(k, N(\log N)^2).$$

Now, let c be as in Lemma 10 and, for $0 \leq t \leq k$, we denote by $\mathcal{E}_t \in \mathcal{F}_t$ the event that \mathcal{T} holds and that the conclusion of Lemma 10 holds for $X^{[t]}$ and $R^{[t]}$ (with the convention $X^{[0]} = X$). If \mathcal{E}_t holds, then for all $z = E + i\eta$ with $|2\sqrt{N} - E| \leq L^{c_0}N^{-1/6}$ and $\eta = N^{-1/6}L^{-c_1}$, we have,

$$\max_{i \neq j} |R_{ij}^{[t]}(z)| \leq \delta = L^{c'}N^{-5/6} \quad \text{and} \quad \max_i |R_{ii}^{[t]}(z)| \leq \delta_0 = cN^{-1/2},$$

where $c' = 1 + c + \max(c_0/2, c_0/4 + c_1/2)$. We define $X_0^{[t]}$ as the symmetric matrix obtained from $X^{[t]}$ by setting to 0 the entries (i_t, j_t) and (j_t, i_t) . By construction $X_0^{[t+1]}$ is \mathcal{F}_t -measurable. We denote by $R_0^{[t]}$ the resolvent of $X_0^{[t]}$. The resolvent identity (4.9) implies that

$$R_0^{[t+1]} = R^{[t]} + R^{[t]}(X^{[t]} - X_0^{[t+1]})R^{[t]} + R^{[t]}(X^{[t]} - X_0^{[t+1]})R^{[t]}(X^{[t]} - X_0^{[t+1]})R_0^{[t+1]}$$

(we omit to write the parameter z for ease of notation). Now, we set for $i \neq j$, $E_{ij}^s = e_i e_j^* + e_j e_i^*$ and $E_{ij}^s = e_i e_i^*$, where e_i denotes the canonical vector of \mathbb{R}^n with all entries equal to 0 except the i -th entry equal to 1. We have

$$X^{[t]} - X_0^{[t+1]} = X_{i_{t+1}j_{t+1}} E_{i_{t+1}j_{t+1}}^s \quad \text{and} \quad X^{[t+1]} - X_0^{[t+1]} = X'_{i_{t+1}j_{t+1}} E_{i_{t+1}j_{t+1}}^s. \quad (5.1)$$

We use that $|X_{ij}| \leq L/4$ and $(R_0^{[t+1]})_{ij} \leq \eta^{-1}$. If \mathcal{E}_t holds, we deduce that for all $z = E + i\eta$ with $|2\sqrt{N} - E| \leq L^{c_0}N^{-1/6}$ and $\eta = N^{-1/6}L^{-c_1}$, we have

$$\max_{i \neq j} |(R_0^{[t+1]})_{ij}| \leq \sqrt{2}\delta \quad \text{and} \quad \max_i |(R_0^{[t+1]})_{ii}| \leq \sqrt{2}\delta_0. \quad (5.2)$$

Similarly, the resolvent identity (4.9) with $R^{[t+1]}$ and $R^{[t]}$ implies that, if \mathcal{E}_t holds, for all $z = E + i\eta$ with $|2\sqrt{N} - E| \leq L^{c_0}N^{-1/6}$ and $\eta = N^{-1/6}L^{-c_1}$, we have

$$\max_{i \neq j} |R_{ij}^{[t+1]}| \leq \sqrt{2}\delta \quad \text{and} \quad \max_i |R_{ii}^{[t+1]}| \leq \sqrt{2}\delta_0. \quad (5.3)$$

Finally, the resolvent identity with $R^{[t+1]}$ and $R_0^{[t+1]}$ gives

$$\begin{aligned} R^{[t+1]} &= R_0^{[t+1]} + R_0^{[t+1]}(X_0^{[t+1]} - X^{[t+1]})R^{[t+1]} \\ &= \sum_{\ell=0}^2 \left(R_0^{[t+1]}(X_0^{[t+1]} - X^{[t+1]}) \right)^\ell R_0^{[t+1]} + \left(R_0^{[t+1]}(X_0^{[t+1]} - X^{[t+1]}) \right)^3 R^{[t+1]}. \end{aligned}$$

Note that, $\mathbb{E}[X'_{i_{t+1}j_{t+1}} | \mathcal{F}_t] = 0$. We use $|X_{i_tj_t}| \leq L/4$, from (5.1)-(5.2)-(5.3), we deduce that

$$\left| \mathbb{E}[R_{ij}^{[t+1]} | \mathcal{F}_t] - (R_0^{[t+1]})_{ij} - s_{ij}^{[t+1]} X_{i_{t+1}j_{t+1}}'^2 \right| \leq a_t \quad \text{and} \quad |R_{ij}^{[t+1]} - (R_0^{[t+1]})_{ij}| \leq b_t, \quad (5.4)$$

where $s_{ij}^{[t]} = ((R_0^{[t]} E_{i_tj_t}^s)^2 R_0^{[t]})_{ij}$ and, if \mathcal{E}_t holds,

$$a_t = L^3 \delta^2 \delta_0^2 + L^3 \delta_0^4 \mathbb{1}_{(t \in T_{ij})} \quad \text{and} \quad b_t = L \delta^2 + L \delta \delta_0 \mathbb{1}_{(t \in T_{ij})} + L \delta_0^2 \mathbb{1}_{((i_t, j_t) = (i, j))}.$$

We rewrite, one last time the resolvent identity with $R_0^{[t+1]}$ and $R^{[t]}$:

$$R^{[t]} = \sum_{\ell=0}^2 \left(R_0^{[t+1]}(X_0^{[t+1]} - X^{[t]}) \right)^\ell R_0^{[t+1]} + \left(R_0^{[t+1]}(X_0^{[t+1]} - X^{[t]}) \right)^3 R^{[t]}.$$

If \mathcal{E}_t holds, we arrive at,

$$\left| \mathbb{E}[R_{ij}^{[t+1]} | \mathcal{F}_t] - (R^{[t]})_{ij} - r_{ij}^{[t+1]} X_{i_{t+1}j_{t+1}} + s_{ij}^{[t+1]} (X_{i_{t+1}j_{t+1}}^2 - X_{i_{t+1}j_{t+1}}'^2) \right| \leq 2a_t,$$

where $r_{ij}^{[t]} = (R_0^{[t]} E_{i_tj_t}^s R_0^{[t]})_{ij}$. We have thus found that

$$R_{ij}^{[k]} - R_{ij} = \sum_{t=0}^{k-1} \left(R_{ij}^{[t+1]} - (R^{[t]})_{ij} \right) = \sum_{t=0}^{k-1} \left(R_{ij}^{[t+1]} - \mathbb{E}[R_{ij}^{[t+1]} | \mathcal{F}_t] \right) + r_{ij} + s_{ij} - s'_{ij} + a_{ij}, \quad (5.5)$$

where we have set, with $Y_{ij} = X_{ij}^2 - \mathbb{E}X_{ij}^2$, $Y'_{ij} = X_{ij}'^2 - \mathbb{E}X_{ij}'^2$,

$$r_{ij} = \sum_{t=1}^k r_{ij}^{[t]} X_{i_tj_t}, \quad s_{ij} = \sum_{t=1}^k s_{ij}^{[t]} Y_{i_tj_t}, \quad s'_{ij} = \sum_{t=1}^k s_{ij}^{[t]} Y'_{i_tj_t}, \quad |a_{ij}| \leq 2 \sum_{t=1}^k a_t.$$

We may now use concentration inequalities to estimate the terms in (5.5). We set

$Z_{t+1} = (R_{ij}^{[t+1]} - \mathbb{E}[R_{ij}^{[t+1]} | \mathcal{F}_t]) \mathbb{1}_{\mathcal{E}_t}$. We write, for any $u \geq 0$,

$$\mathbb{P} \left(\left| \sum_{t=0}^{k-1} \left(R_{ij}^{[t+1]} - \mathbb{E}[R_{ij}^{[t+1]} | \mathcal{F}_t] \right) \right| \geq u \right) \leq \mathbb{P} \left(\left| \sum_{t=1}^k Z_t \right| \geq u \right) + \sum_{t=0}^{k-1} \mathbb{P}(\mathcal{E}_t^c).$$

By Lemma 10, we have for any $c > 0$, $\sum_{t=0}^{k-1} \mathbb{P}(\mathcal{E}_t^c) = O(N^{-c})$. Since $\mathcal{E}_t \in \mathcal{F}_t$, we have by $\mathbb{E}[Z_{t+1}|\mathcal{F}_t] = 0$. Also, from (5.2)-(5.4), $|Z_t| \leq 2b_t$. On the event \mathcal{T} , we have

$$\sqrt{\sum_{t=1}^k b_t^2} \leq L\delta^2\sqrt{k} + L\delta\delta_0\sqrt{\frac{4k'}{N}} + L\delta_0^2 \leq 2L\delta^2\sqrt{k'}.$$

Azuma-Hoeffding martingale inequality implies that, for $u \geq 0$,

$$\mathbb{P}\left(\left|\sum_{t=1}^k Z_t\right| \geq 2uL\delta^2\sqrt{k'}\right) \leq 2\exp\left(-\frac{u^2}{2}\right).$$

We apply the later inequality to $u = \log N$. We deduce that, with overwhelming probability,

$$\sum_{t=0}^{k-1} \left(R_{ij}^{[t+1]} - \mathbb{E}[R_{ij}^{[t+1]}|\mathcal{F}_t]\right) \leq L^2\sqrt{k'}\delta^2. \quad (5.6)$$

We may treat similarly the random variable s'_{ij} in (5.5). We set $Z'_{t+1} = s_{ij}^{[t+1]} Y'_{i_{t+1}j_{t+1}} \mathbb{1}_{\mathcal{E}_t}$. Note that $s_{ij}^{[t+1]}$ is \mathcal{F}_t -measurable and $\mathbb{E}[Y'_{i_{t+1}j_{t+1}}|\mathcal{F}_t] = 0$. Thus $\mathbb{E}[Z'_{t+1}|\mathcal{F}_t] = 0$. Moreover, since $|Y_{ij}| \leq L^2/16$, from (5.2), we find $|Z'_{t+1}| \leq b'_t = L^2(\delta^2\delta_0 + \delta_0^3 \mathbb{1}_{(t \in T_{ij})})$. If \mathcal{T} holds, we get

$$\sqrt{\sum_{t=0}^{k-1} b_t'^2} \leq L^2\delta^2\delta_0\sqrt{k} + L^2\delta_0^3\sqrt{\frac{4k'}{N}} \leq 2L^2\delta^2\delta_0\sqrt{k'}$$

We write, for $u \geq 0$,

$$\mathbb{P}\left(|s'_{ij}| \geq u\right) \leq \mathbb{P}\left(\left|\sum_{t=1}^k Z'_t\right| \geq u\right) + \sum_{t=0}^{k-1} \mathbb{P}(\mathcal{E}_t^c).$$

From Azuma-Hoeffding martingale inequality, we deduce that, with overwhelming probability,

$$|s'_{ij}| \leq \sqrt{k'}\delta^2. \quad (5.7)$$

We now estimate the random variable r_{ij} in (5.5). We will also use Azuma-Hoeffding inequality but we need to introduce a backward filtration. We define \mathcal{F}'_t as the σ -algebra generated by the random variables, X' , S_k and $\{(X_{ij}) : \{i, j\} \notin \{i_s, j_s\}, s \leq t\}$. By construction $X^{[t]}$ and $X_0^{[t]}$ are \mathcal{F}'_t -measurable random variables. Let $\mathcal{E}'_t \in \mathcal{F}'_t$ be the event that \mathcal{T} holds and that the conclusion of Lemma 10 holds for

$X^{[t]}$. If \mathcal{E}'_t holds, then for all $z = E + i\eta$ with $|2\sqrt{N} - E| \leq L^{c_0} N^{-1/6}$ and $\eta = N^{-1/6} L^{-c_1}$, we have,

$$\max_{i \neq j} |R_{ij}^{[t]}(z)| \leq \delta \quad \text{and} \quad \max_i |R_{ii}^{[t]}(z)| \leq \delta_0.$$

Arguing as in (5.2), if \mathcal{E}'_t holds then

$$\max_{i \neq j} |(R_0^{[t]})_{ij}| \leq \sqrt{2}\delta \quad \text{and} \quad \max_i |(R_0^{[t]})_{ii}| \leq \sqrt{2}\delta_0.$$

The variable $r_{ij}^{[t]}$ is \mathcal{F}'_t -measurable and $\mathbb{E}(X_{i,jt} | \mathcal{F}'_t) = 0$. We write, for $u \geq 0$,

$$\mathbb{P}(|r_{ij}| \geq u) \leq \mathbb{P}\left(\left|\sum_{t=0}^{k-1} \tilde{Z}_t\right| \geq u\right) + \sum_{t=0}^{k-1} \mathbb{P}(\mathcal{E}'_t^c),$$

where $\tilde{Z}_{t+1} = r_{ij}^{[t]} X_{i,jt} \mathbb{1}_{\mathcal{E}'_t}$. We have $\mathbb{E}(\tilde{Z}_{t+1} | \mathcal{F}'_t) = 0$ and

$$|\tilde{Z}_t| \leq \tilde{b}_t = L\delta^2 + L\delta\delta_0 \mathbb{1}_{(t \in T_{ij})} + L\delta_0^2 \mathbb{1}_{(i,j,t) = (i,j)}.$$

Arguing as above, from Azuma-Hoeffding martingale inequality, we deduce that with overwhelming probability,

$$|r_{ij}| \leq L^2 \sqrt{k'} \delta^2. \quad (5.8)$$

Similarly, repeating the argument leading to (5.7) with s_{ij} and the filtration (\mathcal{F}'_t) gives with overwhelming probability,

$$|s_{ij}| \leq \sqrt{k'} \delta^2. \quad (5.9)$$

We note also that if \mathcal{T} holds then

$$|a_{ij}| \leq 2 \sum_{t=1}^k a_t \leq 2L^3 \delta^2 \delta_0^2 k + 2L^3 \delta_0^4 \frac{4k'}{N} \leq \sqrt{k'} \delta^2,$$

where the last inequality holds provided that $k \leq N^{5/3}$. So finally, from (5.5)-(5.6)-(5.7)-(5.8)-(5.9), we have proved that for a given $z = E + i\eta$ such that $|E - 2\sqrt{N}| \leq L^{c_0} N^{-1/6}$ and $\eta = N^{-1/6} L^{-c_1}$, with overwhelming probability

$$\left| R_{ij}^{[k]}(z) - R_{ij}(z) \right| \leq 3L^2 \sqrt{k'} \delta^2,$$

where the inequality holds provided that $k \leq N^{5/3}$. Recall that $|R_{ij}(E + i\eta) - R_{ij}(E' + i\eta)| \leq \eta^{-2} |E - E'|$. By a net argument (as in the proof of Lemma 11), we deduce that with overwhelming probability for all $z = E + i\eta$ such that $|2\sqrt{N} - E| \leq L^{c_0} N^{-1/6}$, $\left| R_{ij}^{[k]}(z) - R_{ij}(z) \right| \leq 4L^2 \sqrt{k'} \delta^2$. It concludes the proof of Lemma 13.

5.2 Proof of Lemma 12

Let c_0 be as in Lemma 9 and $c > 0$. We set $c_1 = c_0/2 + 2c$ and let $\eta = N^{-1/6}L^{-c_1}$. Let $p \in \{1, 2\}$. We start with by bounding $\min_j |\lambda_p - \lambda_j^{[k]}|$ and $\min_j |\lambda_p^{[k]} - \lambda_j|$. Since X and $X^{[k]}$ have the same distribution, we only prove that with overwhelming probability

$$\min_{1 \leq j \leq N} |\lambda_p - \lambda_j^{[k]}| \leq 2L^{c_0/2}\eta. \quad (5.10)$$

By Lemma 9, with overwhelming probability, $|\lambda_p - 2\sqrt{N}| \leq L^{c_0}N^{-1/6}$ and for some integer $1 \leq i \leq N$,

$$N\eta^{-1} \operatorname{Im}R(\lambda_p + i\eta)_{ii} \geq \frac{1}{2}\eta^{-2},$$

and,

$$N\eta^{-1} \operatorname{Im}R^{[k]}(\lambda_p + i\eta)_{ii} \leq L^{c_0} \min_{1 \leq j \leq N} (\lambda_p - \lambda_j^{[k]})^{-2}.$$

By Lemma 13, we deduce that if $k \leq N^{5/3}L^{-c_2}$, with overwhelming probability,

$$\frac{1}{4}\eta^{-2} \leq L^{c_0} \min_{1 \leq j \leq N} (\lambda_p - \lambda_j^{[k]})^{-2}.$$

It proves (5.10).

We may now conclude the proof of Lemma 12. Fix $\varepsilon > 0$. As already noticed, from [3, Theorem 2.7], there exists $\delta > 0$ such that, with probability at least $1 - \varepsilon$, $\lambda_2 < \lambda - \delta N^{-1/6}$. From what precedes, with probability at least $1 - 2\varepsilon$, \mathcal{E}_δ holds and for all $k \leq N^{5/3}L^{-c_2}$, we have

$$\max \left(\min_{1 \leq j \leq N} |\lambda_p^{[k]} - \lambda_j|, \min_{1 \leq j \leq N} |\lambda_p - \lambda_j^{[k]}| \right) \leq \alpha,$$

with $\alpha = 2L^{c_0/2}\eta$. On this event, we readily find $|\lambda - \lambda^{[k]}| \leq \alpha$ and for some p , $|\lambda_p - \lambda_2^{[k]}| \leq \alpha$. Assume that this last inequality is false for $p \neq 2$. Since $2\alpha < \delta N^{-1/6}$, if $p \neq 2$, then $p \leq 3$ and we deduce that $\lambda_2 > \lambda_2^{[k]} + \alpha$. We note that, on our event, for some q , we have $|\lambda_2 - \lambda_q^{[k]}| \leq \alpha$. In particular, $\lambda_q^{[k]} > \lambda_2^{[k]}$. So necessarily, $q = 1$ and, from the triangle inequality, $|\lambda_2 - \lambda_1| \leq 2\alpha$. This is a contradiction since $2\alpha < \delta N^{-1/6}$. It concludes the proof of Lemma 12.

References

- [1] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

-
- [2] I. Benjamini, G. Kalai, and O. Schramm. Noise sensitivity of Boolean functions and applications to percolation. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, 90(1):5–43, 1999.
- [3] Paul Bourgade, László Erdős, and Horng-Tzer Yau. Edge universality of beta ensembles. *Comm. Math. Phys.*, 332(1):261–353, 2014.
- [4] J. Bourgain, J. Kahn, G. Kalai, Y. Katznelson, and N. Linial. The influence of variables in product spaces. *Israel Journal of Mathematics*, 77(1-2):55–64, 1992.
- [5] Sourav Chatterjee. *Concentration inequalities with exchangeable pairs (Ph. D. thesis)*. PhD thesis, Stanford University, 2005.
- [6] Sourav Chatterjee. Stein's method for concentration inequalities. *Probab. Theory Related Fields*, 138(1-2):305–321, 2007.
- [7] Sourav Chatterjee. *Superconcentration and related topics*. Springer, 2016.
- [8] László Erdős, Horng-Tzer Yau, and Jun Yin. Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields*, 154(1-2):341–407, 2012.
- [9] László Erdős, Horng-Tzer Yau, and Jun Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.*, 229(3):1435–1515, 2012.
- [10] L. Erdős, B. Schlein, and HT. Yau. Local semicircle law and complete delocalization for wigner random matrices. *Communications in Mathematical Physics*, 287(2):641–655, 2009.
- [11] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American Mathematical Society*, 124:2993–3002, 1996.
- [12] C. Garban. Oded Schramm's contributions to noise sensitivity. *The Annals of Probability*, 39:1702–1767, 2011.
- [13] C. Garban, G. Pete, and O. Schramm. The Fourier spectrum of critical percolation. *Acta Mathematica*, 205:19–104, 2010.
- [14] C. Garban and J.E. Steif. *Noise sensitivity of Boolean functions and percolation*. Cambridge University Press, 2014.
- [15] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.
- [16] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *Proceedings of the 29th Annual Symposium on the Foundations of Computer Science*, pages 68–80. Computer Society Press, 1988.

-
- [17] G. Kalai and S. Safra. Threshold phenomena and influence: perspectives from mathematics, computer science, and economics. *Computational complexity and statistical physics*, pages 25–60, 2006.
- [18] Antti Knowles and Jun Yin. Eigenvector distribution of Wigner matrices. *Probab. Theory Related Fields*, 155(3-4):543–582, 2013.
- [19] Michel Ledoux and Brian Rider. Small deviations for beta ensembles. *Electron. J. Probab.*, 15:no. 41, 1319–1343, 2010.
- [20] R. O’Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.
- [21] Oded Schramm and Jeffrey E Steif. Quantitative noise sensitivity and exceptional times for percolation. *Annals of Mathematics*, 171:619672, 2010.
- [22] M. Talagrand. On russo’s approximate zero-one law. *Annals of Probability*, 22:1576–1587, 1994.
- [23] Terence Tao and Van Vu. Random matrices: universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.*, 298(2):549–572, 2010.