RANDOM WEIGHTED PROJECTIONS, RANDOM QUADRATIC FORMS, RANDOM EIGENVECTORS

VAN VU AND KE WANG

Abstract. In this paper, we present a simple, yet useful, concentration result concerning random (weighted) projections in high dimensional spaces. As application, we prove a general concentration result for random quadratic forms, extended a classical result of Hanson and Wright and improved several recent results. In another application, we show that the infinity norm of most (unit) eigenvectors of a random ± 1 matrix is $O(\sqrt{\log n/n})$, which is optimal, sharpening various earlier estimates. In fact, the estimate holds for a large class of random matrices.

As a by-product, we also obtain an estimate on the threshold for the local semi-circle law. This estimate is tight up to a $\sqrt{\log n}$ factor. It is an interesting open question to see if this factor is necessary.

1. Introduction

1.1. **Projection of a random vector.** Consider $\mathbb{C}^n$ with a subspace H of dimension d. Let $X = (\xi_1, \dots, \xi_n)$ be a random vector. The length of the orthogonal projection of X onto H is an important parameter which plays an essential role in the studies of random matrices and related areas.

In [19], Tao and the first author showed that (under certain conditions) this length is strongly concentrated. In other words, the projection of X onto H lies essentially on a circle centered at the origin. This fact played a crucial role in the computation of the determinant of a random matrix with iid entries. As the absolute value of the determinant is the volume of the parallelepiped spanned by the row vectors, one can expose these vectors in some order and compute the volume as the product of the distances from each vector to the subspace spanned by the previous ones. On the other hand, the distance can be computed from the length of X (which is usually easy to estimate) and the length of the projection. (We only talk about orthogonal projections in this paper and will omit the word "orthogonal" from this point.)

Lemma 1 (Projection Lemma)**.** [19] *Let $X = (\xi_1, \dots, \xi_n)$ by a random vector in $\mathbb{C}^n$ whose coordinates ξ_i are independent random variables with mean 0 and variance 1. Assume furthermore that there is a number K (which may depend on n) such that $|\xi_i| \le K$ with probability 1 and $K \ge \mathbf{E}|\xi_i|^4 + 1$ for all i. Let H be a subspace of dimension d and $\Pi_H X$ be the length of the projection of X onto H*

$$\mathbf{P}(|\Pi_H X - d| \ge t) \le 10 \exp(-\frac{t^2}{20K^2}).$$

The constants 10 and 20 are rather arbitrary. We make no attempt to optimize the constants in this paper.

V. Vu is supported by research grants DMS-0901216 and AFOSAR-FA-9550-09-1-0167.

1.2. **Weighted projections.** Let us fix an orthonormal basis $(u_1, \dots, u_d)$ of H. We can express $\Pi_H X$ as

$$\Pi_H X = (\sum_{i=1}^d |u_i^* X|^2)^{1/2}. \tag{1}$$

In recent studies, we came up with situations when the role of the axis u_i are not uniform. Formally speaking, we need to consider a weighted version of (1) where $(\sum_{i=1}^d |u_i \cdot X|^2)^{1/2}$ is replaced by $(\sum_{i=1}^d c_i |u_i^* X|^2)^{1/2}$ with c_i being non-negative numbers (weights). This motivates us to prove the following generalization of the Projection Lemma.

Lemma 2 (Weighted projection lemma)**.** *There are constants $C, C' > 0$ such that there following holds. Let $X = (\xi_1, \dots, \xi_n)$ by a random vector in $\mathbb{C}^n$ whose coordinates ξ_i are independent random variables with mean 0 and variance 1. Assume furthermore that there is a number K (which may depend on n) such that $|\xi_i| \le K$ with probability 1 for all i. Let H be a subspace of dimension d with an orthonormal basis $\{u_1, \dots, u_d\}$. Then for any $1 \ge c_1, \dots, c_d \ge 0$*

$$\mathbf{P}\left(|\sqrt{\sum_{j=1}^d c_j |u_j^* X|^2} - \sqrt{\sum_{j=1}^d c_j}| \ge t \right) \le C \exp(-\frac{t^2}{C' K^2}).$$

If $|\xi| \le K$ with probability one, we say that ξ is K-bounded. As a matter of fact, we can keep $C = 10, C' = 1/20$ as before, but we prefer this setting for the sake of consistancy.

By squaring, it follows that

$$\mathbf{P}(\sum_{j=1}^d c_j (|u_j^* X|^2 - 1) \le 2t\sqrt{\sum_{j=1}^d c_j} + t^2) \le C \exp(-C' \frac{t^2}{K^2}) \tag{2}$$

and for $t \le \sqrt{\sum_i c_i}$

$$\mathbf{P}(\sum_{j=1}^d c_j (|u_j^* X|^2 - 1) \le -2t\sqrt{\sum_{j=1}^d c_j} + t^2) \le C \exp(-C' \frac{t^2}{K^2}) \tag{3}$$

Furthermore,

$$|\sum_{j=1}^d c_j (|u_j^* X|^2 - 1)| \le 2t\sqrt{\sum_{j=1}^d c_j} + t^2 \tag{4}$$

with probability at least $1 - C \exp(-C' \frac{t^2}{K^2})$.

Remark 3. The values of C and C' may be different in different inequalities.

1.3. **Applications.** We are going to present two applications. The first is a general concentration result for a quadratic form $Q := \sum_{1 \le i,j \le n} a_{ij}\xi_i\xi_j$ where ξ_i are random variables. This can be seen as a quadratic version of the well-known Chernoff bound. Already in 1971, Hanson and Wright [12] proved a strong concentration result for the sub-gaussian random variables. Recently, their result has been extended to other variables (with some loss in the bounds). Our result will generalize or strengthen many of the former results. Details will appear in Section 2.

The second application concerns the (infinity) norm of eigenvectors of a random matrix. Estimates for this norm is important for studies in graph theory [5] and random matrices [9] (see also [6, 24] for surveys). We are going to show that the norm of most eigenvectors of a symmetric random ± 1 matrix is $O(\sqrt{\log n/n})$. This bound seems optimal optimal (notice that the infinity norm of a random vector chosen uniformly from the unit sphere is $\Omega(\sqrt{\log n/n})$) and improves several earlier estimates. Our estimate also holds for many other models of random matrices.

As a by-product of the proof, we obtain an almost tight estimate for the validity threshold of the local semi-circle law for random matrices (see Section **??** for details).

1.4. **Weighted projection lemma for unbounded random variables.** In Lemma 2 we assume the ξ_i are K-bounded. In this section, we present two methods to weaken this assumption.

The first is to consider a notion which is weaker than that of K-bounded.

We say a random vector $X = (\xi_1, \dots, \xi_n)$ is K-concentrated (where K may depend on n) if there are constants $C, C' > 0$ such that for any convex, 1-Lipschitz function $F : \mathbb{C}^n \to \mathbb{R}$ and any $t > 0$

$$\mathbf{P}(|F(X) - M(F(X))| \ge t) \le C \exp(-C' \frac{t^2}{K^2}), \tag{5}$$

where $M(Y)$ denotes the median of a random variable Y (choose an arbitrary one of there are many).

Notice that the notion of K-concentrated is somewhat similar to the notion of threshold in random graph theory in the sense that if X is K-concentrated then it is cK-concentrated for any constant $c > 0$ (similarly, if $p(n)$ is a threshold for a property $\mathcal{P}$ (say, containing a triangle) then $cp(n)$ is also a threshold). One can also replace the median by the expectation (see Lemma 18).

Examples of K-concentrated random variables

- If the coordinates of X are iid standard gaussian (real of complex), then X is 1-concentrated (see [14]).
- If ξ_i are independent and ξ_i are K-bounded for all i, then X is K-concentrated (this is a corollary of Talagrand's inequality; see [14, Chapter 4]; [20, Theorem F.5]).
- If X satisfies the log-Sobolev inequality with parameter K^2, then it is K-concentrated (see [14, Theorem 5.3]).

Lemma 4. *Let $X = (\xi_1, \dots, \xi_n)$ be a K-concentrated random vector in $\mathbb{C}^n$ whose coordinates ξ_i have mean 0 and variance 1. Then there are constants $C, C' > 0$ (which depend on, but could be different from the constants in* (5)*) such that the following holds. Let H be a subspace of dimension d with an orthonormal basis $\{u_1, \dots, u_d\}$. Then for any $1 \ge c_1, \dots, c_d \ge 0$*

$$\mathbf{P}\left(|\sqrt{\sum_{j=1}^{d} c_j |u_j^* X|^2} - \sqrt{\sum_{j=1}^{d} c_j}| \geq t\right) \leq C \exp(-C' \frac{t^2}{K^2}).$$

Another way to weaken the K-bounded assumption is to consider truncation. If ξ is not bounded, but has light tail, then by setting K appropriately, we can show that $\mathbf{P}(|\xi| \geq K)$ is negligible with respect to the probability bound we want to prove. Technically speaking, we would like to replace ξ by its truncation $\xi' := \xi \mathbf{I}_{|\xi| \leq K}$. A technical problem here is that the mean and variance of ξ' will not match those of ξ, but this can be handled by an extra normalization step.

Assume that the ξ_i are independent with mean zero and variance one. Choose a number $K > 1$ and let $\varepsilon_1 := \max_{1 \leq i \leq n} \mathbf{P}(|\xi_i| > K)$. Set $\xi_i' := \xi_i \mathbf{I}_{|\xi_i| \leq K}$ and let μ_i and σ_i^2 denote its mean and variance. Set $\varepsilon_2 := \max_{1 \leq i \leq n} |\mu_i|$ and $\varepsilon_3 := \max_{1 \leq i \leq n} |\sigma_i^2 - 1|$. Assume all $\varepsilon_j \leq 1/2$ (in practice this assumption is satisfied easily). Define $\tilde{\xi}_i := \frac{\xi_i' - \mu_i}{\sigma_i}$. The $\tilde{\xi}_i$ are independent with mean zero and variance 1 and is $2K$-bounded. Let $X' := (\xi_1', \dots, \xi_n')$ and $\tilde{X} := (\tilde{\xi}_1, \dots, \tilde{\xi}_n)$. It is obvious that

$$\mathbf{P}\left(|\sqrt{\sum_{j=1}^{d} c_j |u_j^* X|^2} - \sqrt{\sum_{j=1}^{d} c_j}| \geq t\right) \leq \mathbf{P}\left(|\sqrt{\sum_{j=1}^{d} c_j |u_j^* X'|^2} - \sqrt{\sum_{j=1}^{d} c_j}| \geq t\right) + n\varepsilon_1.$$

The next observation is that if $\varepsilon_2, \varepsilon_3$ are small, then $\sum_{1 \leq i \leq d} c_i |u_i^* X'|^2$ and $\sum_{1 \leq i \leq d} c_i |u_i^* \tilde{X}|^2$ are more or less the same. By definition, we have with probability one

$$|\xi_i' - \tilde{\xi}_i| = |\frac{\xi_i(\sigma_i - 1) + \mu_i}{\sigma_i}| \leq 2(K\epsilon_3 + \epsilon_2).$$

It follows that $D := X' - \tilde{X}$ has norm at most $2n^{1/2}(K\epsilon_3 + \epsilon_2)$ with probability one. On the other hand,

$$\Big| \sum_{1 \leq i \leq d} c_i |u_i^* X'|^2 - \sum_{1 \leq i \leq d} c_i |u_i^* \tilde{X}|^2 \Big| \leq 2 \sum_{1 \leq i \leq d} c_i |u_i^* X_i'| |u_i^* D| + c_i |u_i^* D|^2.$$

As u_i are unit vectors, $|u_i^* X_i'| \leq \|X_i'\| \leq \sqrt{n} K$ and $|u_i^* D_i| \leq \|D_i\| \leq 2\sqrt{n}(K\epsilon_2 + \epsilon_3)$ (these bounds are generous and can be improved by a polynomial factor in certain cases, but in applications such improvement rarely matter). It follows, again rather generously

$$| \sum_{1 \leq i \leq n} c_i |u_i^* X'|^2 - \sum_{1 \leq i \leq n} c_i |u_i^* \tilde{X}|^2 | \leq 4n \sum_{i=1}^{d} c_i K^2 (\epsilon_2 + \epsilon_3) \leq 4n^2 K^2 (\varepsilon_2 + \varepsilon_3).$$

Applying Lemma 2 for $\tilde{X}$, we obtain

Lemma 5. *There are constants $C, C > 0$ such that the following holds. Let $X = (\xi_1, \dots, \xi_n)$ by a random vector in $\mathbb{C}^n$ whose coordinates ξ_i are independent random variables with mean 0 and variance 1. Under the above notation, we have, for any $1 \geq c_1, \dots, c_n \geq 0$ and $t > 0$*

$$\mathbf{P}\left(|\sqrt{\sum_{j=1}^{d} c_j |u_j^* X|^2} - \sqrt{\sum_{j=1}^{d} c_j}| \geq t + 4n^2K^2(\epsilon_2 + \epsilon_3)\right) \leq C\exp(-C'\frac{t^2}{K^2}) + n\epsilon_1. \tag{6}$$

In practice, ε_j are typically super-polynomially small, which yields $4n^2K^2(\epsilon_2+\epsilon_3) = o(1)$. This term can be ignored (by slightly changing the values of C, C' if necessary) and we end up with a more friendly inequality

$$\mathbf{P}(|\sqrt{\sum_{j=1}^{d} c_j |u_j^* X|^2} - \sqrt{\sum_{j=1}^{d} c_j}| \geq t) \leq C\exp(-C'\frac{t^2}{K^2}) + n\epsilon_1. \tag{7}$$

As an illustration, let us consider the following tail-decay assumption, which comes up frequently in practice.

Definition 6. We say that ξ is sub-exponential with exponent α if there are constants $a, b > 0$ such that for all $t > 0$

$$\mathbf{P}(|\xi - \mathbf{E}\xi| \geq t^\alpha) \leq a\exp(-bt). \tag{8}$$

If $\alpha = 1/2$ then ξ is sub-gaussian.

For a sufficiently large K (compared to a and b), $\varepsilon_j \leq \exp(-\frac{b}{2}K^{1/\alpha})$ for $j = 1, 2, 3$. For $K = \omega(\log^\alpha n)$, $n^2K^2\exp(-\frac{b}{2}K^{1/\alpha}) = o(1)$ and (7) yields

$$\mathbf{P}\left(|\sqrt{\sum_{j=1}^{d} c_j |u_j^* X|^2} - \sqrt{\sum_{j=1}^{d} c_j}| \geq t\right) \leq C\exp(-C'\frac{t^2}{K^2}) + n\exp(-\frac{b}{2}K^{1/\alpha}). \tag{9}$$

1.5. **Structure of the paper. Notation.** We use standard assumption notation such as O, o, Θ, etc under the assumption that $n \to \infty$. For a vector X, $\|X\|$ is its Euclidean norm and $\|X\|_\infty$ its infinity norm. For a matrix $A \in \mathbb{C}^{n\times n}$, $\|A\|_F$ and $\|A\|_2$ denote the Frobenius and spectral norm, respectively. All eigenvectors will have unit length.

2. Concentration inequalities for quadratic forms

Consider a quadratic form $Y := X^*AX$ where $X = (\xi_1, \dots, \xi_n)$ is, as usual, a random vector and $A = (a_{ij})_{1\leq i,j\leq n}$ a deterministic matrix. In this section, we aim to prove a large deviation result for Y, which can be seen as the quadratic version of the standard Chernoff bound. Quadratic forms of random variables appear frequently in applications and the large deviation problem been considered by several researchers.

In 1971, Hanson and Wright [12] obtained the first important inequality for sub-gaussian random variables.

Theorem 7 (Hanson-Wright inequality). *Let $X = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n$ be random vector with ξ_i being iid symmetric and sub-gaussian random variables with mean 0 and variance 1. There exist constants $C, C' > 0$ (which may depend on the constants in Definition 8) such that the following hold. Let A be a real matrix of size n with entries a_{ij} and $B := (|a_{ij}|)$. Then*

$$\mathbf{P}(|X^T AX - \text{trace}(A)| \geq t) \leq C\exp(-C' \min\{\frac{t^2}{\|A\|_F^2}, \frac{t}{\|B\|_2}\}) \tag{10}$$

for any $t > 0$.

Later, Wright [26] extended Theorem 7 to non-symmetric random variables. Recently, Hsu, Kakade and Zhang [13] showed that one can obtain a better upper tail (notice that $\|B\|_2$ is replaced by $\|A\|_2$)

$$\mathbf{P}(X^T AX - \text{trace}(A) \geq t) \leq C\exp(-C' \min\{\frac{t^2}{\|A\|_F^2}, \frac{t}{\|A\|_2}\}) \tag{11}$$

under a weaker assumption. On the other hand, their method does not cover the lower tail. Let us pause here to point out a strong distinction from the linear case and the quadratic case: In the linear case (Chernoff type bounds), the lower tail follows from the upper tail by simply switching ξ_i to $-\xi_i$, but this trick is useless in the quadratic case.

In the previous papers, the random variables ξ_i are required to be real. Few years ago, motivated by the delocalization problem for random matrices, Erdös, Schlein and Yau [9] considered the complex case. By assuming either both the real and imaginary parts of x_i are iid sub-gaussian or the distribution of x_i is rotationally symmetric, they proved

$$\mathbf{P}(|X^* AX - \text{trace}(A)| \geq t) \leq C\exp(-C' \frac{t}{\|A\|_F}). \tag{12}$$

Later, Erdös, Yau and Yin [11] showed that if ξ_i are independent sub-exponential random variables with exponent $\alpha > 0$, having mean 0 and variance 1, then

$$\mathbf{P}(|X^* AX - \text{trace}(A)| \geq t) \leq C\exp(-C' (\frac{t}{\|A\|_F})^{\frac{1}{2+2\alpha}}). \tag{13}$$

Using Lemma 4, we will prove the following result

Theorem 8. *Let X be a K-concentrated random vector in $\mathbb{C}^n$ whose entries have mean 0 and variance 1. Then there are constants $C, C' > 0$ such that for any matrix A*

$$\mathbf{P}(|X^* AX - \text{trace}(A)| \geq t) \leq C\log n\exp(-C' K^{-2} \min\{\frac{t^2}{\|A\|_F^2 \log n}, \frac{t}{\|A\|_2}\}). \tag{14}$$

To simplify the comparison with other results, let us ignore the $\log n$ terms (which play little role in practice). If $K = O(1)$, then the main difference between Theorem 7 of Hanson and Wright and Theorem **??** is that the term $\|B\|_2$ in Theorem 7 is now replaced by $\|A\|_2$. It is easy to see that $\|B\|_2 \geq \|A\|_2$ for any real matrix A. In fact, in many cases, $\|B\|_2$ is significantly larger than $\|A\|_2$. For instance, a random matrix A with entries of order 1 typically has spectral norm of order $\sqrt{n}$, but in this case it is clear that $\|B\|$ has spectral norm of order n (as all row sums are of this order). The same holds for several classical explicit matrices, such as the Hadamard matrix. In these cases, our bound improves Hanson-Wright's significantly. Furthermore, our result applies in the complex case while it seems that the approach used by Hanson and Wright is restricted to the real case.

Comparing to (12), we do not need the fairly restricted assumption that either both the real and imaginary parts of x_i are iid sub-gaussian or the distribution of x_i is rotationally symmetric. In the case $K = O(1)$, both terms $\frac{t^2}{\|A\|_F^2 \log n}$ and $\frac{t}{\|A\|_2}$ in our bound can be considerably larger than $\frac{t}{\|A\|_F}$. For instance, $\frac{t}{\|A\|_2}$ and $\frac{t}{\|A\|_F}$ differ by a factor $\sqrt{n}$ in both the random and Hadamard cases.

Next, we make use of Lemma 5. We keep the parameters $\varepsilon_1, \varepsilon_2, \varepsilon_3$ as defined in this lemma.

Theorem 9. *There are constants $C, C' > 0$ such that the following holds. Assume $n^2K^2(\epsilon_2+\epsilon_3)) = o(1)$, then*

$$\mathbf{P}(|X^*AX - \mathrm{trace}(A)| \geq t) \leq C \log n \exp(-C'K^{-2} \min\{\frac{t^2}{\|A\|_F^2 \log n}, \frac{t}{\|A\|_2}\}) + n\epsilon_1.$$

As an illustration, let us consider the case when ξ_i are sub-exponential with exponent $\alpha > 0$ (with accompanying constants a and b). We obtain an analogue of (**??**)

$$\mathbf{P}(|x^*Ax - \mathrm{trace}(A)| \geq t) \leq C \exp(-C'K^{-2} \min\{\frac{t^2}{\|A\|_F^2 \log n}, \frac{t}{\|A\|_2}\}) + n \exp(-\frac{b}{2} K^{1/\alpha}), \tag{15}$$

under the assumption that $K = \omega(\log^\alpha n)$.

To optimize the bound, we choose K such that $K^{-2} \min\{\frac{t^2}{\|A\|_F^2 \log n}, \frac{t}{\|A\|_2}\}) = K^{1/\alpha}$. This leads to setting $K := \min\{(\frac{t}{\|A\|_F \sqrt{\log n}})^{\frac{2}{2+1/\alpha}}, (\frac{t}{\|A\|_2})^{\frac{1}{2+1/\alpha}}\}$. Assume

$$t = \omega(\log^{\alpha+1} n(\|A\|_F + \log^{\alpha n} \|A\|_2).$$

This assumption guarantees $K = \omega(\log^\alpha n)$. It also implies $n \exp(-\frac{b}{2} K^{1/\alpha}) \leq \exp(-\frac{b}{3} K^{1/\alpha}$. It follows that

Corollary 10. *Assume that ξ_i are independent sub-exponential with exponent $\alpha > 0$ with mean 0 and variance 1. Then there are constants $C, C' > 0$ such that for any $t = \omega(\log^{\alpha+1} n(\|A\|_F + \log^{\alpha n} \|A\|_2)$*

$$\mathbf{P}(|x^*Ax - \mathrm{trace}(A)|| \geq t) \leq C \exp(-C' \min\{(\frac{t}{\|A_F\| \sqrt{\log n}})^{\frac{1}{\alpha+1/2}}, (\frac{t}{\|A\|_2})^{\frac{1}{2\alpha+1}}\}. \tag{16}$$

Notice that in order to make a Hanson-Wright type bound non-trivial, we need to assume $t \geq (\|A\|_F + \|A\|_2)$. In many applications, we want the probability bound to be polynomially or even super-polynomially small. This requires a lower bound $\log^{\Omega(1)} n(\|A\|_F + \|A\|_2)$ on t, which is consistent with the assumption in the corollary.

(16) compares favorably to (13). For the term $\frac{t}{\|A\|_F}$, the exponent $\frac{1}{\alpha+1/2}$ is superior to $\frac{1}{2\alpha+2}$ (notice that we are talking about a double exponent, so an improvement here could improve the quality of the bound quite a lot).

For the term $\frac{t}{\|A\|_2}$, the exponent $\frac{1}{2\alpha+1}$ is still better than $\frac{1}{2\alpha+2}$. Furthermore, $\|A\|_2$ can be significantly smaller than $\|A\|_F$, as discussed above.

3. Norm of random eigenvectors

Let M_n be a symmetric ± 1 matrix (the upper diagonal entries are iid Bernoulli random variables taking values ± 1 with probability 1/2). This is an important object in both probabilistic combinatorics and the theory of random matrices. Let u be an arbitrary eigenvector of M_n (of unit lenght). We would like to study the natural question, raised by Dekel et. al. [5]

How big is $\|u\|_\infty$ *?*

A good bound on the infinity norm of the eigenvectors plays a critical in spectral analysis and many other applications, such as the studies of nodal domains (see for instance [5] and the references therein). Recently, it also plays a crucial role in the study of local statistics of random matrices (see [24, 6] for surverys).

It is well known ([16]) that if we replace the entries of M_n by iid standard gaussian variables, then a random eigenvector distributes like a random vector v (with respect to the uniform distribution) from the unit sphere. Such a vector, with high probability, has norm $\Theta(\sqrt{\log n/n})$. It is natural to conjecture that the same bound holds in the Bernoulli case. As a matter of fact, several bounds of the type $n^{-1/2+\epsilon}$ or $n^{-1/2}\log^C n$ have been proved recently for various models of random matrices. They are usually referred to as delocalization results. The first such result was obtained by Erdös et. al. [9, Corollary 3.2] for a random matrix with entries having continuous distribution satisfying certain decay assumption, using (12). The next result, by Tao and the first author, handles the case when the entries are K-bounded (including the Bernoulli case), using the approach in [9] with Lemma 1 (see [22, Proposition 62]. Later, these results were extended to many other models (see [6, 24] for surveys). However, in all results, the constant C in the log term (if any) is either far from the conjectural value 1/2 or was not even determined (in these cases, if one tries to follow all steps to track down this value, the result would turn out to be rather disappointing).

As an application of the Weighted projection lemma, we are going to show that most of the eigenvectors of M_n have norm $O(\sqrt{\log n/n})$, with high probability. To explain what we mean by "most of", we first mention some basic facts about the eigenvalues. A corner stone of random matrix theory is the Wigner semi-circle law, which describes the limiting distribution of the eigenvalues. Denote by ρ_{sc} the semi-circle density function with support on $[-2, 2]$,

$$\rho_{sc}(x) := \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}, & |x| \leq 2 \\ 0, & |x| > 2. \end{cases} \tag{17}$$

Theorem 11 (Semi-circular law)**.** *Let M_n be a random Hermitian matrix whose entries on and above the diagonal are iid bounded random variables with zero mean and unit variance and $W_n = \frac{1}{\sqrt{n}} M_n$. Then for any real number x,*

$$\lim_{n\to\infty} \frac{1}{n} |\{1 \le i \le n : \lambda_i(W_n) \le x\}| = \int_{-2}^{x} \rho_{sc}(y)\ dy$$

in the sense of probability, where we use $|I|$ to denote the cardinality of a finite set I.

Remark 12. This is result is the famous Wigner's semi-circle law. It was first proved by Wigner for some class of random matrices and later extended to the general case above by many researchers; we refer to [17, 3] for detailed discussions. By Wigner's law, we expect most of the eigenvalues of W_n to lie in the interval $(-2+\varepsilon, 2+\varepsilon)$ for a fixed, small ϵ. Following random matrix literature, we refer to this region as the *bulk* of the spectrum.

Now we are ready to state our result

Theorem 13 (Optimal infinity norm of eigenvectors)**.** *For any constant $C_1 > 0$ there is a constant $C_2 > 0$ such that the following holds.*

- *(Bulk case) For any $\epsilon > 0$ and any $1 \le i \le n$ with $\lambda_i(W_n) \in [-2+\epsilon, 2-\epsilon]$, let $u_i(W_n)$ denote the corresponding unit eigenvector, then*
$$\|u_i(W_n)\|_\infty \le \frac{C_2 \log^{1/2} n}{\sqrt{n}}$$
with probability at least $1 - n^{-C_1}$.
- *(Edge case) For any $\epsilon > 0$ and any $1 \le i \le n$ with $\lambda_i(W_n) \in [-2-\epsilon, -2+\epsilon] \cup [2-\epsilon, 2+\epsilon]$, let $u_i(W_n)$ denote the corresponding unit eigenvector, then*
$$\|u_i(W_n)\|_\infty \le \frac{C_2 \log n}{\sqrt{n}}$$
with probability at least $1 - n^{-C_1}$.

By Wigner's semi-circle law, with probability $1 - O(\varepsilon^{3/2})$ a randomly selected eigenvector corresponds to an eigenvalue in the interval $[-2+\epsilon, 2-\epsilon]$. By letting ϵ tends to zero, we can conclude that with high probability, a randomly selected eigenvector u satisfies $\|u\|_\infty = O(\sqrt{\frac{\log n}{n}})$. It is an interesting open problem to reduce the $\log n$ term in the edge case to $\sqrt{\log n}$. If this holds, then all eigenvectors u satisfies the optimal bound $\|u\|_\infty = O(\sqrt{\frac{\log n}{n}})$.

For numerical simulation in Figure 1, we plot the cumulative distribution function of the (normalized) infinity norm of eigenvector v for symmetric random Bernoulli matrix, and compare it with the vector u chosen uniformly from the unit sphere.

This simulation suggests a tantalizing conjecture that $\|u\|_\infty$ and $\|v\|_\infty$, after a proper normalization, have the same distribution. However, this conjecture is beyond our reach at this moment.

We stated our result for random Bernoulli matrix since this is the most popular model in combinatorics. One can easily extend the result to the following more general setting. Let Z_i be the ith row vector of the matrix and X_i be the $n-1$ dimensional vector obtained from Z_i by deleting the ith (diagonal) entry.

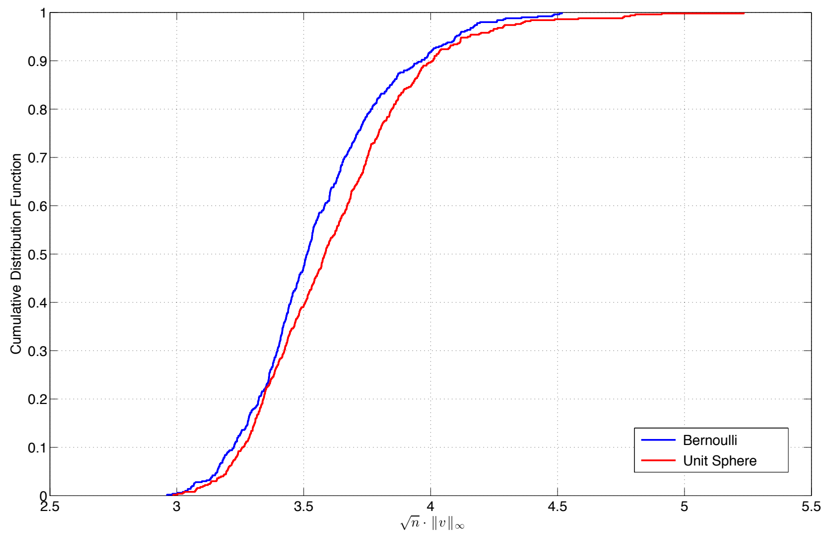

FIGURE 1. Plotted above are the empirical cumulative distribution functions of the distribution of $\sqrt{n} \cdot \|v\|_\infty$ for $n = 2000$, evaluated from 500 samples. In the blue curve, u is a unit eigenvector for symmetric random Bernoulli matrix. The red curve is generated for v to have a uniform distribution on the unit sphere S_n.

Theorem 14 (Optimal infinity norm of eigenvectors)**.** . *Let M_n be a Hermitian matrix whose upper diagonal entries are independent random variables with mean 0 and variance 1. Assume furthermore that for any index $1 \le i \le n$, X_i is K-concentrated. Then for any constant $C_1 > 0$ there is a constant $C_2 > 0$ such that the following holds*

- *(Bulk case) For any $\epsilon > 0$ and any $1 \le i \le n$ with $\lambda_i(W_n) \in [-2+\epsilon, 2-\epsilon]$, let $u_i(W_n)$ denote the corresponding unit eigenvector, then*
$$\|u_i(W_n)\|_\infty \le \frac{C_2 K \log^{1/2} n}{\sqrt{n}}$$
with probability at least $1 - n^{-C_1}$.
- *(Edge case) For any $\epsilon > 0$ and any $1 \le i \le n$ with $\lambda_i(W_n) \in [-2-\epsilon, -2+\epsilon] \cup [2-\epsilon, 2+\epsilon]$, let $u_i(W_n)$ denote the corresponding unit eigenvector, then*
$$\|u_i(W_n)\|_\infty \le \frac{C_2 K^2 \log n}{\sqrt{n}}$$
with probability at least $1 - n^{-C_1}$.

Since any K-concentrate implies K-bounded, Theorem 14 implies Theorem 13.

4. THE LOCAL SEMI-CIRCLE LAW

The key tool for bounding the infinity norm of a eigenvector is a statement of the following type: Any short interval in the spectrum contains an eigenvalue, with high probability. The quality of the bound will depend on how short the interval is. This approached was developed by Erdös, Schlein and Yau in [8, 7, 9], leading to the bounds $n^{-2/3}, n^{-3/4}$ and finally $n^{-1+o(1)}$. An argument of the same spirit was developed by Tao and the first author in [20] (see [20, Chapter 4] for a problem concerning random non-hermitian matrices.

The leading idea is that one expects the semi-circle law to hold for small intervals (or at small scale). Intuitively, we would like to have with high probability that

$$|N_I - n\int_I \rho_{sc}(x)\ dx| \leq \delta n|I|,$$

for any interval I and fixed $\delta > 0$, where N_I denotes the number of eigenvalues of $W_n := \frac{1}{\sqrt{n}}M_n$ on the interval I. Of course, the reader can easily see that I cannot be arbitrarily short (since N_I is an integer). Following [9], we call a statement of this kind a local semi-circle law (LSCL).

A natural question is: how short can I be ? Formally, we say that the LSCL holds at a scale $f(n)$ if with probability $1 - o(1)$

$$|N_I - n\int_I \rho_{sc}(x)\ dx| \leq \delta n|I|,$$

for any interval I in the bulk of length $\omega(f(n))$ and any fixed $\delta > 0$. Furthermore, we say that $f(n)$ is a *threshold scale* if the LSCL holds at scale $f(n)$ but does not holds at scale $g(n)$ for any function $g(n) = o(f(n))$. (The reader may notice a similarity between this definition and the definition of threshold functions for random graphs.) We would like to raise the following problem.

Problem 15. *Determine the threshold scale (if exists).*

A recent result by Ben Arous and Bourgart [1] shows that the maximum gap between two consecutive (bulk) eigenvalues of GUE (random matrix with complex gaussian entries) is of order $\Theta(\sqrt{\log n}/n)$, with high probability. Thus, if we partition the bulk into intervals of length $\alpha\sqrt{\log n}/n$ for a sufficiently small α, one of these intervals contains at most one eigenvalue. Thus, we expect that the LSCL do not hold below the $\sqrt{\log n}/n$ scale, at least for a large class of random matrices. In [9, 22], upper bound of the form $\log^C n/n$ was proved for some large value of C. Here we are going to show

Theorem 16. *Let M_n be a random matrix with K-concentrated entries. Then the threshold scale for LSCL bounded from above by $K^2 \log n/n$.*

Theorem 16, on the other hand, is a consequence of the following more quantitative statement.

Theorem 17. *For any constants $\epsilon, \delta, C_1 > 0$ there is a constant $C_2 > 0$ such that the following holds. Let M_n be a Hermitian random matrix whose upper diagonal entries are independent K-concentrated random variables with mean 0 and variance 1. Then with probability at least $1 - n^{-C_1}$, we have*

$$|N_I - n\int_I \rho_{sc}(x)\ dx| \leq \delta n \int_I \rho_{sc}(x)\ dx,$$

for all interval $I \subset (-2+\epsilon, 2-\epsilon)$ of length at least $C_2K^2 \log n/n$.

By Theorem 17, we now know (at least for random matrices with bounded entries) that the right scale is $\log n/n$. We can now formulate a sharp threshold question. Let us fix δ and δ'. Then for each n, let C_n be the infimum of those C such that with probability $1 - \delta'$

$$|N_I - n\int_I \rho_{sc}(x)\ dx| \leq \delta n|I|$$

holds for any $I, |I| \geq C\log n/n$. Is it true that $\lim_{n\to\infty} C_n$ exist? If so, can we compute its value as a function of δ and δ'?

5. Proof of Lemma 4

Set $f(X) := \sqrt{\sum_{j=1}^d c_j |u_j^* X|^2}$. Thus, f is a function from $\mathbb{C}^n$ to $\mathbb{R}$.

We first observe that $f(X)$ is convex. Indeed, for $0 \leq \lambda, \mu \leq 1$ where $\lambda+\mu = 1$ and any $X, Y \in \mathbb{C}^n$, by Cauchy-Schwardz inequality,

$$\begin{aligned} f(\lambda X + \mu Y) &\leq \sqrt{\sum_{j=1}^d c_j(\lambda |u_j^* X| + \mu |u_j^* Y|)^2} \\ &\leq \lambda \sqrt{\sum_{j=1}^d c_j |u_j^* X|^2} + \mu \sqrt{\sum_{j=1}^d c_j |u_j^* Y|^2} = \lambda f(X) + \mu f(Y) \end{aligned}$$

Next, we show that $f(X)$ is *1-Lipschitz*. Noticed that $f(X) \leq \sqrt{\sum_{j=1}^d |u_j^* X|^2} \leq \|X\|$. Since $f(X)$ is convex, one has

$$\frac{1}{2}f(X) = f(\frac{1}{2}X) = f(\frac{1}{2}(X-Y) + \frac{1}{2}Y) \leq \frac{1}{2}f(X-Y) + \frac{1}{2}f(Y).$$

Thus $f(X) - f(Y) \leq f(X-Y)$ and $f(Y) - f(X) \leq f(Y-X) = f(X-Y)$, which implies

$$|f(X) - f(Y)| \leq f(X-Y) \leq \|X - Y\|.$$

Thus, by the definition of the K-concentrated property,

$$\mathbf{P}(|f(X) - M(f(X))| \geq t) \leq C\exp(-C'\frac{t^2}{K^2}), \tag{18}$$

for some constants $C, C' > 0$.

To conclude the proof, it suffices to show $|M(f(X)) - \sqrt{\sum_{j=1}^d c_j}| = O(K)$.

Lemma 18. *Let Y be a real random variable. Assume $\mathbf{P}(|Y-\mu| \geq t) \leq f(t)$, where $\int_0^\infty f(x)dx = O(1)$. Then $|\mathbf{E}Y - \mu| = O(1)$.*

Assume further more that Y is non-negative, $\mathbf{E}Y^2 = \sigma^2$, $\mu + \sigma = \Omega(1)$ and $\int_0^\infty xf(x) = O(1)$, then $|\mathbf{E}Y - \sigma| = O(1)$.

Proof. By symmetry, it suffices to show that $\mathbf{E}(Y) \leq \mu + O(1)$. Notice that for any variable Y and any number L

$$\mathbf{E}Y \leq L + \int_0^\infty \mathbf{P}(Y \geq L + x)dx.$$

Taking $L = \mu$, the desired bound follows from the assumption on $f(x)$.

To prove the second part, notice that

$$\mathbf{E}Y^2 \leq L^2 + \int_0^\infty x\mathbf{P}(Y \geq L + x)dx.$$

Taking $L = \mu$, and use the second assumption on $f(x)$, we have

$$\sigma^2 \leq \mu^2 + O(1).$$

As $\sigma + \mu = \Omega(1)$ we can conclude that $\sigma \leq \mu + O(1) = \mathbf{E}Y + O(1)$. Furthermore, by convexity $\sigma \geq \mathbf{E}Y$, concluding the proof. $\square$

To apply this lemma, set $c_i' := \frac{c_i}{\max_{1\leq i\leq n} c_i}$, $Y := \frac{1}{K}\sqrt{\sum_{i=1}^n c_i'|u_i^* X|^2}$ and $\mu := M(Y)$. We have, by the K-concentration property

$$\mathbf{P}(|Y - \mu| \geq t) = \mathbf{P}(|\sqrt{\sum_{i=1}^n c_i'|u_i^* X|^2} - M(\sqrt{\sum_{i=1}^n c_i'|u_i^* X|^2})| \geq tK) \leq C\exp(-C't^2).$$

Set $f(x) = C\exp(-C'x^2)$. The assumptions on $f(x)$ in Lemma 18 are trivially satisfied. Since $\mathbf{E}Y^2 = \sigma^2 = \sum_{i=1}^n c_i' \geq 1$, it follows from Lemma 18 that

$$M(Y) = \sqrt{\sum_{i=1}^n c_i'} + O(1).$$

Renormalizing, we obtain

$$M(\sqrt{\sum_{i=1}^n c_i|u_i^* X|^2}) = \sqrt{\sum_{i=1}^n c_i + O(K\sqrt{\max_{1\leq i\leq n} c_i})},$$

which concludes the proof of Lemma 4.

6. Proof of Theorem 8

Notice that if $Y = X^*AX$, then $Y + \bar{Y} = X^*(A + A^*)X$ and $Y - \bar{Y} = X^*(A - A^*)X$. Since

$$Y - \operatorname{trace} A = \frac{1}{2}\big([(Y + \bar{Y}) - (\operatorname{trace}(A + A^*)] + \sqrt{-1}[(Y - \bar{Y}) - \operatorname{trace}(A - A^*)],$$

we have

$$\mathbf{P}(|Y - \text{trace}\, A| \geq t) \leq \mathbf{P}(|(Y + \bar{Y}) - (\text{trace}(A + A^*)| \geq t) + \mathbf{P}(|(Y - \bar{Y}) - \text{trace}(A - A^*)| \geq t).$$

Moreover, as $\|A + A^*\|_F, \|A - A^*\|_F = O(\|A\|_F)$ and $\|A + A^*\|_2, \|A - A^*\|_2 = O(\|A\|_2)$, it suffices to prove the theorem in the case A is Hermitian.

Next, we observe that any Hermitian matrix A can be written as $A := A_1 - A_2$ where A_i are positive semi-definite and $\max_i \|A_i\| \leq \|A\|, \max_i \|A_i\|_F \leq \|A\|_F$. (In fact, the positive eigenvalues of A_1 are the positive eigenvalues of A and the positive eigenvalues of A_2 are the absolute values of the negative eigenvalues of A.) This enables us to further reduce the problem to the case when A is positive semi-definite.

Finally, as the content of the theorem is invariant under scaling, we can assume that $\|A\| = 1$. Let $c_1 = 1, 1 \geq c_2, \ldots, c_n \geq 0$ be the eigenvalues of A together with corresponding eigenvectors $\{u_1, \ldots, u_n\}$, we have

$$X^*AX - \text{trace}(A) = \sum_{j=1}^{n} c_j |u_j^* x|^2 - \sum_{j=1}^{n} c_j. \tag{19}$$

This is precisely the setting of the projection lemmas. Using Lemma 4 together with (4), we know that for any numbers $0 \leq d_j \leq 1$, $j \in J$,

$$\mathbf{P}(|\sum_{j \in J} d_j |u_j^* X|^2 - \sum_{j \in J} d_j| \geq 2t\sqrt{\sum_i d_i} + t^2) \leq C\exp(-C' K^{-2} t^2). \tag{20}$$

However, it is wasteful to apply this to (19). We will need an extra partition step. Set

$$J_k := \{1 \leq j \leq n : \frac{1}{4^{k+1}} \leq c_j \leq \frac{1}{4^k}\}, 0 \leq k \leq k_0 := 10 \log n,$$

and let J_{k_0+1} be the collection of the remaining indices.

For each $0 \leq k \leq k_0 + 1$, apply Lemma 2 to $d_i := 4^k c_i, c_i \in J_k$, we have, for any $s \geq 0$

$$\mathbf{P}(|\sum_{i \in J_k} 4^k (c_i |u_i \cdot X|^2 - 1)| \geq 2s\sqrt{\sum_{i \in J_k} 4^k c_i} + s^2) \leq C\exp(-C' K^{-2} s^2).$$

Set $s := \frac{t}{\|A\|_F}$ and simplify by 4^k, the above inequality becomes

$$\mathbf{P}(|(c_i |u_i \cdot X|^2 - 1)| \geq \frac{2t}{2^k \|A\|_F}\sqrt{\sum_{i \in J_k} c_i} + \frac{t^2}{4^k \|A_F\|^2}) \leq C\exp(-C' K^{-2} \frac{t^2}{\|A\|_F^2}).$$

Apparently, $\sum_{k=0}^{k_0+1} \frac{t^2}{4^k \|A_F\|^2} \leq 2\frac{t^2}{|A_F|^2}$. Moreover, $\sum_{i \in J_{k_0+1}} c_i \leq n \times n^{-5} = n^{-4}$ and

$$\begin{aligned}\sum_{0\le k\le k_0} 2^{-k}\sqrt{\sum_{i\in J_k} c_i} &\le k_0^{1/2}(\sum_{k=0}^{k_0} 4^{-k}\sum_{i\in J_k} c_i)^{1/2}\\ &\le 4\log^{1/2} n(\sum_{k=0}^{k_0}\sum_{i\in J_k} c_i^2)^{1/2}\\ &\le 4\log^{1/2} n\|A\|_F,\end{aligned}$$

by Cauchy-Schwartz.

Putting the above estimates together and using the union bound, we obtain

$$\mathbf{P}(|\sum_{i=1}^n c_i(|u_i^* X|^2-1)|\ge 4\log^{1/2} n\|A\|_F t+2\frac{t^2}{\|A_F\|^2}+n^{-2})\le C\log n\exp(-C'K^{-2}\frac{t^2}{\|A\|_F^2}).$$

We can ignore the small term n^{-2}. Reset $t := 5\log^{1/2} n\|A\|_F t+2\frac{t^2}{\|A_F\|^2}$, the desired bound follows.

Remark 19. If we have more information about A, the extra $\log n$ term can be improved. For instance of all eigenvalues of A are comparable, then one can remove this term.

The proof of Theorem **??** is left as an exercise.

7. Random matrices and the Stieltjes transform

In this section, we recall some facts about random matrices. The empirical spectral distribution (ESD) of the $n\times n$ Hermitian matrix $W_n := \frac{1}{\sqrt{n}}M_n$ is a one-dimensional function

$$F^{\mathbf{W_n}}(x)=\frac{1}{n}|\{1\le j\le n:\lambda_j(W)\le x\}|,$$

where $|\mathbf{I}|$ denotes the cardinality of a set $\mathbf{I}$. We are going to focus on the case when the entries of M_n are K-bounded; it is easy to extend this assumption to K-concentrated (see Remark 24).

The Stieltjes transform of a real measure $\mu(x)$ is defined for any complex number z not in the support of μ as

$$s(z)=\int_{\mathbb{R}}\frac{1}{x-z}d\mu(x).$$

Thus, the Stieltjes transform $s_n(z)$ of W_n is

$$s_n(z)=\int_{\mathbb{R}}\frac{1}{x-z}dF^{\mathbf{W_n}}(x)=\frac{1}{n}\sum_{i=1}^n\frac{1}{\lambda_i(W_n)-z}.$$

Furthermore, the Stieltjes transform $s(z)$ of the semi-circle distribution is

$$s(z) := \int_{\mathbb{R}} \frac{\rho_{sc}(x)}{x-z} dx = \frac{-z + \sqrt{z^2-4}}{2},$$

where $\sqrt{z^2-4}$ is the branch of square root with a branch cute in $[-2,2]$ and asymptotically equals z at infinity [3].

The beauty (and power) of the Stieltjes transform is the fact that it has clear a linear algebra content; $s_n(z)$ of W_n is exactly the trace of the matrix $(W_n - zI)^{-1}$. This allows us to compute the Stieltjes transform by looking at the diagonal entries of $(W_n - zI)^{-1}$. In matrix theory, Stieltjes transform plays the role Fourier transform in analysis. If the Stieltjes transforms of two spectral measures are close to each other (for all z), then the two measures are more or less the same. In particular, if $s_n(z)$ is close to $s(z)$, then the spectral distribution of W_n is close to the semi-circle distribution (see for instance [3, Chapter 11], [8]). We are going to use the following lemma.

Lemma 20. *Let M_n be a random Hermitian matrix with independent K-bounded entries with mean 0 and variance 1. Let $1/n < \eta < 1/10$ and $L, \varepsilon, \delta > 0$. For any constant $C_1 > 0$, there exists a constant $C > 0$ such that if one has the bound*

$$|s_n(z) - s(z)| \le \delta$$

with probability at least $1 - n^{-C}$ uniformly for all z with $|Re(z)| \le L$ and $Im(z) \ge \eta$, then for any interval I in $[-L+\varepsilon, L-\varepsilon]$ with $|I| \ge max(2\eta, \frac{\eta}{\delta}\log\frac{1}{\delta})$, one has

$$|N_I - n\int_I \rho_{sc}(x)\,dx| \le \delta n |I|$$

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

This is [22, Lemma 64], which, in turn, is a variant of [8, Corollary 4.3]).

An appropriate application of Lemma 20 will imply Theorem **??**. In order to use this lemma, we set $L = 4, \varepsilon = 1$, and critically

$$\eta := \frac{K^2 C^2 \log n}{n\delta^6},$$

where $C = C_1 + 10^4$. We are going to show that

$$|s_n(z) - s(z)| = o(\delta) \tag{21}$$

holds with probability at least $1 - n^{-C}$ for any fixed z in the region $\{z \in \mathbb{C} : |\mathrm{Re}(z)| \le 4, \mathrm{Im}(z) \ge \eta\}$. Notice that in this statement we fix z. However, it is simple to strengthen the statement to hold for all z, using an ϵ-net argument, exploiting the fact that $s_n(z)$ is Lipschitz continuous with the Lipschitz constant $O(n^2)$ (for details, we refer to [7, Theorem 1.1] or [22]; **Ke: find the exact part of this paper**).

In order to show that $s_n(z)$ is close to $s_{sc}(z)$, the key observation is that $s_{sc}(z)$ can also be defined by the equation

$$s(z) = -\frac{1}{z + s(z)}. \tag{22}$$

This equation is stable, so if we can show $s_n(z) \approx -\frac{1}{z+s_n(z)}$ then it follows that $s_n(z) \approx s_{sc}(z)$. This observation was due to Bai et. al. [], who used it to prove the $n^{-1/2}$ rate of convergence of $s_n(z)$ to $s_{sc}(z)$. In [8, 7, 9], Erdős et. al. refined Bai's approach to prove local semi-circle law at scales finer than $n^{-1/2}$, ultimately to $n^{-1}\log^C n$ [8]. Our main contribution here is to push the scale further down to $n^{-1}\log n$, which we believe is (at most) a factor $\sqrt{\log n}$ from the truth.

Recall that $s_n(z)$ is the trace of $(W_n - zI)^{-1}$. By computing the diagonal entires, one can show (see [3, Chapter 11], [8], [22, Lemma 39])

$$s_n(z) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{-\frac{\zeta_{kk}}{\sqrt{n}} - z - Y_k} \tag{23}$$

where

$$Y_k = a_k^*(W_{n,k} - zI)^{-1}a_k,$$

and $W_{n,k}$ is the matrix W_n with the k^{th} row and column removed, and a_k is the k^{th} row of W_n with the k^{th} element removed.

The entries of a_k are independent of each other and of $W_{n,k}$, and have mean zero and variance $1/n$. By linearity of expectation we have

$$\mathbf{E}(Y_k|W_{n,k}) = \frac{1}{n}\text{Trace}(W_{n,k} - zI)^{-1} = (1 - \frac{1}{n})s_{n,k}(z)$$

where

$$s_{n,k}(z) = \frac{1}{n-1}\sum_{i=1}^{n-1} \frac{1}{\lambda_i(W_{n,k}) - z}$$

is the *Stieltjes transform* of $W_{n,k}$. From the Cauchy interlacing law, we can get

$$|s_n(z) - (1 - \frac{1}{n})s_{n,k}(z)| = O(\frac{1}{n}\int_{\mathbb{R}} \frac{1}{|x - z|^2}\, dx) = O(\frac{1}{n\eta}) = o(\delta^2)$$

and thus

$$\mathbf{E}(Y_k|W_{n,k}) = s_n(z) + o(\delta^2).$$

The heart of the matter now is the following concentration result

Lemma 21. *Let M_n be as in Lemma 20. For $1 \le k \le n$, $Y_k = \mathbf{E}(Y_k|W_{n,k}) + o(\delta^2)$ holds with probability at least $1 - O(n^{-C})$ for any z with $|Re(z)| \le 4$ and $Im(z) \ge \eta$.*

To prove this lemma, we are going to make an essential use of Projection Lemma 2.

8. Proofs of Lemma 21 and Theorem ??

First, we record a lemma that provides a crude upper bound on the number of eigenvalues in short intervals.

Lemma 22. *Let M_n be a random Hermitian matrix with independent K-bounded entries with mean 0 and variance 1. For any constant $C_1 > 0$, there exists a constant $C_2 > 0$ such that for any interval $I \subset \mathbb{R}$ with $|I| \ge \frac{C_2K^2\log n}{n}$,*

$$N_I \ll n|I|$$

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

This Lemma is Proposition 66 in [22], which is a variant of [9, Theorem 5.1]. Notice that

$$Y_k = a_k^*(W_{n,k} - zI)^{-1}a_k = \sum_{j=1}^{n-1} \frac{|u_j(W_{n,k})^* a_k|^2}{\lambda_j(W_{n,k}) - z}. \tag{24}$$

Therefore,

$$|Y_k - \mathbf{E}(Y_k|W_{n,k})| = \frac{1}{n}|\sum_{j=1}^{n-1} \frac{|u_j(W_{n,k})^* X_k|^2 - 1}{\lambda_j(W_{n,k}) - z}| = \frac{1}{n}|\sum_{j=1}^{n-1} \frac{R_j}{\lambda_j(W_{n,k}) - x - \sqrt{-1}\eta}|, \tag{25}$$

where $R_j := |u_j(W_{n,k})^* X_k|^2 - 1$. By symmetry, we can restrict the sum to those indices j where $\lambda_j(W_{n,k}) - x \geq 0$.

Let J be the set of indices j such that $0 \leq \lambda_j(W_{n,k}) - x \leq \eta$. Since $x = \text{Re}z, \eta = \text{Im}z$, we have

$$\begin{aligned} &\tfrac{1}{n}|\textstyle\sum_{j\in J} \frac{R_j}{\lambda_j(W_{n,k}) - x - \sqrt{-1}\eta}| \\ &\leq \tfrac{1}{n}|\textstyle\sum_{j\in J} \frac{\lambda_j(W_{n,k}) - x}{(\lambda_j(W_{n,k}) - x)^2 + \eta^2} R_j| + \tfrac{1}{n}|\textstyle\sum_{j\in J} \frac{\eta}{(\lambda_j(W_{n,k}) - x)^2 + \eta^2} R_j| \\ &\leq \tfrac{1}{n\eta}|\textstyle\sum_{j\in J} \frac{(\lambda_j(W_{n,k}) - x)\eta}{(\lambda_j(W_{n,k}) - x)^2 + \eta^2} R_j| + \tfrac{1}{n\eta}|\textstyle\sum_{j\in J} \frac{\eta^2}{(\lambda_j(W_{n,k}) - x)^2 + \eta^2} R_j|. \end{aligned}$$

Consider the sum $S_1 := |\sum_{j\in J} \frac{(\lambda_j(W_{n,k}) - x)\eta}{(\lambda_j(W_{n,k}) - x)^2 + \eta^2} R_j|$. As $0 \leq \frac{(\lambda_j(W_{n,k}) - x)\eta}{(\lambda_j(W_{n,k}) - x)^2 + \eta^2} \leq 1$, we are in position to apply Lemma 2. Taking $t = C_4 K\sqrt{\log n}$ with a sufficiently large constant C_4, by (4) we have

$$S_1 \leq \frac{C_4}{n\eta}(\sqrt{|J|\log n} + K^2 \log n)$$

with probability at least $1 - C\exp(-C'C_4 \log n) \geq 1 - n^{-C_4/2}$. By Lemma 22, $|J| \leq Bn\eta$ with probability at least $1 - n^{-C_4}$, for some sufficiently large constant $B > 0$. Recall $\eta := \frac{K^2 C_3^2 \log n}{n\delta^6}$; it follows that with probability at least $1 - 2n^{-C_4/2}$ we have

$$S_1 \leq C_4 C_3^{-2} B\delta^6 \log n.$$

Thus, for C_3 sufficiently large compared to C_4 and B, then $S_1 \leq \delta^3$. Similarly, we can prove the same bound for $S_2 := \frac{1}{n\eta}|\sum_{j\in J} \frac{\eta^2}{(\lambda_j(W_{n,k}) - x)^2 + \eta^2} R_j|$.

For the other eigenvalues, we divide the real line into small intervals. For integer $l \geq 0$, let J_l be the set of eigenvalues $\lambda_j(W_{n,k})$ such that $(1+\alpha)^l \eta < \lambda_j(W_{n,k}) - x \leq (1+\alpha)^{l+1}\eta$. We use the parameters $a = (1+\alpha)^l \eta$ and $\alpha = 10$ (say). The number of such J_l is $O(\log n)$. By Lemma 22 one has, $|J_l| \ll na\alpha$. Again by Lemma 2 (take $t = K\sqrt{C(l+1)}\sqrt{\log n}$),

$$\begin{aligned} &\frac{1}{n}|\sum_{j\in J_l}\frac{R_j}{\lambda_j(W_{n,k})-x-\sqrt{-1}\eta}| \\ &\leq \frac{1}{n}|\sum_{j\in J_l}\frac{\lambda_j-x}{(\lambda_j-x)^2+\eta^2}R_j| + \frac{1}{n}|\sum_{j\in J_l}\frac{\eta}{(\lambda_j-x)^2+\eta^2}R_j| \\ &\leq \frac{1+\alpha}{na}|\sum_{j\in J_l}\frac{a(\lambda_j-x)}{(1+\alpha)((\lambda_j-x)^2+\eta^2)}R_j| + \frac{\eta}{na^2}|\sum_{j\in J}\frac{a^2}{(\lambda_j-x)^2+\eta^2}R_j| \\ &\leq (\frac{1+\alpha}{na}+\frac{\eta}{na^2})(K\sqrt{C(l+1)}\sqrt{\log n}\sqrt{n\alpha a}+K^2C(l+1)\log n) \\ &\leq \frac{20\delta^3}{\sqrt{C}}\frac{l+1}{(1+\alpha)^{l/2}} \end{aligned}$$

with probability at least $1-10n^{-C(l+1)}$.

Summing over l, we have

$$\frac{1}{n}|\sum_l\sum_{j\in J_l}\frac{R_j}{\lambda_j(W_{n,k})-x-\sqrt{-1}\eta}| \leq \frac{40\delta^3}{\sqrt{C}} = o(\delta^2),$$

with probability at least $1-10n^{-C}$. This completes the proof of Proposition 21.

Inserting the bounds into (23), one has

$$s_n(z) + \frac{1}{n}\sum_{k=1}^{n}\frac{1}{s_n(z)+z+o(\delta^2)} = 0$$

with probability at least $1-10n^{-C}$. The term $|\zeta_{kk}/\sqrt{n}| = o(\delta^2)$ as $|\zeta_{kk}| \leq K$ by assumption. Comparing this equation with (22), one can use a continuity argument (see [21] for details) to obtain $|s_n(z)-s(z)| \leq \delta$ with probability at least $1-n^{-C+100}$.

Applying Lemma 20, we have

Theorem 23. *For any constant $C_1 > 0$, there exists a constant $C_2 > 0$ such that for $0 \leq \delta \leq 1/2$ any interval $I \subset (-3,3)$ of length at least $C_2K^2\log n/n\delta^8$,*

$$|N_I - n\int_I \rho_{sc}(x)\,dx| \leq \delta n|I|$$

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

In particular, Theorem 17 follows.

Remark 24. The only (nominal) difference is that we replace the K-bounded assumption in [22] by K-concentrated. Since in [22], we only used the K-bounded to guarantee K-concentration, the proof remains the same.

9. Proof of Theorem 14

With the concentration theorem for ESD, we are able to derive the eigenvector delocalization results thanks to the next lemma:

Lemma 25 (Eq (4.3), [7] or Lemma 41, [22]). *Let*

$$B_n = \begin{pmatrix} a & X^* \\ X & B_{n-1} \end{pmatrix}$$

be an $n \times n$ symmetric matrix for some $a \in \mathbb{C}$ and $X \in \mathbb{C}^{n-1}$, and let $\begin{pmatrix} x \\ v \end{pmatrix}$ be an eigenvector of B_n with eigenvalue $\lambda_i(B_n)$, where $x \in \mathbb{C}$ and $v \in \mathbb{C}^{n-1}$. Suppose that none of the eigenvalues of B_{n-1} are equal to $\lambda_i(B_n)$. Then

$$|x|^2 = \frac{1}{1 + \sum_{j=1}^{n-1} (\lambda_j(B_{n-1}) - \lambda_i(B_n))^{-2} |u_j(B_{n-1})^* X|^2},$$

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

First, for the bulk case, for any $\lambda_i(W_n) \in (-2+\varepsilon, 2-\varepsilon)$, by Theorem 17, one can find an interval $I \subset (-2+\varepsilon, 2-\varepsilon)$, centered at $\lambda_i(W_n)$ and $|I| = K^2 C \log n/n$, such that $N_I \geq \delta_1 n |I|$ ($\delta_1 > 0$ small enough) with probability at least $1 - n^{-C_1 - 10}$. By Cauchy interlacing law, we can find a set $J \subset \{1, \ldots, n-1\}$ with $|J| \geq N_I/2$ such that $|\lambda_j(W_{n-1}) - \lambda_i(W_n)| \leq |I|$ for all $j \in J$.

By Lemma 25, we have

$$\begin{aligned} |x|^2 &= \frac{1}{1 + \sum_{j=1}^{n-1} (\lambda_j(W_{n-1}) - \lambda_i(W_n))^{-2} |u_j(W_{n-1})^* \frac{1}{\sqrt{n}} X|^2} \\ &\leq \frac{1}{1 + \sum_{j \in J} (\lambda_j(W_{n-1}) - \lambda_i(W_n))^{-2} |u_j(W_{n-1})^* \frac{1}{\sqrt{n}} X|^2} \\ &\leq \frac{1}{1 + n^{-1} |I|^{-2} \sum_{j \in J} |u_j(W_{n-1})^* X|^2} \\ &\leq \frac{1}{1 + 100^{-1} n^{-1} |I|^{-2} |J|} \leq 200 |I|/\delta_1 \leq \frac{K^2 C_2^2 \log n}{n} \end{aligned} \tag{26}$$

for some constant C_2 with probability at least $1 - n^{-C_1 - 10}$. The third inequality follows from Lemma 2 by taking $t = \delta_1 K \sqrt{C} \log n/\sqrt{n}$ (say).

Thus, by union bound and symmetry, $\|u_i(W_n)\|_\infty \leq \frac{C_2 K \log^{1/2} n}{\sqrt{n}}$ holds with probability at least $1 - n^{-C_1}$.

Appendix A. Proof for the Edge case of Theorem 14

For the edge case in Theorem 14, we use a different approach based on the next lemma:

Lemma 26 (Interlacing identity, Lemma 37, [21]). *If $u_j(W_{n-1})^* Y$ is non-zero for every j, then*

$$\sum_{j=1}^{n-1} \frac{|u_j(W_{n-1})^* Y|^2}{\lambda_j(W_{n-1}) - \lambda_i(W_n)} = \frac{1}{\sqrt{n}} \zeta_{nn} - \lambda_i(W_n). \tag{27}$$

By symmetry, it suffices to consider the case $\lambda_i(W_n) \in [2-\epsilon, 2+\epsilon]$ for $\epsilon > 0$ small. By Lemma 25, in order to show $|x|^2 \leq C^4 K^4 \log^2 n/n$ (for some constant $C > C_1 + 100$) with a high probability, it is enough to show

$$\sum_{j=1}^{n-1} \frac{|u_j(W_{n-1})^* X|^2}{(\lambda_j(M_{n-1}) - \lambda_i(M_n))^2} \geq \frac{n}{C^4 K^4 \log^2 n}.$$

By the projection lemma, $|u_j(W_{n-1})^* X| \leq K\sqrt{C \log n}$ with probability at least $1 - 10n^{-C}$. It suffices to show that with probability at least $1 - n^{-C_1 - 100}$,

$$\sum_{j=1}^{n-1} \frac{|u_j(W_{n-1})^* X|^4}{(\lambda_j(M_{n-1}) - \lambda_i(M_n))^2} \geq \frac{n}{C^3 K^2 \log n}.$$

Let $Y = \frac{1}{\sqrt{n}} X$, by Cauchy-Schwardz inequality, it is enough to show for some integers $1 \leq T_- < T_+ \leq n-1$ that

$$\sum_{T_- \leq j \leq T_+} \frac{|u_j(W_{n-1})^* Y|^2}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|} \geq \frac{\sqrt{T_+ - T_-}}{C^{1.5} K \sqrt{\log n}}.$$

And by Lemma 26, we are going to show for $T_+ - T_{-1} = O(\log n)$ (the choice of T_+, T_- will be given later) that

$$|\sum_{j \geq T_+ \text{or} j \leq T_-} \frac{|u_j(W_{n-1})^* Y|^2}{\lambda_j(W_{n-1}) - \lambda_i(W_n)}| \leq 2 - \epsilon - \frac{\sqrt{T_+ - T_-}}{C^{1.5} K \sqrt{\log n}} + o(1), \tag{28}$$

with probability at least $1 - n^{-C_1 - 100}$.

Now we divide the real line into disjoint intervals I_k for $k \geq 0$. Let $|I| = \frac{K^2 C \log n}{n\delta^8}$ with constant $\delta \leq \epsilon/1000$. Denote $\beta_k = \sum_{s=0}^{k} \delta^{-8s}$. Let $I_0 = (\lambda_i(W_n) - |I|, \lambda_i(W_n) + |I|)$. For $1 \leq k \leq k_0 = \log^{0.9} n$ (say),

$$I_k = (\lambda_i(W_n) - \beta_k |I|, \lambda_i(W_n) - \beta_{k-1}|I|] \cup [\lambda_i(W_n) + \beta_{k-1}|I|, \lambda_i(W_n) + \beta_k |I|),$$

thus $|I_k| = 2\delta^{-8k}|I| = o(1)$ and the distance from $\lambda_i(W_n)$ to the interval I_k satisfies

$$\text{dist}(\lambda_i(W_n), I_k) \geq \beta_{k-1}|I|.$$

For each such interval, by Theorem 17, the number of eigenvalues $|J_k| = N_{I_k} \leq n\alpha_{I_k}|I_k| + \delta^k n |I_k|$ with probability at least $1 - n^{-C_1 - 100}$, where $\alpha_{I_k} = \int_{I_k} \rho_{sc}(x)dx/|I_k|$.

By Lemma 2, for the kth interval, with probability at least $1 - n^{-C_1 - 100}$,

$$\begin{aligned}
&\frac{1}{n} \sum_{j \in J_k} \frac{|u_j(W_{n-1})^* X|^2}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|} \leq \frac{1}{n} \frac{1}{\text{dist}(\lambda_i(W_n), I_k)} \sum_{j \in J_k} |u_j(W_{n-1})^* X|^2 \\
&\leq \frac{1}{n} \frac{1}{\text{dist}(\lambda_i(W_n), I_k)} (|J_k| + K\sqrt{|J_k|}\sqrt{C \log n} + CK^2 \log n) \\
&\leq \frac{\alpha_{I_k}|I_k|}{\text{dist}(\lambda_i(W_n), I_k)} + \frac{\delta^k |I_k|}{\text{dist}(\lambda_i(W_n), I_k)} + \frac{CK^2 \log n}{n\text{dist}(\lambda_i(W_n), I_k)} \\
&+ \frac{K\sqrt{n\alpha_{I_k} + n\delta^k}\sqrt{|I_k|}\sqrt{C \log n}}{n\text{dist}(\lambda_i(W_n), I_k)} \\
&\leq \frac{\alpha_{I_k}|I_k|}{\text{dist}(\lambda_i(W_n), I_k)} + 2\delta^{k-16} + \delta^{8k-8} + \delta^{4k-15}.
\end{aligned}$$

For $k \geq k_0 + 1$, let the interval I_k's have the same length of $|I_{k_0}| = 2\delta^{-8k_0}|I|$. The number of such intervals within $[2 - 2\varepsilon, 2 + 2\varepsilon]$ is bounded crudely by $o(n)$. And the distance from $\lambda_i(W_n)$ to the interval I_k satisfies

$$\mathrm{dist}(\lambda_i(W_n), I_k) \geq \beta_{k_0-1}|I| + (k-k_0)|I_{k_0}|.$$

The contribution of such intervals can be computed similarly by

$$\frac{1}{n}\sum_{j\in J_k}\frac{|u_j(W_{n-1})^*X|^2}{|\lambda_j(W_{n-1})-\lambda_i(W_n)|} \leq \frac{1}{n}\frac{1}{\mathrm{dist}(\lambda_i(W_n), I_k)}\sum_{j\in J_k}|u_j(W_{n-1})^*X|^2$$
$$\leq \frac{1}{n}\frac{1}{\mathrm{dist}(\lambda_i(W_n), I_k)}(|J_k| + K\sqrt{|J_k|}\sqrt{C\log n} + CK^2\log n)$$
$$\leq \frac{\alpha_{I_k}|I_k|}{\mathrm{dist}(\lambda_i(W_n), I_k)} + \frac{100\delta^{k_0}}{k-k_0}$$

with probability at least $1-n^{-C_1-100}$.

Sum over all intervals for $k \geq 18$ (say), then

$$|\sum_{j\geq T_+ \mathrm{or} j\leq T_-}\frac{|u_j(W_{n-1})^*Y|^2}{\lambda_j(W_{n-1})-\lambda_i(W_n)}| \leq |\sum_{I_k}\frac{\alpha_{I_k}|I_k|}{\mathrm{dist}(\lambda_i(W_n), I_k)}| + \delta. \tag{29}$$

Using Riemann integration of the principal value integral,

$$\sum_{I_k}\frac{\alpha_{I_k}|I_k|}{\mathrm{dist}(\lambda_i(W_n), I_k)} = p.v.\int_{-2}^{2}\frac{\rho_{sc}(x)}{\lambda_i(W_n)-x}\,dx + o(1),$$

where

$$p.v.\int_{-2}^{2}\frac{\rho_{sc}(x)}{\lambda_i(W_n)-x}\,dx := \lim_{\varepsilon\to 0}\int_{-2\leq x\leq 2, |x-\lambda_i(W_n)|\geq\varepsilon}\frac{\rho_{sc}(x)}{\lambda_i(W_n)-x}\,dx,$$

and using the explicit formula for the Stieltjes transform and from residue calculus, one obtains

$$p.v.\int_{-2}^{2}\frac{\rho_{sc}(x)}{x-\lambda_i(W_n)}\,dx = -\lambda_i(W_n)/2$$

for $|\lambda_i(W_n)| \leq 2$, with the right-hand side replaced by $-\lambda_i(W_n)/2 + \sqrt{\lambda_i(W_n)^2-4}/2$ for $|\lambda_i(W_n)| > 2$. Finally, we always have

$$|\sum_{I_k}\frac{\alpha_{I_k}|I_k|}{\mathrm{dist}(\lambda_i(W_n), I_k)}| \leq 1+\delta \leq 1+\epsilon/1000.$$

Now for the rest of eigenvalues such that $|\lambda_i(W_n)-\lambda_j(W_{n-1})| \leq |I_0|+|I_1|+\ldots+|I_{18}| \leq |I|/\delta^{60}$, the number of eigenvalues is given by $T_+ - T_- \leq n|I|/\delta^{60} = CK^2\log n/\delta^{68}$. Thus

$$\frac{\sqrt{T_+ - T_-}}{CK\sqrt{\log n}} \leq \frac{1}{\delta^{34}\sqrt{C}} \leq \epsilon/1000,$$

by choosing C sufficiently large. Thus, with probability at least $1-n^{-C_1-10}$,

$$|x| \leq \frac{C_2K^2\log n}{\sqrt{n}}.$$

Appendix B. Local Marchenko-Pastur law for random covariance matrix and delocalization of singular vectors

In this Appendix, we extend the results obtained for random Hermitian matrices discussed in the previous sections to random covariance matrices, focusing on the changes needed for the proofs. Interested reader can refer to closely related papers [23] and [25] (see also [10], [18]).

Let $M_{n,p} = (\zeta_{ij})_{1\le i\le p, 1\le j\le n}$ be a random $p\times n$ matrix, where $p = p(n)$ is an integer such that $p \le n$ and $\lim_{n\to\infty} p/n = y$ for some $y \in (0, 1]$. Assume the atom variables ζ_{ij} are jointly independent, K-concentrated and have mean zero and variance one. The matrix ensemble M is said to obey condition **C1** if the random variables ζ_{ij} are jointly independent, have mean zero and variance one, and obey the moment condition $\sup_{i,j} \mathbf{E}|\zeta_{ij}|^{C_0} \le C$ for some constant C independent of n, p.

For such a $p \times n$ random matrix M, we form the $n \times n$ *(sample) covariance matrix* $W = W_{n,p} = \frac{1}{n}M^*M$. This (non-negative definite) matrix has at most p non-zero eigenvalues which are ordered as

$$0 \le \lambda_1(W) \le \lambda_2(W) \le \ldots \le \lambda_p(W).$$

Denote $\sigma_1(M), \ldots, \sigma_p(M)$ to be the singular values of M. Notice that $\sigma_i(M) = \sqrt{n}\lambda_i(W)^{1/2}$. From the singular value decomposition, there exist orthonormal bases $u_1, \ldots, u_p \in \mathbb{C}^n$ and $v_1, \ldots, v_p \in \mathbb{C}^p$ such that $Mu_i = \sigma_i v_i$ and $M^*v_i = \sigma_i u_i$.

The first fundamental result concerning the asymptotic limiting behavior of ESD for large covariance matrices is the *Marchenko-Pastur Law* (see [15] and [2]).

Theorem 27. *(Marchenko-Pastur Law) Assume a $p \times n$ random matrix M obeys condition* **C1** *with $C_0 \ge 4$, and $p, n \to \infty$ such that $\lim_{n\to\infty} p/n = y \in (0, 1]$, the empirical spectral distribution of the matrix $W_{n,p} = \frac{1}{n}M^*M$ converges in distribution to the Marchenko-Pastur Law with a density function*

$$\rho_{MP,y}(x) := \frac{1}{2\pi xy}\sqrt{(b-x)(x-a)}\mathbf{1}_{[a,b]}(x),$$

where

$$a := (1-\sqrt{y})^2, b := (1+\sqrt{y})^2.$$

The hard edge of the limiting support of spectrum refers to the left edge a when $y = 1$ where it gives rise to a singularity of $x^{-1/2}$. The cases of left edge a when $y < 1$ and the right edge b regardless of the value of y are called the soft edges. Recent progress on studying the local convergence to *Marchenko-Pastur Law* include [10], [18],[23],[25] for the soft edge and [20], [4] for the hard edge. In this paper, we focus on improving the previous results for the soft edge case.

Our main results for the random covariance matrices are the following local Marchenko-Pastur law (LMPL) and the delocalization property of singular vectors.

Theorem 28. *For any constants $\epsilon, \delta, C_1 > 0$ there exists $C_2 > 0$ such that the following holds. Assume that $p/n \to y$ for some $0 < y \le 1$. Let $M = (\zeta_{ij})_{1\le i\le p,1\le j\le n}$ be a random matrix whose entries are independent K-concentrated random variables with mean zero and variance 1. Consider the covariance matrix $W_{n,p} = \frac{1}{n}M^*M$. Then with probability at least $1 - n^{-C_1}$, one has*

$$|N_I(W_{n,p}) - p\int_I \rho_{MP,y}(x)\,dx| \le \delta p|I|.$$

for any interval $I \subset (a+\epsilon, b-\epsilon)$ of length at least $C_2K^2\log n/n$.

Theorem 29 (Delocalization of singular vectors)**.** *For any constant $C_1 > 0$ there is a constant $C_2 > 0$ such that the following holds.*

- *(Bulk case) For any $\epsilon > 0$ and any $1 \le i \le p$ with $\sigma_i(M_{n,p})^2/n \in [a+\epsilon, b-\epsilon]$, let u_i denote the corresponding (left or right) unit singular vector, then*
$$\|u_i\|_\infty \le \frac{C_2 K \log^{1/2} n}{\sqrt{n}}$$
with probability at least $1 - n^{-C_1}$.
- *(Edge case) For any $\epsilon > 0$ and any $1 \le i \le p$ with $\sigma_i(M_{n,p})^2/n \in [a-\epsilon, a+\epsilon] \cup [b-\epsilon, b+\epsilon]$ if $a \neq 0$ and $\sigma_i(M_{n,p})^2/n \in [4-\epsilon, 4]$ if $a = 0$, let u_i denote the corresponding (left or right) unit singular vector, then*
$$\|u_i\|_\infty \le \frac{C_2 K^2 \log n}{\sqrt{n}}$$
with probability at least $1 - n^{-C_1}$.

B.1. **Proof of Theorem 28.** Similarly to the Hermitian case, we compare the *Stieltjes transform* of the ESD of matrix W
$$s(z) := \frac{1}{p} \sum_{i=1}^{p} \frac{1}{\lambda_i(W) - z},$$
with the *Stieltjes transform* of *Marchenko-Pastur Law*
$$s_{MP,y}(z) := \int_{\mathbb{R}} \frac{1}{x-z} \rho_{MP,y}(x)\,dx = \int_a^b \frac{1}{2\pi x y (x-z)} \sqrt{(b-x)(x-a)}\,dx,$$
which is the unique solution to the equation
$$s_{MP,y}(z) + \frac{1}{y + z - 1 + yz s_{MP,y}(z)} = 0$$
in the upper half plane. We will show that $s(z)$ satisfies a similar equation.

The analogue of Lemma 20 is the following:

Proposition 30. (Lemma 29, [23]) *Let $1/10 \ge \eta \ge 1/n$, and $L_1, L_2, \varepsilon, \delta > 0$. For any constant $C_1 > 0$, there exists a constant $C > 0$ such that if one has the bound*
$$|s(z) - s_{MP,y}(z)| \le \delta$$
with (uniformly) probability at least $1 - n^{-C}$ for all z with $L_1 \le Re(z) \le L_2$ and $Im(z) \ge \eta$. Then for any interval I in $[L_1 - \varepsilon, L_2 + \varepsilon]$ with $|I| \ge max(2\eta, \frac{\eta}{\delta} \log \frac{1}{\delta})$, one has
$$|N_I - n \int_I \rho_{MP,y}(x)\,dx| \le \delta n |I|$$
with probability at least $1 - n^{-C_1}$.

Our objective is to show
$$|s(z) - s_{MP,y}(z)| = o(\delta) \tag{30}$$
with probability at least $1 - n^{-C}$for all z in the region R_y, where
$$R_y = \{z \in \mathbb{C} : |z| \le 10, a - \epsilon \le \mathrm{Re}(z) \le b + \epsilon, \mathrm{Im}(z) \ge \eta\}$$
if $y \neq 1$, and
$$R_y = \{z \in \mathbb{C} : |z| \le 10, \epsilon \le \mathrm{Re}(z) \le 4 + \epsilon, \mathrm{Im}(z) \ge \eta\}$$

if $y = 1$. We use the parameter $\eta = \frac{K^2C^2 \log n}{n\delta^6}$. In the defined region R_y, $|s_{MP,y}(z)| = O(1)$.

First, by Schur's complement, one can rewrite

$$s(z) = \frac{1}{p}\mathrm{Tr}(W - zI)^{-1} = \frac{1}{p}\sum_{k=1}^{p} \frac{1}{\xi_{kk} - z - Y_k} \tag{31}$$

where $Y_k = a_k^*(W_k - zI)^{-1}a_k$, and W_k is the matrix $W^* = \frac{1}{n}MM^*$ with the k^{th} row and column removed, and a_k is the k^{th} row of W with the k^{th} element removed. Let M_k be the $(p-1) \times n$ minor of M with the k^{th} row removed and $X_i^* \in \mathbb{C}^n$ $(1 \le i \le p)$ be the rows of M. Thus $\xi_{kk} = X_k{}^*X_k/n = ||X_k||^2/n, a_k = \frac{1}{n}M_kX_k, W_k = \frac{1}{n}M_kM_k^*$. And

$$Y_k = \sum_{j=1}^{p-1} \frac{|a_k^*v_j(M_k)|^2}{\lambda_j(W_k) - z} = \sum_{j=1}^{p-1} \frac{1}{n}\frac{\lambda_j(W_k)|X_k^*u_j(M_k)|^2}{\lambda_j(W_k) - z}$$

where $u_1(M_k), \ldots, u_{p-1}(M_k) \in \mathbb{C}^n$ and $v_1(M_k), \ldots, v_{p-1}(M_k) \in \mathbb{C}^{p-1}$ are orthonormal right and left singular vectors of M_k. Here we used the facts that $a_k^*v_j(M_k) = \frac{1}{n}X_k^*M_k^*v_j(M_k) = \frac{1}{n}\sigma_j(M_k)X_k^*u_j(M_k)$ and $\sigma_j(M_k)^2 = n\lambda_j(W_k)$.

The entries of X_k are independent of each other and of W_k, and have mean 0 and variance 1. Noticed $u_j(M_k)$ is a unit vector. By linearity of expectation we have

$$\mathbf{E}(Y_k|W_k) = \sum_{j=1}^{p-1} \frac{1}{n}\frac{\lambda_j(W_k)}{\lambda_j(W_k) - z} = \frac{p-1}{n} + \frac{z}{n}\sum_{j=1}^{p-1}\frac{1}{\lambda_j(W_k) - z} = \frac{p-1}{n}(1 + zs_k(z))$$

where

$$s_k(z) = \frac{1}{p-1}\sum_{i=1}^{p-1}\frac{1}{\lambda_i(W_k) - z}$$

is the *Stieltjes transform* for the ESD of W_k. From the Cauchy interlacing law, we can get

$$|s(z) - (1 - \frac{1}{p})s_k(z)| = O(\frac{1}{p}\int_{\mathbb{R}} \frac{1}{|x - z|^2}\,dx) = O(\frac{1}{p\eta})$$

and thus

$$\mathbf{E}(Y_k|W_k) = \frac{p-1}{n} + z\frac{p}{n}s(z) + O(\frac{1}{n\eta}) = \frac{p-1}{n} + z\frac{p}{n}s(z) + o(\delta^2).$$

In fact a similar estimate holds for Y_k itself:

Proposition 31. *For $1 \le k \le n$, $Y_k = \mathbf{E}(Y_k|W_k) + o(\delta^2)$ holds with probability at least $1 - 20n^{-C}$ uniformly for all z in the region R_y.*

To prove Proposition 31, we decompose

$$Y_k - \mathbf{E}(Y_k|W_k) = \sum_{j=1}^{p-1}\frac{\lambda_j(W_k)}{n}\left(\frac{|X_k^*u_j(M_k)|^2 - 1}{\lambda_j(W_k) - z}\right) := \frac{1}{n}\sum_{j=1}^{p-1}\frac{\lambda_j(W_k)}{\lambda_j(W_k) - x - \sqrt{-1}\eta}R_j. \tag{32}$$

The estimation of (32) is a repetition of the calculation in (25). Therefore, inserting the bounds to (31), we have

$$s(z) + \frac{1}{y + z - 1 + yzs(z) + o(\delta^2)} = 0,$$

with probability at least $1 - 10n^{-C}$. By a continuity argument (see [25] for details), one has $|s(z) - s_{MP,y}(z)| = o(\delta)$ with probability at least $1 - n^{-C}$. By Proposition 30, one can derive the following LMPL for random covariance matrices.

Theorem 32. *For any constants $\epsilon, \delta, C_1 > 0$, there exists $C_2 > 0$ such that the following holds. Assume that $p/n \to y$ for some $0 < y \leq 1$. Let $M = (\zeta_{ij})_{1\leq i\leq p, 1\leq j\leq n}$ be a random matrix with entries bounded by K where K may depend on n. Consider the covariance matrix $W_{n,p} = \frac{1}{n}M^*M$. Then with probability at least $1 - n^{-C_1}$, one has*

$$|N_I - p\int_I \rho_{MP,y}(x)\,dx| \leq \delta p|I|,$$

for any interval $I \subset (a-\epsilon, b+\epsilon)$ if $a \neq 0$ and $I \subset (\epsilon, 4+\epsilon)$ if $a = 0$ of length at least $C_2K^2\log n/n\delta^8$.

In particular, Theorem 28 follows.

B.2. **Proof of Theorem 29.** To prove the delocalization of singular vectors, we need the following formula that expresses an entry of a singular vector in terms of the singular values and singular vectors of a minor. It is enough to prove the delocalization for the right unit singular vectors.

Lemma 33 (Corollary 25, [23])**.** *Let $p, n \geq 1$, and let*

$$M_{p,n} = \begin{pmatrix} M_{p,n-1} & X \end{pmatrix}$$

be a $p \times n$ matrix for some $X \in \mathbb{C}^p$, and let $\begin{pmatrix} u \\ x \end{pmatrix}$ be a right unit singular vector of $M_{p,n}$ with singular value $\sigma_i(M_{p,n})$, where $x \in \mathbb{C}$ and $u \in \mathbb{C}^{n-1}$. Suppose that none of the singular values of $M_{p,n-1}$ are equal to $\sigma_i(M_{p,n})$. Then

$$|x|^2 = \frac{1}{1 + \sum_{j=1}^{min(p,n-1)} \frac{\sigma_j(M_{p,n-1})^2}{(\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2)^2}|v_j(M_{p,n-1})^*X|^2},$$

where $v_1(M_{p,n-1}), \ldots, v_{min(p,n-1)}(M_{p,n-1}) \in \mathbb{C}^p$ is an orthonormal system of left singular vectors corresponding to the non-trivial singular values of $M_{p,n-1}$.

In a similar vein, if

$$M_{p,n} = \begin{pmatrix} M_{p-1,n} \\ Y^* \end{pmatrix}$$

for some $Y \in \mathbb{C}^n$, and $\begin{pmatrix} v \\ y \end{pmatrix}$ is a left unit singular vector of $M_{p,n}$ with singular value $\sigma_i(M_{p,n})$, where $y \in \mathbb{C}$ and $v \in \mathbb{C}^{p-1}$, and none of the singular values of $M_{p-1,n}$ are equal to $\sigma_i(M_{p,n})$, then

$$|y|^2 = \frac{1}{1 + \sum_{j=1}^{min(p-1,n)} \frac{\sigma_j(M_{p-1,n})^2}{(\sigma_j(M_{p-1,n})^2 - \sigma_i(M_{p,n})^2)^2}|u_j(M_{p-1,n})^*Y|^2},$$

where $u_1(M_{p-1,n}), \ldots, u_{min(p-1,n)}(M_{p-1,n}) \in \mathbb{C}^n$ is an orthonormal system of right singular vectors corresponding to the non-trivial singular values of $M_{p-1,n}$.

First, if $\lambda_i(W_{p,n})$ lies within the bulk of spectrum, by Theorem 32, one can find an interval $I \subset (a+\varepsilon, b-\varepsilon)$, centered at $\lambda_i(W_{p,n})$ and with length $|I| = K^2C_2^2\log n/2n$ such that $N_I \geq \delta_1 n|I|$ ($\delta_1 > 0$ small constant) with probability at least $1 - n^{-C_1-10}$. By Cauchy interlacing law, we can find a set $J \subset \{1, \ldots, n-1\}$ with $|J| \geq N_I/2$ such that $|\lambda_j(W_{n-1}) - \lambda_i(W_n)| \leq |I|$ for all $j \in J$. Applying Lemma 25, one has

$$\begin{aligned}
&\sum_{j=1}^{\min(p,n-1)} \frac{\sigma_j(M_{p,n-1})^2}{(\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2)^2} |v_j(M_{p,n-1})^* X|^2 \\
&\geq \frac{1}{n} \sum_{j\in J} \frac{\lambda_j(W_{p,n-1})}{(\lambda_j(W_{p,n-1}) - \lambda_i(W_{p,n}))^2} |v_j(M_{p,n-1})^* X|^2 \\
&\geq \sum_{j\in J} n^{-1}|I|^{-2}|v_j(M_{p,n-1})^* X|^2 \gg n^{-1}|I|^{-2}|J| \gg |I|^{-1}
\end{aligned}$$

with probability at least $1 - n^{-C_1-10}$.

Thus, by the union bound and Lemma 33, $\|u_i(M_{p,n})\|_\infty \leq \frac{C_2 K \log^{1/2} n}{\sqrt{n}}$ holds with probability at least $1 - n^{-C_1}$.

For the edge case, where $|\lambda_i(W_{p,n}) - a| = o(1)$ $(a \neq 0)$ or $|\lambda_i(W_{p,n}) - b| = o(1)$, we refer to the analogue of Lemma 26.

Lemma 34 (Interlacing identity for singular values, Lemma 3.5 [25]). *Assume the notations in Lemma 33, then for every i,*

$$\sum_{j=1}^{min(p,n-1)} \frac{\sigma_j(M_{p,n-1})^2 |v_j(M_{p,n-1})^* X|^2}{\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2} = ||X||^2 - \sigma_i(M_{p,n})^2. \tag{33}$$

Similarly, we have

$$\sum_{j=1}^{min(p-1,n)} \frac{\sigma_j(M_{p-1,n})^2 |u_j(M_{p-1,n})^* Y|^2}{\sigma_j(M_{p-1,n})^2 - \sigma_i(M_{p,n})^2} = ||Y||^2 - \sigma_i(M_{p,n})^2. \tag{34}$$

By the union bound and Lemma 25, in order to show $|x|^2 \leq C^4 K^2 \log^2 n/n$ with probability at least $1 - n^{-C_1-10}$ for some large constant $C > C_1 + 100$, it is enough to show

$$\sum_{j=1}^{\min(p,n-1)} \frac{\sigma_j(M_{p,n-1})^2}{(\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2)^2} |v_j(M_{p,n-1})^* X|^2 \geq \frac{n}{C^4 K^4 \log^2 n}.$$

By the projection lemma, $|v_j(M_{p,n-1})^* X| \leq K\sqrt{C \log n}$ with probability at least $1 - 10n^{-C}$.

It suffices to show that with probability at least $1 - n^{-C_1-100}$,

$$\sum_{j=1}^{\min(p,n-1)} \frac{\sigma_j(M_{p,n-1})^2}{(\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2)^2} |v_j(M_{p,n-1})^* X|^4 \geq \frac{n}{C^3 K^2 \log n}.$$

By Cauchy-Schwardz inequality and the fact $|\sigma_i(M_{p,n-1})| = O(\sqrt{n})$, it is enough to show for some integers $1 \leq T_- < T_+ \leq \min(p, n-1)$ (the choice of T_-, T_+ will be given later),

$$\sum_{T_- \leq j \leq T_+} \frac{\frac{1}{n}\sigma_j(M_{p,n-1})^2}{|\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2|} |v_j(M_{p,n-1})^* X|^2 \geq \frac{\sqrt{T_+ - T_-}}{C^{1.5} K \sqrt{\log n}}$$

On the other hand, by the projection lemma, with probability at least $1-n^{-C_1-100}$, $\|X\|^2/n = y+o(1)$. By (33) in Lemma 34,

$$\sum_{j=1}^{\min(p,n-1)} \frac{1}{n}\frac{\sigma_j(M_{p,n-1})^2|v_j(M_{p,n-1})^*X|^2}{\sigma_j(M_{p,n-1})^2-\sigma_i(M_{p,n})^2} = y+o(1)-\lambda_i(W_{p,n}). \tag{35}$$

It is enough to evaluate

$$\sum_{j\geq T_+ \text{or} j\leq T_-} \frac{\lambda_j(W_{p,n-1})|v_j(M_{p,n-1})^*X|^2}{\lambda_j(W_{p,n-1})-\lambda_i(W_{p,n})}. \tag{36}$$

Now we divide the real line into disjoint intervals I_k for $k\geq 0$. Let $|I| = \frac{K^2C\log n}{n\delta^8}$ with constant $\delta \leq \epsilon/1000$. Denote $\beta_k = \sum_{s=0}^k \delta^{-8s}$. Let $I_0 = (\lambda_i(W_{p,n})-|I|, \lambda_i(W_{p,n})+|I|)$. For $1\leq k\leq k_0 = \log^{0.9} n$ (say),

$$I_k = (\lambda_i(W_{p,n})-\beta_k|I|, \lambda_i(W_{p,n})-\beta_{k-1}|I|]\cup[\lambda_i(W_{p,n})+\beta_{k-1}|I|, \lambda_i(W_{p,n})+\beta_k|I|),$$

thus $|I_k| = 2\delta^{-8k}|I| = o(1)$ and the distance from $\lambda_i(W_{p,n})$ to the interval I_k satisfies

$$\text{dist}(\lambda_i(W_{p,n}), I_k) \geq \beta_{k-1}|I|.$$

For each such interval, by Theorem 17, the number of eigenvalues $|J_k| = N_{I_k} \leq n\alpha_{I_k}|I_k| + \delta^k n|I_k|$ with probability at least $1-n^{-C_1-100}$, where $\alpha_{I_k} = \int_{I_k}\rho_{MP,y}(x)dx/|I_k|$.

By Lemma **??**, for the kth interval, with probability at least $1-n^{-C_1-100}$,

$$\begin{aligned}
&\frac{1}{n}\sum_{j\in J_k}\frac{|\lambda_j(W_{p,n-1})||v_j(M_{p,n-1})^*X|^2}{|\lambda_j(W_{p,n-1})-\lambda_i(W_{p,n})|} \leq \frac{1}{n}(1+\frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_n),I_k)})\sum_{j\in J_k}|v_j(W_{p,n-1})^*X|^2\\
&\leq \frac{1}{n}(1+\frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_{p,n}),I_k)})(|J_k|+K\sqrt{|J_k|}\sqrt{C\log n}+CK^2\log n)\\
&\leq \frac{1}{n}(1+\frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_{p,n}),I_k)})(p\alpha_{I_k}|I_k|+\delta^k p|I_k|+4K\sqrt{C\log n}\sqrt{n}\sqrt{|I_k|}+CK^2\log n)\\
&\leq y(1+\frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_{p,n}),I_k)})\alpha_{I_k}|I_k|+100\delta^{-7k-4}|I|.
\end{aligned}$$

For $k_0+1\leq k\leq N$, let the interval I_k's have the same length of $|I_{k_0}| = 2\delta^{-8k_0}|I|$. And the distance from $\lambda_i(W_{p,n})$ to the interval I_k satisfies

$$\text{dist}(\lambda_i(W_{p,n}), I_k) \geq \beta_{k_0-1}|I| + (k-k_0)|I_{k_0}|.$$

The contribution of such intervals can be computed similarly by

$$\begin{aligned}
&\frac{1}{n}\sum_{j\in J_k}\frac{|\lambda_j(W_{p,n-1})||v_j(M_{p,n-1})^*X|^2}{|\lambda_j(W_{p,n-1})-\lambda_i(W_{p,n})|} \leq \frac{1}{n}(1+\frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_n),I_k)})\sum_{j\in J_k}|v_j(W_{p,n-1})^*X|^2\\
&\leq \frac{1}{n}(1+\frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_{p,n}),I_k)})(|J_k|+K\sqrt{|J_k|}\sqrt{C\log n}+CK^2\log n)\\
&\leq y(1+\frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_{p,n}),I_k)})\alpha_{I_k}|I_k|+\frac{100\delta^{k_0-8}}{k-k_0},
\end{aligned}$$

with probability at least $1-n^{-C_1-100}$.

Sum over all intervals for $k \geq 20$ (say) and notice that $N\delta^{-8k_0}|I| = O(1)$. We have

$$\sum_{k=0}^{k_0} 100\delta^{-7k-4}|I| + \sum_{k=k_0}^{N} \frac{100\delta^{k_0-8}}{k-k_0} = o(1).$$

Using Riemann integration of the principal value integral,

$$y\sum_{I_k}(1+\frac{\lambda_i(W_{p,n})}{\mathrm{dist}(\lambda_i(W_{p,n}),I_k)})\alpha_{I_k}|I_k| = |p.v.\int_a^b y\frac{x\rho_{MP,y}(x)}{x-\lambda_i(W_{p,n})}\ dx| + o(1) \tag{37}$$

where (see [25] for details)

$$p.v.\int_a^b y\frac{x\rho_{MP,y}(x)}{x-\lambda_i(W_{p,n})}\ dx = \begin{cases} \sqrt{y}+o(1), & \text{if } |\lambda_i(W_{p,n})-a| = o(1), \\ -\sqrt{y}+o(1), & \text{if } |\lambda_i(W_{p,n})-b| = o(1). \end{cases} \tag{38}$$

by using the explicit formula for the Stieltjes transform and from residue calculus.

Now for the rest of eigenvalues such that $|\lambda_i(W_{p,n})-\lambda_j(W_{p,n-1})| \leq |I_0|+|I_1|+\ldots+|I_{20}| \leq |I|/\delta^{60}$. The number of eigenvalues is given by $T_+ - T_- \leq n|I|/\delta^{60} = CK^2\log n/\delta^{68}$. Thus

$$\frac{\sqrt{T_+ - T_-}}{C^{1.5}K\sqrt{\log n}} \leq \frac{1}{\delta^{34}C} \leq \epsilon/1000,$$

by choosing C sufficiently large. By comparing (35), (36) and (38), one can conclude with probability at least $1-n^{-C_1-10}$,

$$|x| \leq \frac{C_2K^2\log n}{\sqrt{n}}.$$

References

[1] G.B. Arous and P. Bourgade. Extreme gaps between eigenvalues of random matrices. *Annals of Probability.*, to appear.

[2] Z. Bai and J.W. Silverstein. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis*, 54(2):175–192, 1995.

[3] Z. Bai and J.W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Verlag, 2010.

[4] C. Cacciapuoti, A. Maltsev, and B. Schlein. Local marchenko-pastur law at the hard edge of sample covariance matrices. *Arxiv preprint arXiv:1206.1730*, 2012.

[5] Y. Dekel, J.R. Lee, and N. Linial. Eigenvectors of random graphs: Nodal domains. *Random Structures & Algorithms*, 39(1):39–58, 2011.

[6] L. Erdös. Universality of wigner random matrices: a survey of recent results. *Russian Mathematical Surveys*, 66:507, 2011.

[7] L. Erdös, B. Schlein, and H.T. Yau. Local semicircle law and complete delocalization for Wigner random matrices. *Communications in Mathematical Physics*, 287(2):641–655, 2009.

[8] L. Erdös, B. Schlein, and H.T. Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *The Annals of Probability*, 37(3):815–852, 2009.

[9] L. Erdös, B. Schlein, and H.T. Yau. Wegner estimate and level repulsion for Wigner random matrices. *International Mathematics Research Notices*, 2010(3):436, 2010.

[10] L. Erdös, B. Schlein, H.T. Yau, and J. Yin. The local relaxation flow approach to universality of the local statistics for random matrices. *Mathematical Physics*, 15(82B44), 2009.

[11] L. Erdős, H.T. Yau, and J. Yin. Bulk universality for generalized wigner matrices. *Probability Theory and Related Fields*, 154(1-2):341–407, 2012.

[12] D.L. Hanson and F.T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3):1079–1083, 1971.

[13] D. Hsu, S. M. Kakade, and T. Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *arXiv preprint arXiv:1110.2842*, 2011.

[14] Michel Ledoux. *The concentration of measure phenomenon*, volume 89. American Mathematical Soc., 2001.

[15] V.A. Marčenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1:457, 1967.
[16] M.L. Mehta. *Random matrices*, volume 142. Academic press, 2004.
[17] L.A. Pastur. On the spectrum of random matrices. *Theoretical and Mathematical Physics*, 10(1):67–74, 1972.
[18] N.S. Pillai and J. Yin. Universality of covariance matrices. *Arxiv preprint arXiv:1110.2501*, 2011.
[19] T. Tao and V. Vu. On random± 1 matrices: singularity and determinant. *Random Structures & Algorithms*, 28(1):1–23, 2006.
[20] T. Tao and V. Vu. Random matrices: The distribution of the smallest singular values. *Geometric And Functional Analysis*, 20(1):260–297, 2010.
[21] T. Tao and V. Vu. Random matrices: Universality of local eigenvalue statistics up to the edge. *Communications in Mathematical Physics*, pages 1–24, 2010.
[22] T. Tao and V. Vu. Random matrices: Universality of local eigenvalue statistics. *Acta mathematica*, 206(1):127–204, 2011.
[23] T. Tao and V. Vu. Random covariance matrices: Universality of local statistics of eigenvalues. *The Annals of Probability*, 40(3):1285–1315, 2012.
[24] T. Tao and V. Vu. Random matrices: The universality phenomenon for wigner ensembles. *arXiv preprint arXiv:1202.0068*, 2012.
[25] K. Wang. Random covariance matrices: Universality of local statistics of eigenvalues up to the edge. *Random Matrices: Theory and Appl.*, 1(1150005), 2012.
[26] F.T. Wright. A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *The Annals of Probability*, 1(6):1068–1070, 1973.

Department of Mathematics, Yale University, New Haven , CT 06520, USA

E-mail address: van.vu@yale.edu

Department of Mathematics, Rutgers University, Piscataway, NJ 08854

E-mail address: wkelucky@math.rutgers.edu