Largest eigenvalues and eigenvectors of band or sparse random matrices

Florent Benaych-Georges[*]   Sandrine Péché[†]

Abstract

In this text, we consider an random $N \times N$ matrix $\mathbf{X}$ such that all but $o(N)$ rows of $\mathbf{X}$ have W non identically zero entries, the other rows having less than W entries (such as, for example, standard or cyclic band matrices). We always suppose that $1 \ll W \ll N$. We first prove that if the entries are independent, centered, have variance one, satisfy a certain tail upper-bound condition and $W \gg (\log N)^{6(1+\alpha)}$, where α is a positive parameter depending on the distribution of the entries, then the largest eigenvalue of $\mathbf{X}/\sqrt{W}$ converges to the upper bound of its limit spectral distribution, that is 2, as for Wigner matrices. This extends some previous results by Khorunzhiy and Sodin where less hypotheses were made on W, but more hypotheses were made about the law of the entries and the structure of the matrix. Then, under the same hypotheses, we prove a delocalization result for the eigenvectors of $\mathbf{X}$. More precisely we show that eigenvectors associated to eigenvalues "far enough" from zero cannot be essentially localized on less than $W/\log(N)$ entries. This lower bound on the localization length has to be compared to the recent result by Steinerberger, which states that the localization length in the edge is $\ll W^{7/5}$ or there is strong interaction between two eigenvectors in an interval of length $W^{7/5}$.

Keywords: random matrices ; sparse matrices ; largest eigenvalues ; localization.
AMS MSC 2010: 15A52 ; 60F05.
Submitted to ECP on September 19, 2013, final version accepted on January 27, 2014.

Random band matrices (*i.e.* random Hermitian matrices with independent entries vanishing out of a band around the diagonal) have raised lots of attention recently. Indeed, varying the bandwidth W from 1 to the full size shows (in the large size limit) a crossover between a strongly disordered regime, with localized eigenfunctions and weak eigenvalue correlation, and a weakly disordered regime, with extended eigenfunctions and strong eigenvalue repulsion. It is conjectured (and explained on a Physics level of rigor by Fyodorov and Mirlin in [9]) that for Gaussian band matrices, the *localization strength* (*i.e.* the typical number of coordinates bearing most of the ℓ^2 mass) of a typical eigenvector in the bulk of the spectrum shall be of order $L \sim N \wedge W^2$, so that eigenvectors of the bulk should be localized (resp. extended) if $W \ll \sqrt{N}$ (resp. $\gg \sqrt{N}$). The only rigorous result in the direction of localization is by Schenker [13]. Therein it is proved that $L \ll W^8$ for Gaussian band matrices. On the other hand, delocalization in the bulk is proved by Erdös, Knowles, Yau and Yin [7] when $W \gg N^{4/5}$. In both regimes, it is known from Erdös and Knowles [5, 6] that typically (*i.e.* disregarding a negligible proportion of eigenvectors) $L \gg W^{7/6} \wedge N$ for a certain class of random band

---

[*]MAP 5, UMR CNRS 8145 - Université Paris Descartes, France.
  E-mail: florent.benaych-georges@parisdescartes.fr
[†]LPMA, Université Paris Diderot, France.
  E-mail: sandrine.peche@math.univ-paris-diderot.fr

matrices (with sub-exponential tails and symmetric distribution). We refer the reader to Spencer [15] and Erdös, Schlein and Yau [8] for a more detailed discussion on the localized/delocalized regime.

Regarding the edges of the spectrum, little is known about the behavior of the extreme eigenvalues and the typical localization length of the associated eigenvectors. As far as the limit of the largest eigenvalue is concerned, Khorunzhiy proved in [11] that for Gaussian band matrices, if $(\log N)^{3/2} \ll W \ll N$, then the extreme eigenvalues converge to the bounds $u_\pm$ of the support of the limiting spectral measure (which is the semicircle law). For matrices with cyclic band structure and Bernoulli entries, Sodin extended this result to the case where $\log N \ll W \ll N$ in [14], where he proved important results about the fluctuations of the extreme eigenvalues around their limits. Concerning the localization length L of the eigenvectors associated to the extreme eigenvalues, one could conjecture the following on the basis of the Thouless argument as explained in [9]. For eigenvectors associated to eigenvalues λ close to the bottom edge e.g. u_-, the localization strength should behave as $L \sim N \wedge W^2(\lambda - u_-)$. Sodin's statement [14] combined with Erdös-Knowles-Yau-Yin's results [7] suggest that this should hold true as soon as $W \gg N^{5/6}$. Moreover, Steinerberger proved recently in [16] that for matrices with Bernoulli entries and cyclic band structure, with probability tending to one, we have either $L \ll W^{7/5}$ or there is strong interaction between two eigenvectors in an interval of length $W^{7/5}$. Let us also mention that in the quite different framework of band matrices with heavy tailed entries, a transition between the localized and the delocalized regime at the edge was proved by the authors of the present text in [3].

In this text, we consider an random $N \times N$ Hermitian matrix $\mathbf{X}$ such that rows of $\mathbf{X}$ have W non identically zero entries (such as, for example, standard or cyclic band matrices). We always suppose that $1 \ll W \ll N$. We first prove that if the entries are independent, centered, have variance one, satisfy a certain tail upper-bound condition and $W \gg (\log N)^{6(1+\alpha)}$, where α is a positive parameter depending on the distribution of the entries, then the largest eigenvalue of $\mathbf{X}/\sqrt{W}$ converges to the upper bound of its limit spectral distribution, that is 2, as for Wigner matrices. This extends the above mentioned results by Khorunzhiy and Sodin [11, 14] where less hypotheses were made on W, but more hypotheses were made about the law of the entries (they use in a crucial way the fact that the entries are symmetrically distributed) and about the structure of the matrix (in our result, we only need that most rows have W non zero entries, no matter what position the entries have on the matrix). Then, under some close hypotheses, we prove a delocalization result for the eigenvectors of $\mathbf{X}$, precisely that most of them cannot be essentially localized on less than $W/\log N$ entries. This lower bound on the localization length has to be compared to the recent result by Steinerberger in [16], which states that the localization *length* (here we use the word *length* rather than *strength* because in [16], the author considers *intervals* carrying most of the ℓ^2-mass of the eigenvectors) in the edge is $\ll W^{7/5}$ or there is strong interaction between two eigenvectors in an interval of length $W^{7/5}$. The paper is organized as follows: our main results are stated in the next section, Theorem 1.4 is proved in Section 2, Theorem 1.9 is proved in Section 3, and some technical results needed here are proved in Section 4 and in the appendix.

Notation. Here, $A \ll B$ means that $A/B \longrightarrow 0$ as $N \to \infty$. Let $\|\mathbf{X}\|$ denote the spectral radius of the Hermitian matrix $\mathbf{X}$ and $\lambda_{\max}(\mathbf{X})$ denote its largest eigenvalue.

1 Main results

We make the following hypotheses.

Hypothesis 1.1. *The matrix* $\mathbf{X} = (X_{ij})$ *is an* $N \times N$ *Hermitian random matrix (depend-*

ing implicitly on N) with independent entries (modulo the symmetry).

Hypothesis 1.2. *There is $W = W(N)$ such that*

$$1 \ll W \ll N \tag{1.1}$$

and such that on each row of $\mathbf{X}$, the number of non identically zero entries is $\leq W$, with equality on all but $o(N)$ rows. All non identically zero entries of $\mathbf{X}$ are centered with variance one. Moreover, there exist constants $C \in [0, +\infty)$ and $\alpha \in [0, +\infty)$ such that for all $k \geq 2$,

$$\mathbb{E}[|X_{ij}|^k] \leq (Ck)^{\alpha k}, \tag{1.2}$$

uniformly on N, i, j.

Then the following theorem has been proved under weaker moment hypotheses in [4] (using the resolvent approach), but can easily be reproved here using a standard moment method as in [1, 2].

Theorem 1.3. *Under Hypotheses 1.1 and 1.2, as $N \to \infty$, the empirical spectral law of $\frac{\mathbf{X}}{\sqrt{W}}$ converges weakly in probability towards the law $\frac{1}{2\pi}\sqrt{4 - x^2}\mathrm{d}x$, with support $[-2, 2]$.*

Our first result is the following one.

Theorem 1.4. *Under Hypotheses 1.1 and 1.2, suppose that W satisfies*

$$W \gg (\log N)^{6(1+\alpha)}, \tag{1.3}$$

with α the constant of (1.2). Then as $N \to \infty$, we have the convergence in probability

$$\frac{\lambda_{\max}(\mathbf{X})}{\sqrt{W}} \longrightarrow 2. \tag{1.4}$$

Remark 1.5. This theorem extends some results of [11] and [14]. In these papers, the convergence of (1.4) is proved under the respective hypotheses $W \gg (\log N)^{3/2}$ and $W \gg \log N$, but for some particular models of matrices: in [11] the matrices considered are Gaussian and in [14], they have Bernoulli distributed entries. Both make a crucial use of the fact that the entries are symmetrically distributed. Moreover, in both papers, the authors also suppose and rely heavily on a particular position of the non zero entries of $\mathbf{X}$. We do not make such a hypothesis here.

To state our main result, a lower bound on the localization length of eigenvectors of $\mathbf{X}$, we slightly modify the hypotheses.
Let $\mathbf{X}$ be a random matrix satisfying Hypothesis 1.1. We make the following two assumptions.

Hypothesis 1.6. *For a certain sequence $W = W(N) \gg 1$, we have the convergence in probability $\frac{\|\mathbf{X}\|}{2\sqrt{W}} \underset{N \to \infty}{\longrightarrow} 1$.*

For example, Hypothesis 1.6 is satisfied if $\mathbf{X}$ satisfies the hypotheses of Theorem 1.4 or those the papers [11] and [14] (see Remark 1.5 above). However we emphasize that Hypothesis 1.6, focused on the extreme eigenvalues, does not make (at least directly) any assumption on the maximal number of non zero entries per row of $\mathbf{X}$ (it may be N), neither on the relative growth of W with respect to N.

We also reinforce the assumption on the tail of the distribution of the entries. Let $C > 0$ be fixed.

Hypothesis 1.7. *The entries X_{ij} belong to the set $\mathcal{E}_C$ of complex random variables X such that*

$$\mathbb{E}X = 0, \qquad \mathbb{E}|X|^k \leq (Ck)^{k/2}, \, \forall k \in \mathbb{N}.$$

Note that this assumption reinforces (1.2) as $\alpha \leq 1/2$. It is equivalent to the fact that there exists $\delta > 0$ and $K > 0$ such that

$$\mathbb{E}e^{\delta|X|^2} \leq K. \tag{1.5}$$

With a slight abuse of notation, we denote $\mathcal{E}_C$ by $\mathcal{E}_{\delta,K}$, as our proof mostly uses assumption (1.5).

The following theorem is the main result of this text.

We use the following Definition 7.1 from Erdös, Schlein and Yau [8]: for L a positive integer and $\eta > 0$, a unit vector $\mathbf{v} = (v_1, \ldots, v_N) \in \mathbb{C}^N$ is said to be (L, η)-*localized* if there exists a set $S \subset \{1, \ldots, N\}$ such that $|S| = L$ and $\sum_{j \in S^c} |v_j|^2 \leq \eta$.

Remark 1.8. The largest L and η are, the strongest the statement *"there is no (L, η)-localized eigenvector"* is.

Theorem 1.9. *We suppose Hypotheses 1.1, 1.6 and 1.7. Fix $\eta \in (0, 1/2)$ and choose $L = L(N)$ such that*

$$L \ll \frac{W}{\log N}. \tag{1.6}$$

Let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of $\mathbf{X}$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_N$ be some associated normalized eigenvectors. Then for any κ such that $\sqrt{\eta/(1-\eta)} < \kappa < 1$,

$$\mathbb{P}(\exists i, \; |\lambda_i| \geq 2\kappa\sqrt{W} \text{ and } \mathbf{v}_i \text{ is } (L, \eta)\text{-localized}) \underset{N \to \infty}{\longrightarrow} 0.$$

Remark 1.10. The same proof can also bring to a version of this theorem where $\eta = \eta(N) \longrightarrow 0$. In this case, $\kappa = \kappa(N)$ is allowed to tend to zero, thus the theorem allows to lower bound the localization length of most eigenvectors of $\mathbf{X}$.

Remark 1.11. The estimate in Theorem 1.9 is almost sharp, as shown by the case where $\mathbf{X}$ is the block diagonal matrix formed with $[N/W](+1)$ GUE matrices of size W (or at most W for the last block).

2 Proof of Theorem 1.4

The proof goes along the same lines as the proof of Theorem 2 in the paper [10] by Füredi and Komlós (see also Theorem 2.1.22 in [1]). First note that by Theorem 1.3, we already know that for any $\eta > 0$, $\mathbb{P}(\lambda_{\max}(\mathbf{X}) < (2 - \eta)\sqrt{W}) \longrightarrow 0$.

For any $\eta > 0$, for any $k \geq 1$,

$$\mathbb{P}(\lambda_{\max}(\mathbf{X}) > (2 + \eta)\sqrt{W}) \leq \mathbb{P}(\operatorname{Tr}\mathbf{X}^{2k} \geq (2 + \eta)^{2k}W^k) \leq W^{-k}(2 + \eta)^{-2k}\mathbb{E}\operatorname{Tr}\mathbf{X}^{2k},$$

hence it suffices to find a sequence $k = k(N)$ such that for any $\eta > 0$,

$$\mathbb{E}\operatorname{Tr}\mathbf{X}^{2k} \ll W^k(2 + \eta)^{2k}. \tag{2.1}$$

We have

$$\mathbb{E}\operatorname{Tr}\mathbf{X}^{2k} = \sum \mathbb{E}X_{i_1 i_2} \cdots X_{i_{2k} i_1},$$

where the sum is over collections $\mathbf{i} = (i_1, \ldots, i_{2k})$ such that for all ℓ, $i_\ell \in \{1, \ldots, N\}$. For each $\mathbf{i}$, let $G_\mathbf{i}$ be the simple, non oriented graph with vertex set $\{i_1, \ldots, i_{2k}\}$ and edges $\{i_\ell, i_{\ell+1}\}$ ($1 \leq \ell \leq 2k$, with the convention $i_{2k+1} = i_1$). For the expectation

in the RHT above to be non zero, we need all edges to be visited at least twice by the path **i** (because the X_{ij}'s are centered) and the edges $\{i_\ell, i_{\ell+1}\}$ to be such that $X_{i_\ell, i_{\ell+1}}$ is non identically zero. The symmetric group S_N acts on the set of **i**'s by $\sigma \cdot (i_1, \ldots, i_{2k}) := (\sigma(i_1), \ldots, \sigma(i_{2k}))$. Following Section 2.1.3 of [1], we denote by $\mathcal{W}_{2k,t}$ the set of equivalence classes, under the action of S_N, of **i**'s such that all edges of $G_{\mathbf{i}}$ are visited at least twice by the path **i** and $G_{\mathbf{i}}$ has exactly t vertices (this set is actually stable under this action).

Note that for $\mathcal{W}_{2k,t}$ to be non empty, we need to have $t \le k+1$. Indeed, $G_{\mathbf{i}}$ is always connected hence its number of vertices is at most its number of edges plus one.

Note that for any $w \in \mathcal{W}_{2k,t}$, the number of **i**'s in the class w is at most NW^{t-1}.

It follows from the previous remarks that

$$\mathbb{E} \operatorname{Tr} \mathbf{X}^{2k} \le N \sum_{t=1}^{k+1} W^{t-1} \sum_{w \in \mathcal{W}_{2k,t}} \max_{\mathbf{i} \in w} \mathbb{E} X_{i_1 i_2} \cdots X_{i_{2k-1} i_1}.$$

Now, let us fix $t \in \{1, \ldots, k+1\}$, $w \in \mathcal{W}_{2k,t}$ and $\mathbf{i} \in w$. Let us denote by l (resp. m) the number of edges of $G_{\mathbf{i}}$ visited exactly twice (resp. at least three times). Obviously, $2l + 3m \le 2k$. Moreover, as $l + m$ is the number of edges of the $G_{\mathbf{i}}$, hence by connectedness of $G_{\mathbf{i}}$ again, we have $t \le l + m + 1$. So

$$6t \le 6m + 6l + 6 = 2(3m + 2l) + 2l + 6 \le 4k + 2l + 6,$$

so

$$2k - 2l \le 6(k - t + 1).$$

Now, notice that as the X_{ij}'s have variance one, $\mathbb{E} X_{i_1 i_2} \cdots X_{i_{2k} i_1}$ can be reduced to the expectation of a product of $2k - 2l$ X_{ij}'s, hence by (1.2) and Hölder's inequality,

$$\mathbb{E} X_{i_1 i_2} \cdots X_{i_{2k} i_1} \le \left(C(2k - 2l) \right)^{\alpha(2k-2l)} \le \{6C(k-t+1)\}^{6\alpha(k-t+1)}.$$

As a consequence,

$$\mathbb{E} \operatorname{Tr} \mathbf{X}^{2k} \le N \sum_{t=1}^{k+1} W^{t-1} \# \mathcal{W}_{2k,t} \times \{6C(k-t+1)\}^{6\alpha(k-t+1)}.$$

Now, we shall use Lemma 2.1.23 of [1], which states that $\# \mathcal{W}_{2k,t} \le 4^k (2k)^{6(k-t+1)}$ as soon as $t \le k+1$ (the case $t = k+1$ is technically not contained in Lemma 2.1.23 of [1], but follows from Equation (2.1.20) and Lemma 2.1.3 of the same book). It follows that

$$
\begin{aligned}
\mathbb{E} \operatorname{Tr} \mathbf{X}^{2k} &\le N 4^k \sum_{t=1}^{k+1} W^{t-1} (2k)^{6(k-t+1)} \{6C(k-t+1)\}^{6\alpha(k-t+1)} \\
&= N W^k 4^k \sum_{i=0}^{k} W^{-i} (2k)^{6i} (6Ci)^{6\alpha i} \\
&\le N W^k 4^k \left(1 - \frac{(2k(6Ck)^\alpha)^6}{W} \right)^{-1},
\end{aligned}
$$

where the last inequality is true as soon as $W > (2k(6Ck)^\alpha)^6$. Then it is easy to see that (2.1) holds for $k = k(N)$ such that $\log N \ll k \ll W^{\frac{1}{6(1+\alpha)}}$.

3 Proof of Theorem 1.9

Before proving Theorem 1.9, we shall need the following theorem and its corollary. The proof of Theorem 3.1 is postponed to Section 4.

Theorem 3.1. *Under Hypotheses 1.1 and 1.7, there are constants $c_2 = c_2(\delta, K) > 0$ and $C_2 = C_2(\delta, K) < \infty$ independent of all the other parameters such that for all $t > 0$,*

$$\mathbb{P}(\|\mathbf{X}\| > t\sqrt{N}) \leq e^{-c_2(t^2 - C_2)N}. \tag{3.1}$$

Let us denote by $\rho(\mathbf{X})$ the spectral radius of $\mathbf{X}$ and, for $L \geq 1$, by $\rho_L(\mathbf{X})$ the maximum spectral radius of its $L \times L$ principal submatrices (a *principal* submatrix is a submatrix chosen by extracting a certain set of columns and *the same* set of rows, but this set does not need to be an interval).

Corollary 3.2. *Under Hypotheses 1.1 and 1.7, there exists $t < \infty$ and $c_3 > 0$ such that*

$$\mathbb{P}(\rho_L(\mathbf{X}) \geq t\sqrt{L \log N}) \leq e^{-c_3 L \log N}. \tag{3.2}$$

Proof. The number of ways to choose an $L \times L$ principal submatrix is $\leq N^L = e^{L \log N}$. For each submatrix $\mathbf{S}$, $\mathbb{P}(\rho(\mathbf{S}) \geq t\sqrt{N \log N}) \leq \exp\{-c_2(t^2 \log N - C_2)L\}$, hence by the union bound,

$$\mathbb{P}(\rho_L(\mathbf{X}) \geq t\sqrt{L \log N}) \leq \exp\{[-c_2(t^2 \log N - C_2) + \log N]L\},$$

thus if $c_2 t^2 > 1$, then (3.2) holds for a certain $c_3 > 0$. □

To prove Theorem 1.9, we shall also need the following lemma (see Lemma 4.2 in [3]).

Lemma 3.3. *For all i, if $\mathbf{v}_i$ is (L, η)-localized, then $|\lambda_i| \leq \frac{\rho_L(\mathbf{X}) + \sqrt{\eta}\rho(\mathbf{X})}{\sqrt{1-\eta}}$.*

Let us now prove Theorem 1.9.

Proof. Let us choose $\varepsilon > 0$ such that $\sqrt{\eta/(1-\eta)}(1+\varepsilon) < \kappa$ and set

$$\delta := \kappa - \sqrt{\eta/(1-\eta)}(1+\varepsilon).$$

We know, by (1.6), that with probability tending to 1, $\rho(\mathbf{X}) \leq (1+\varepsilon)2\sqrt{W}$, *i.e.*

$$\frac{\sqrt{\eta}\rho(\mathbf{X})}{\sqrt{1-\eta}} \leq (\kappa - \delta)2\sqrt{W}.$$

Moreover, by Corollary 3.2, there is $t < \infty$ such that with probability tending to one, $\rho_L(\mathbf{X}) \leq t\sqrt{L \log N}$. But by (1.6), for N large enough,

$$\frac{t\sqrt{L \log N}}{\sqrt{1-\eta}} \leq 2\delta\sqrt{W},$$

so the theorem is proved. □

4 Proof of Theorem 3.1

4.1 A preliminary lemma

Lemma 4.1. *Under Hypotheses 1.1 and 1.7, there are constants c_1, C_1 depending only on δ, K such that for any $\mathbf{z} \in \mathbb{C}^N$ with $|\mathbf{z}| \leq 1$,*

$$\mathbb{P}(\mathbf{z}^* \mathbf{X}^* \mathbf{X} \mathbf{z} \geq Nt) \leq e^{-c_1(t-C_1)N}. \tag{4.1}$$

Remark 4.2. If $\mathbf{X}$ is a random $N \times N$ matrix whose maximum number of non identically zero entries per row is W (like for a *band matrix* with band width W), then (4.1) remains true with N replaced by W everywhere (for some constants still depending only on δ and K).

Proof. We denote by $\mathbf{X}_1, \ldots, \mathbf{X}_N$ the columns of $\mathbf{X}$. We have , for any τ, C as in Lemma 5.2 of the appendix,

$$\mathbb{E}e^{\tau^2 \mathbf{z}^* \mathbf{X}^* \mathbf{X} \mathbf{z}} \;=\; \mathbb{E}e^{\tau^2 |\mathbf{X}\mathbf{z}|^2} \;=\; \mathbb{E}e^{\tau^2 \sum_j |\mathbf{X}_j \cdot \mathbf{z}|^2} \;\leq\; e^{NC\tau^2}$$

Hence

$$\mathbb{P}(\mathbf{z}^* \mathbf{X}^* \mathbf{X} \mathbf{z} \geq Nt) \leq \mathbb{E}e^{\tau^2 \mathbf{z}^* \mathbf{X}^* \mathbf{X} \mathbf{z}} e^{-\tau^2 Nt} \leq e^{-\tau^2 (t-C)N}.$$

So the lemma is proved. $\qquad\square$

4.2 Proof of Theorem 3.1

Lemma 4.3. *Let $N \geq 1$. For any fixed $0 < \varepsilon < 1/4$, there exists a family $(\mathbf{z}_i)_{i \in I}$ of elements of the unit ball of $\mathbb{C}^N$ such that $|I| \leq (2/\varepsilon)^{2N}$ and any element of the unit sphere of $\mathbb{C}^N$ is within a distance at most ε of one of the $\mathbf{z}_i$'s. Moreover, for any positive $N \times N$ Hermitian matrix P,*

$$\lambda_{\max}(P) \leq \frac{\max_i \mathbf{z}_i^* P \mathbf{z}_i}{1 - 2\varepsilon}.$$

The first part of this lemma is well known (see e.g. [12]), whereas its second part follows from the fact that for any vectors of the unit ball $\mathbf{z}, \mathbf{z}_i$,

$$\left| \mathbf{z}_i^* P \mathbf{z}_i - \mathbf{z}^* P \mathbf{z} \right| \leq 2\|P\| \times \|\mathbf{z} - \mathbf{z}_i\|$$

and specifying $\mathbf{z}$ to be an eigenvector associated to $\lambda_{\max}$. Let us now prove Theorem 3.1.

Proof. By Lemma 4.1, we know that there are constants there are constants c_1, C_1 depending only on δ, K such that for any $\mathbf{z} \in \mathbb{C}^N$ with $\|\mathbf{z}\| \leq 1$,

$$\mathbb{P}(\mathbf{z}^* \mathbf{X}^* \mathbf{X} \mathbf{z} \geq Nt) \leq e^{-c_1(t-C_1)N}.$$

Now, by Lemma 4.3, we have

$$
\begin{aligned}
\mathbb{P}(\lambda_{\max}(\mathbf{X}^* \mathbf{X}) \geq Nt) \;&\leq\; \mathbb{P}(\max_i \mathbf{z}_i^* \mathbf{X}^* \mathbf{X} \mathbf{z}_i \geq Nt(1 - 2\varepsilon)) \\
&\leq\; (2/\varepsilon)^{2N} e^{-c_1(t(1-2\varepsilon) - C_1)N} \\
&=\; e^{(-c_1(1-2\varepsilon)(t - \frac{C_1}{1-2\varepsilon}) + 2\log(2/\varepsilon))N} \\
&=\; e^{-c_2(t-C_2)N}
\end{aligned}
$$

As a consequence,

$$\mathbb{P}(\|\mathbf{X}\| > t\sqrt{N}) = \mathbb{P}(\lambda_{\max}(\mathbf{X}^* \mathbf{X}) \geq t^2 N) \leq e^{-c_2(t^2 - C_2)N}.$$

$\qquad\square$

5 Appendix: technical results

Lemma 5.1. *For any real centered random variable X and any $r \in \mathbb{R}$, we have*

$$\mathbb{E}e^{rX} \leq 1 + 3\mathbb{E}[e^{\delta X^2}](e^{r^2/\delta} - 1) \leq e^{3r^2 \mathbb{E}[e^{\delta X^2}]/\delta}$$

for any $\delta > 0$.

Proof. The second inequality follows from the fact that for any $y \geq 1$, we have the inequality $1 + y(e^{r^2/\delta} - 1) \leq e^{r^2 y/\delta}$ (this is obvious with the series expansion of $\exp$).

So let us prove the first inequality. Note that up to a replacement of X by rX and of δ by δ/r^2, we shall suppose that $r = 1$.

The case where $\mathbb{E}[e^{\delta X^2}] = \infty$ is obvious, hence we focus on the other case, which allows to expend all sums with the moments of X.

Claim : for all $x \geq 0$, $e^x \leq 1 + x + 3\frac{e^x + e^{-x} - 2}{2}$. Indeed, both terms are equal for $x = 0$ and the derivative of 2RHT$-$2LHT is $e^x - 3e^{-x} + 2$, which is increasing, hence has the same sign as x.

It follows that

$$\mathbb{E}e^X \leq 1 + 3\sum_{n \geq 1} \frac{\mathbb{E}X^{2n}}{(2n)!} \leq 1 + 3\sum_{n \geq 1} \frac{n!\delta^{-n}\mathbb{E}e^{\delta X^2}}{(2n)!} \leq 1 + 3\mathbb{E}[e^{\delta X^2}](e^{1/\delta} - 1),$$

where we used first $\mathbb{E}X = 0$, then $\frac{\delta^n \mathbb{E}X^{2n}}{n!} \leq \mathbb{E}e^{\delta X^2}$ and at last $\frac{n!}{(2n)!} \leq \frac{1}{n!}$. $\qquad\square$

Let $\mathbf{a} \cdot \mathbf{b}$ denote the standard scalar product of two complex vectors and let $|\cdot|$ denote the associated norm.

Lemma 5.2. *Let us fix $\delta, K > 0$. Then there is $\tau = \tau(\delta, K) > 0$ and $C = C(\delta, K) > 0$ such that for all $N \geq 1$, all $\mathbf{z} \in \mathbb{C}^N$ such that $|\mathbf{z}| \leq \tau$, for any $\mathbf{Y}$ random vector taking values in $\mathbb{C}^N$ with independent components in the set $\mathcal{E}_{\delta,K}$ defined at Hypothesis 1.7,*

$$\mathbb{E}e^{|\mathbf{Y} \cdot \mathbf{z}|^2} \leq e^{C|\mathbf{z}|^2}.$$

Proof. First step: Let us first prove the result for $\mathbf{Y}$ having independent components in

$$\mathcal{E}_{\delta,K}^{\mathbb{R}} := \{Y \in \mathcal{E}_{\delta,K} \, ; \, Y \text{ is real-valued}\}$$

and $\mathbf{z} \in \mathbb{R}^N$. Let $\tau_{\mathbb{R}} > 0$ be such that for any $t \in [0, \tau_{\mathbb{R}})$, we have

$$12t^2 K/\delta < 1 \qquad \text{and} \qquad \left(1 - 12t^2 K/\delta\right)^{-1/2} \leq e^{12t^2 K/\delta}.$$

Let g be a standard real Gaussian variable, independent of the other variables, let $\mathbb{E}_g$ denote the expectation with respect to g and let $\mathbb{E}$ denote the expectation with respect to all other variables than g.

For any $\tau > 0$ and $\mathbf{z} \in \mathbb{R}^N$ such that $|\mathbf{z}| \leq \tau$, using the formula $e^{x^2} = \mathbb{E}_g e^{\sqrt{2}xg}$, we have

$$\mathbb{E}e^{|\mathbf{Y} \cdot \mathbf{z}|^2} = \mathbb{E}\mathbb{E}_g e^{\sqrt{2}g\mathbf{Y} \cdot \mathbf{z}} = \mathbb{E}_g \mathbb{E}e^{\sqrt{2}g\mathbf{Y} \cdot \mathbf{z}} = \mathbb{E}_g \prod_i \mathbb{E}e^{\sqrt{2}gY_i z_i}.$$

Hence by Lemma 5.1,

$$\mathbb{E}e^{|\mathbf{Y} \cdot \mathbf{z}|^2} \leq \mathbb{E}_g \prod_i e^{6g^2 z_i^2 K/\delta} = \mathbb{E}_g e^{6g^2 |\mathbf{z}|^2 K/\delta} = \left(1 - 12|\mathbf{z}|^2 K/\delta\right)^{-1/2} \leq e^{12|\mathbf{z}|^2 K/\delta}.$$

Second step: To extend this result to the complex case, just decompose Y and $\mathbf{z}$ into real and imaginary parts and use Hölder inequality to see that the constants $\tau = \frac{\tau_{\mathbb{R}}}{\sqrt{8}}$ and $C = 4C_{\mathbb{R}}$ are convenient in the general case. $\qquad\square$

References

[1] G. Anderson, A. Guionnet, O. Zeitouni *An Introduction to Random Matrices.* Cambridge studies in advanced mathematics, **118** (2009). MR-2760897

[2] Z.D. Bai, J.W. Silverstein *Spectral analysis of large dimensional random matrices.* Second Edition, Springer, New York, 2009. xvi+551 pp. MR-2567175

[3] F. Benaych-Georges, S. Péché *Localization and delocalization for heavy tailed band matrices*, to appear in Ann. Inst. Henri Poincaré Probab. Stat. arXiv:1210.767

[4] L. V. Bogachev, S. A. Molchanov, L. A. Pastur *On the density of states of random band matrices*, Mat. Zametki 50 (1991), 31–42, **157**. MR-1150631

[5] L. Erdös, A. Knowles *Quantum diffusion and eigenfunction delocalization in a random band matrix model*. Comm. Math. Phys. **303** (2011), no. 2, 509–554. MR-2782623

[6] L. Erdös, A. Knowles *Quantum diffusion and delocalization for band matrices with general distribution.* Ann. Henri Poincaré **12** (2011), no. 7, 1227–1319. MR-2846669

[7] L. Erdös, A. Knowles, H-T. Yau, J. Yin *Delocalization and Diffusion Profile for Random Band Matrices*. Comm. Math. Phys. **323** (2013), no. 1, 367–416. MR-3085669

[8] L. Erdös, B. Schlein, H.T. Yau *Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices*, Ann. Prob. **37** (2009), no. 3, 815–852. MR-2537522

[9] Y.V. Fyodorov, A.D. Mirlin *Scaling properties of localization in random band matrices: a σ-model approach*, Phys. Rev. Lett. **67** (1991), no. 18, 2405-2409. MR-1130103

[10] Z. Füredi, J. Komlós *The eigenvalues of random symmetric matrices.* Combinatorica **1** (1981), no. 3, 233-241. MR-0637828

[11] O. Khorunzhiy *Estimates for moments of random matrices with Gaussian elements* in Séminaire de Probabilités XLI, Lecture Notes in Math. 1934, Springer-Verlag, New York, 2008, pp. 51–92. MR-2483726

[12] M. Ledoux, M. Talagrand *Probability in Banach spaces. Isoperimetry and processes.* Reprint of the 1991 edition. Classics in Mathematics. Springer-Verlag, Berlin, 2011. xii+480 pp. MR-2814399

[13] J. Schenker *Eigenvector localization for random band matrices with power law band width.* Comm. Math. Phys. **290** (2009), no. 3, 1065-1097. MR-2525652

[14] S. Sodin *The spectral edge of some random band matrices*, Ann. of Math. **172** (2010), 2223–2251. MR-2726110

[15] T. Spencer *Random banded and sparse matrices* The Oxford handbook of random matrix theory. Oxford University Press (2011) 471–488. MR-2932643

[16] S. Steinerberger *On Eigenvectors of Random Band Matrices with Large Band*, arXiv:1307.5753

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

Economical model of EJP-ECP

- Low cost, based on free software (OJS[1])

- Non profit, sponsored by IMS[2], BS[3], PKP[4]

- Purely electronic and secure (LOCKSS[5])

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]OJS: Open Journal Systems http://pkp.sfu.ca/ojs/
[2]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[3]BS: Bernoulli Society http://www.bernoulli-society.org/
[4]PK: Public Knowledge Project http://pkp.sfu.ca/
[5]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm