RANDOM COVARIANCE MATRICES: UNIVERSALITY OF LOCAL STATISTICS OF EIGENVALUES UP TO THE EDGE

KE WANG

ABSTRACT. We study the universality of the eigenvalue statistics of the covariance matrices $\frac{1}{n}M^*M$ where M is a large $p \times n$ matrix with independent entries that have mean zero, variance one and sufficiently high finite moments. In particular, as an application, we prove a variant of universality results regarding the smallest singular value of $M_{p,n}$. This paper is an extension of the results in [11] from the bulk of the spectrum up to the edge.

1. INTRODUCTION

The goal of this paper is to extend the Four Moment theorem established by Tao and Vu [11] for iid covariance matrices from the bulk of spectrum to the edge. Let us first specify the matrix ensembles that will be studied.

Definition 1.1 (Condition **C1**). 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 $y \in (0, 1]$. 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

$$M^*v_i = \sigma_i u_i.$$

For a square matrix W, the empirical spectral distribution (ESD) is the probability distribution that puts equal weight on each eigenvalue of W. In particular, if W is Hermitian and thus has real eigenvalues, the ESD is a one-dimensional function

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

where $|\mathbf{I}|$ denotes the cardinality of a set $\mathbf{I}$.

The first fundamental result concerning the asymptotic limiting behavior of ESD for large covariance matrices is the $Marchenko - Pastur\ Law$ due to [10] (see also [1]).

Theorem 1.2. *(Marchenko-Pastur Law) Assume a $p \times n$ random matrix M obeys condition* **C1** *with $C_0 \geq 4$, 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

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

We adapt these definitions introduced in Tao and Vu's paper [14].

Definition 1.3 (Frequent events). [14] Let E be an event depending on n.

- E holds *asymptotically almost surely* if $\mathbf{P}(E) = 1 - o(1)$.
- E holds *with high probability* if $\mathbf{P}(E) \geq 1 - O(n^{-c})$ for some constant $c > 0$ (independent of n).
- E holds *with overwhelming probability* if $\mathbf{P}(E) \geq 1 - O(n^{-C})$ for *every* constant $C > 0$ (or equivalently, that $\mathbf{P}(E) \geq 1 - \exp(-\omega(\log n))$).
- E holds *almost surely* if $\mathbf{P}(E) = 1$.

Definition 1.4 (Matching). [14] We say that two complex random variables ζ, ζ' *match to order k* for some integer $k \geq 1$ if one has $\mathbf{E}\mathrm{Re}(\zeta)^m \mathrm{Im}(\zeta)^l = \mathbf{E}\mathrm{Re}(\zeta')^m \mathrm{Im}(\zeta')^l$ for all $m, l \geq 0$ with $m + l \leq k$.

Our main result is the following Four Moment theorem, which extends the result (Theorem 6) in [11] to the edge of the spectrum. The proof is inspired by the proofs in [14], [13] and [11] and will be presented in Section 5.

Theorem 1.5 (Four Moment Theorem). *For sufficiently small $c_0 > 0$ and sufficiently large $C_0 > 0$ ($C_0 = 10^4$ will suffice) the following holds for every $k \geq 1$. Let $M = (\zeta_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ and $M' = (\zeta'_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ be two random matrices satisfying condition* **C1** *with the indicated constant C_0, and assume that for each i, j that ζ_{ij} and ζ'_{ij} match to order 4. Let W, W' be the associated covariance matrices. Assume also that $p/n \to y$ for some $0 < y \leq 1$.*

Let $G : \mathbb{R}^k \to \mathbb{R}$ be a smooth function obeying the derivative bounds

$$(1.1) \qquad\qquad\qquad\qquad |\nabla^j G(x)| \leq n^{c_0}$$

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

Then for any $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, and for n sufficiently large depending on k, c_0, we have

$$(1.2) \qquad |\mathbf{E}(G(n\lambda_{i_1}(W), \ldots, n\lambda_{i_k}(W))) - \mathbf{E}(G(n\lambda_{i_1}(W'), \ldots, n\lambda_{i_k}(W')))| \leq n^{-c_0}.$$

If ζ_{ij} and ζ'_{ij} only match to order 3 rather 4, then the conclusion (1.2) still holds provided that one strengthens (1.1) to

$$|\nabla^j G(x)| \leq n^{-jc_1}$$

for all $0 \leq j \leq 5$ and $x \in \mathbb{R}^k$ and any $c_1 > 0$, provided that c_0 is sufficiently small depending on c_1.

 The next theorem is an extension of Theorem 17 in [11], which is used in the proof of Theorem 1.5 and is of independent interest as well. The proof is delayed to Section 5.

Definition 1.6 (Gap property up to the edge). Let M be a random matrix obeying condition **C1**. We say M obeys the *gap property* if for every $c > 0$ and every $1 \leq i \leq p$, one has $|\lambda_{i+1}(W) - \lambda_i(W)| \geq n^{-1-c}$ with high probability.

Theorem 1.7 (Gap theorem up to the edge). *Let M be a random matrix satisfying condition **C1**. Then M obeys the gap property.*

 We would like to point out that a version of universality results for sample covariance matrices was proved in Erdős, Schlein, Yau and Yin's paper (Theorem 3.1, [8]). Roughly speaking, the k-point correlation functions of M^*M converge to those of *Gaussian* models in the sense of averaging over the energy parameter. For technical reasons, the ensembles are required to have symmetric distribution and satisfy logarithmic Sobolev inequality, besides mean 0, variance 1 and exponential decay. On the other hand, the Four Moment theorem implies the k-point correlation functions of M^*M, with M obeying condition **C1**, converge to those of *Gaussian* ensembles in the weak sense for any fixed energy parameter, provided the four moment matching (see Theorem 38 in [14] for details).

Remark 1.8. When $y = 1$, the singular value statistics around $a = 0$ turn out to be different since the density function $\rho_{\mathrm{MP},y}(x)$ has a singularity at $x = 0$. In this case, the normalized singular values $\frac{1}{\sqrt{n}}\sigma_1(M), \ldots, \frac{1}{\sqrt{n}}\sigma_n(M)$ are asymptotically distributed to the so-called *quarter-circle* distribution $\frac{1}{\pi}\sqrt{4 - x^2}\mathbf{1}_{[0,2]}(x)$ (Girko and Repin, 1995). The density function does not go to zero around $x = 0$, thus this is not actually an edge case. In this paper, we will focus on the edge case when $a > 0$.

Remark 1.9. Throughout this paper, we consider n as an asymptotic parameter that tends to infinity. We use the notation $X \ll Y$ $(Y \gg X)$ or $X = O(Y)$ if $X \leq CY$ for some constant C as long as n is sufficiently large. And $X = o(Y)$ or $Y = \omega(X)$ means that $X/Y \to 0$ as $n \to \infty$. We write $X = \Theta(Y)$ for $Y \ll X \ll Y$. The vectors $x \in \mathbf{C}^n$ are column vectors with the Euclidean norm defined as $\|x\| := (x^*x)^{1/2}$.

 This paper is organized as follows: in Section 2, we show a variant of universality result regarding the smallest singular value as an application of the Four Moment theorem. In Section 3, we collect a few basic results from linear algebra and probability. In Section 4, we provide the proofs of two technical lemmas, which are the major content of this paper. Finally, the object of Section 5 is to give the proofs of the Gap theorem (Theorem 1.7) and Four Moment theorem (Theorem 1.5). The argument draws heavily from those in [14], [13] and [11], thus we only focus on the changes needed to complete the proofs.

2. Applications

 In a similar way as [11] (Section 1.3), equipped with the Four Moment theorem, we can obtain universality results for large classes of random matrices. Let us demonstrate through some examples, stressing on the results for the lower edge of the spectrum. Recall $\sigma_1(M_{p,n})$ denotes the smallest singular value of $M_{p,n}$.

We adapt the notation in [12]. In this section, $M_{p,n}(\zeta)$ denotes the random $n \times p$ matrix whose entries are iid copies of a (real or complex-valued) random variable ζ. We say ζ is $\mathbb{R}$-*normalized* ($\mathbb{C}$-*normalized*) if ζ is real-valued with $\mathbf{E}\zeta = 0$ and $\mathbf{E}\zeta^2 = 1$ (complex-valued with $\mathbf{E}\zeta = 0$, $\mathbf{E}\mathrm{Re}(\zeta)^2 = \mathbf{E}\mathrm{Im}(\zeta)^2 = 1/2$, and $\mathbf{E}\mathrm{Re}(\zeta)\mathrm{Im}(\zeta) = 0$). A complex random variable ζ of mean 0 and variance 1 is *Gaussian divisible* if it has the same distribution as $(1-t)^{1/2}\zeta' + t^{1/2}\zeta''$ for some $0 < t < 1$, where ζ', ζ'' are independent with mean 0 and variance one, with ζ'' complex Gaussian.

For the case when $p = n$, the limiting distribution for $\sigma_1(M_{p,n})$ for the Gaussian models was computed by Edelman [4]. Recently a universality result has been established by Tao and Vu [12] requiring the entries to have bounded sufficiently high moments.

For the non-square case, where $p \leq n$, the distribution of the smallest singular value of $\mathbb{R}$-*normalized* ($\mathbb{C}$-*normalized*) entries has been studied by Borodin and Forrester [3]. Later, Feldheim and Sodin [9] proved a universality result for certain sample covariance matrices:

Theorem 2.1. [9] *Let $M_{p,n} = (\zeta_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ be a random covariance matrix, where $p = p(n) \leq n$ tends to infinity as $n \to \infty$ and $\limsup_{n \to \infty} p/n < 1$. Let ζ_{ij} be independent for all i, j.*

If ζ_{ij} are $\mathbb{R}$-normalized, exponential decaying and symmetric (that is, ζ_{ij} and $-\zeta_{ij}$ have the same distribution), then

$$(2.1) \qquad \frac{\sigma_1(M)^2 - (p^{1/2} - n^{1/2})^2}{(p^{1/2} - n^{1/2})(p^{-1/2} - n^{-1/2})^{1/3}} \to TW_1;$$

If ζ_{ij} are $\mathbb{C}$-normalized, exponential decaying and symmetric, then

$$(2.2) \qquad \frac{\sigma_1(M)^2 - (p^{1/2} - n^{1/2})^2}{(p^{1/2} - n^{1/2})(p^{-1/2} - n^{-1/2})^{1/3}} \to TW_2,$$

where TW_1, TW_2 denote the Tracy-Widom distributions.

In a same way as the authors proving ([14], Theorem 9) and ([11], Theorem 11), one can get the following (also see Figure 1 for numerical simulations):

Theorem 2.2. *The conclusions of Theorem 2.1 can be extended to the case when $p = p(n) \leq n$ tends to infinity as $n \to \infty$ and $\lim_{n \to \infty} p/n = y \in (0, 1)$, and when $M_{p,n} = (\zeta_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ obey condition $\mathbf{C1}$ with sufficiently large constant C_0, and ζ_{ij} have vanishing third moment.*

On the other hand, if the universality at the edge for random matrices $M_{p,n}(\zeta)$ with i.i.d. entries of *Gaussian divisible* distribution holds, then with the matching theorem (Corollary 30, [14]) and comparing with the *Gaussian divisible* ensembles, we can drop the third moment condition whereas ζ is assumed to be supported on at least three points, that is, the conclusion (2.2) of Theorem 2.1 can be extended to the case when $p = p(n) \leq n$ tends to infinity as $n \to \infty$ and $\lim_{n \to \infty} p/n = y \in (0, 1)$, and when $M_{p,n} = (\zeta_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ obey condition $\mathbf{C1}$ with sufficiently large constant C_0, and ζ_{ij} are supported on at least three points.

3. General Tools

In this section, we collect some basic tools from linear algebra and probability that will be used repeatedly in the sequel.
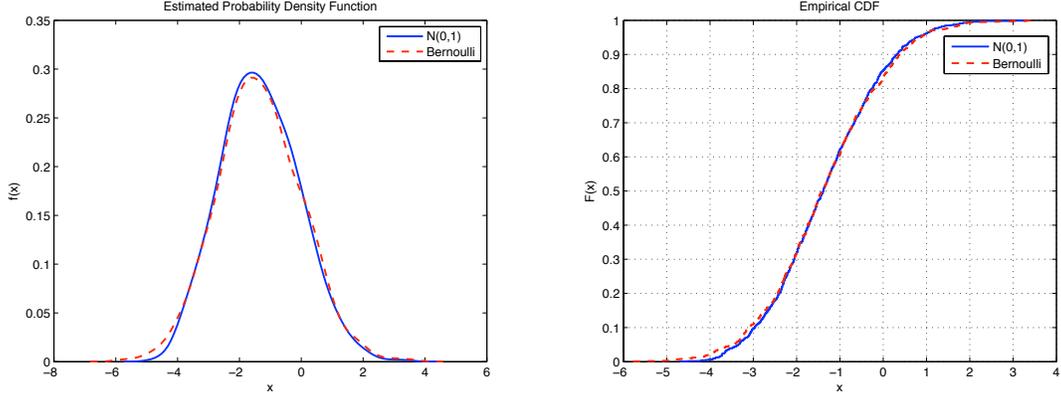
FIGURE 1. Plotted above are the empirical PDF and CDF of the distribution of $\sigma_1(M_{p,n}(\zeta))^2$ (normalized as in (2.1)) for $n = 800, p = 600$, based on data from 1000 random matrices. The blue solid curves were generated with $\zeta = N(0,1)$, while the red dashed curves were generated with ζ a random Bernoulli variable, taking values ± 1 with probability $1/2$ each.

3.1. **Tools from linear algebra.** We start with the Cauchy interlacing law and the Weyl inequalities.

Lemma 3.1 (Cauchy interlacing law). [11] *Let* $1 \leq p \leq n$.

(i) *If* A_n *is an* $n \times n$ *Hermitian matrix, and* A_{n-1} *is an* $(n-1) \times (n-1)$ *minor, then* $\lambda_i(A_n) \leq \lambda_i(A_{n-1}) \leq \lambda_{i+1}(A_n)$ *for all* $1 \leq i < n$.

(ii) *If* $M_{p,n}$ *is a* $p \times n$ *matrix, and* $M_{p-1,n}$ *is an* $(p-1) \times n$ *minor, then* $\sigma_i(M_{p,n}) \leq \sigma_i(M_{p-1,n}) \leq \sigma_{i+1}(M_{p,n})$ *for all* $1 \leq i < p$.

(iii) *If* $p < n$, *if* $M_{p,n}$ *is a* $p \times n$ *matrix, and* $M_{p,n-1}$ *is a* $p \times (n-1)$ *minor, then* $\sigma_{i-1}(M_{p,n}) \leq \sigma_i(M_{p,n-1}) \leq \sigma_i(M_{p,n})$ *for all* $1 \leq i \leq p$, *with the understanding that* $\sigma_0(M_{p,n}) = 0$. *(For* $p = n$, *one can of course use the transpose of (ii) instead.)*

Lemma 3.2 (Weyl inequality). [11] *Let* $1 \leq p \leq n$.

- *If* A, B *are* $n \times n$ *Hermitian matrices, then* $\|\lambda_i(A) - \lambda_i(B)\| \leq \|A - B\|_{op}$ *for all* $1 \leq i \leq n$.
- *If* M, N *are* $p \times n$ *matrices, then* $\|\sigma_i(M) - \sigma_i(N)\| \leq \|M - N\|_{op}$ *for all* $1 \leq i \leq p$.

The following formula for an entry of a singular vector, in terms of the singular values and singular vectors of a minor, is very useful:

Lemma 3.3 (Corollary 25, [11]). *Let* $p, n \geq 1$, *and let*

$$M_{p,n} = \left(\begin{array}{cc} M_{p,n-1} & X \end{array} \right)$$

be a $p \times n$ *matrix for some* $X \in \mathbb{C}^p$, *and let* $\left(\begin{array}{c} u \\ x \end{array} \right)$ *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 $M_{p-1,n}$.

The next lemma is the well-known Cauchy interlacing identities:

Lemma 3.4 (Lemma 40, [14]). *Let A_n be a $n \times n$ Hermitian matrix, and*

$$A_n = \begin{pmatrix} A_{n-1} & X \\ X^* & a_{nn} \end{pmatrix}.$$

Let $\lambda_i(A_n), 1 \le i \le n$ be the eigenvalues of A_n and $\lambda_j(A_{n-1}), 1 \le j \le n-1$ be the eigenvalues of A_{n-1}. Suppose that X is not orthogonal to any of the unit eigenvectors $u_j(A_{n-1})$ of A_{n-1}. Then we have

$$\sum_{j=1}^{n-1} \frac{|u_j(A_{n-1})^* X|^2}{\lambda_j(A_{n-1}) - \lambda_i(A_n)} = a_{nn} - \lambda_i(A_n)$$

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

From this lemma, one immediately gets an interlacing identity for singular values:

Lemma 3.5 (Interlacing identity for singular values). *Assume the notations in Lemma 3.3, then for every i,*

$$(3.1) \qquad \sum_{j=1}^{min(p,n-1)} \frac{\sigma_j(M_{p,n-1})^2 |v_j(M_{p,n-1})^* X|^2}{\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2} = ||X||^2 - \sigma_i(M_{p,n})^2.$$

Similarly, we have

$$(3.2) \qquad \sum_{j=1}^{min(p-1,n)} \frac{\sigma_j(M_{p-1,n})^2 |u_j(M_{p-1,n})^* Y|^2}{\sigma_j(M_{p-1,n})^2 - \sigma_i(M_{p,n})^2} = ||Y||^2 - \sigma_i(M_{p,n})^2.$$

Proof. Apply Lemma 3.4 to the matrix

$$M_{p,n}^* M_{p,n} = \begin{pmatrix} M_{p,n-1}^* M_{p,n-1} & M_{p,n-1}^* X \\ X^* M_{p,n-1} & ||X||^2 \end{pmatrix}$$

with eigenvalue $\sigma_i(M_{p,n})^2$.

Since we have $\lambda_j(M_{p,n-1}^* M_{p,n-1}) = \sigma_j(M_{p,n-1})^2$ and

$$u_j(M_{p,n-1}^* M_{p,n-1})^* M_{p,n-1}^* = \sigma_j(M_{p,n-1}) v_j(M_{p,n-1})^*,$$

(3.1) follows. Similarly, to show (3.2), apply Lemma 3.5 to the matrix

$$M_{p,n}M_{p,n}^* = \begin{pmatrix} M_{p-1,n}^* M_{p-1,n} & M_{p-1,n}Y \\ Y^* M_{p-1,n}^* & ||Y||^2 \end{pmatrix}.$$

$\square$

The *Stieltjes transform* $s_n(z)$ of a Hermitian matrix W is defined for $z \in \mathbb{C}$ by the formula

$$s_n(z) := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i(W) - z}.$$

By Schur's complement, it has the following alternate representation:

Lemma 3.6 (Lemma 39, [14]). *Let $W = (\zeta_{ij})_{1 \leq i,j \leq n}$ be a Hermitian matrix, and let z be a complex number not in the spectrum of W. Then we have*

$$s_n(z) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\zeta_{kk} - z - a_k^*(W_k - zI)^{-1}a_k}$$

where W_k is the $(n-1) \times (n-1)$ matrix with the kth row and column of W removed, and $a_k \in \mathbb{C}^{n-1}$ is the kth column of W with the kth entry removed.

3.2. **Tools from probability theory.**

We will rely frequently on the next concentration of measure result for projections of random vectors.

Lemma 3.7 (Lemma 43, [14]). *Let $X = (\zeta_1, \ldots, \zeta_n) \in \mathbb{C}^n$ be a random vector whose entries are independent with mean zero, variance 1, and are bounded in magnitude by K almost surely for some K, where $K \geq 10(\mathbf{E}|\xi|^4 + 1)$. Let H be a subspace of dimension d and π_H the orthogonal projection onto H. Then*

$$\mathbf{P}(| \| \pi_H(X) \| - \sqrt{d}| \geq t) \leq 10 \exp(-\frac{t^2}{10K^2}).$$

In particular, one has

$$\| \pi_H(X) \| = \sqrt{d} + O(K \log n)$$

with overwhelming probability.

Lemma 3.8 (Theorem 44, [14]). *Let $1 \leq N \leq n$ be integers, and let $A = (a_{ij})_{1 \leq i \leq N; 1 \leq j \leq n}$ be an $N \times n$ complex matrix whose N rows are orthonormal in $\mathbb{C}^n$, and obeying the incompressibility condition*

$$(3.3) \qquad \sup_{1 \leq i \leq N; 1 \leq j \leq n} |a_{ij}| \leq \sigma$$

for some $\sigma > 0$. Let $\zeta_1, \ldots, \zeta_n$ be independent complex random variables with mean zero, variance one, and $\mathbf{E}|\zeta_i|^3 \leq C$ for some $C \geq 1$. For each $1 \leq i \leq N$, let S_i be the complex random variable

$$S_i := \sum_{j=1}^{n} a_{ij}\zeta_j$$

and let $\vec{S}$ be the $\mathbb{C}^N$-valued random variable with coefficients $S_1, \ldots, S_N$.

- *(Upper tail bound on S_i) For $t \geq 1$, we have $\mathbf{P}(|S_i| \geq t) \ll \exp(-ct^2) + C\sigma$ for some absolute constant $c > 0$.*
- *(Lower tail bound on $\vec{S}$) For any $t \leq \sqrt{N}$, one has $\mathbf{P}(|\vec{S}| \leq t) \ll O(t/\sqrt{N})^{\lfloor N/4 \rfloor} + CN^4 t^{-3}\sigma$.*

The same claim holds if one of the ζ_i is assumed to have variance c instead of 1 for some absolute constant $c > 0$.

4. Main technical lemmas

Recall in the proofs of the Four Moment Theorem and the Gap Theorem as in [11], [14] and [13], a crucial input was the *Delocalization Theorem* of Erdős, Schlein, and Yau ([7], [6] and [5]). The material in this section is analogous to Section 3 of [11]. We will first extend the concentration of ESD result to the edge of the spectrum and use this concentration theorem to show the delocalization of singular vectors. The proof of the delocalization result in the edge of spectrum is significantly different from that in the bulk of spectrum as in [11]. However, similar to [13], the Cauchy interlacing identities for singular values in Theorem 3.5 will help us to deal with this problem.

First observe that if $M = (\zeta_{ij})$ obeys *condition* **C1** for some constant $C_0 > 0$, then by Markov's inequality and the union bound, one has $|\zeta_{ij}| \leq n^{10/C_0}$ for all i, j with probability $1 - O(n^{-8})$. By a truncation technique (see [2] for details) and Lemma 3.2, one may assume that

$$|\zeta_{ij}| \leq K := n^{10/C_0}$$

almost surely for all i, j.

We will derive the eigenvalue concentration theorem (up to the edge) which is an analogue of Theorem 19 in [11]:

Theorem 4.1 (Concentration up to the edge). *Suppose that $p/n \to y$ for some $0 < y \leq 1$. Assume $a > 0$. Let $M = (\zeta_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ obey condition **C1** for some $C_0 \geq 2$ and the probability distribution of ζ_{ij} be continuous. Assume further $|\zeta_{ij}| \leq K$ almost surely for some $K = o(n^{1/2}\delta^2 \log^{-1} n)$ for all i, j, where $0 < \delta < 1/2$ (which can depend on n). Then for any interval $I \subset [a, b]$ of length $|I| \geq \frac{K^2 \log^{4.5} n}{\delta^9 n}$, one has with overwhelming probability(uniformly in I) that the number of eigenvalues N_I of W in I obeys the concentration estimate*

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

As a consequence of Theorem 4.1, one can deduce the following delocalization theorem:

Theorem 4.2 (Delocalization of singular vectors up to edge). *Let the hypothesis be as in Theorem 4.1, then with overwhelming probability, all the unit left and right singular vectors of M have all coefficients uniformly of size at most $K^2 n^{-1/2} \log^{O(1)} n$.*

Remark 4.3. *The continuity hypothesis in the above theorems, which guarantees the singular values are almost surely simple, is only a technical one. In practice we are able to eliminate this hypothesis by a limiting argument using Lemma 3.2.*

4.1. **Proof of Theorem 4.2:** Let $u_1(M_{p,n}), \ldots, u_p(M_{p,n}) \in \mathbb{C}^n$ be the right singular vectors of $M_{p,n}$. By the union bound and symmetry, it suffices to show that

$$|u_i(M_{p,n})^* e_1| \leq K^2 n^{-1/2} \log^{O(1)} n$$

with overwhelming probability. The delocalization of left singular vectors can be proved similarly.

The "bulk" case is treated in [11]. Now we consider the edge case when $1 \leq i \leq \log^8 n$ or $n - \log^8 n \leq i \leq n$ (say). Using Theorem 4.2, we have with overwhelming probability that

$$|\lambda_i(W_{n,p}) - a| \leq o(1) \qquad \text{or} \qquad |\lambda_i(W_{n,p}) - b| \leq o(1).$$

By Lemma 3.3, it suffices to show with overwhelming probability that

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

From Lemma 3.7, we conclude that $|v_j(M_{p,n-1})^* X| \ll K \log n$ with overwhelming probability for each j (and hence for all j, by the union bound). Then it is enough to show that with overwhelming probability

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

By the Cauchy-Schwarz inequality, it thus suffices to show that

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

with overwhelming probability for some $1 \leq T_- < T_+ \ll K^2 \log^{O(1)} n$. Notice that $\sigma_j(M_{p,n-1})^2 = \lambda_j(M_{p,n-1}^* M_{p,n-1}) = \Theta(n)$, we thus need to show

$$(4.1) \qquad \sum_{T_- \leq j \leq T_+} \frac{\sigma_j(M_{p,n-1})^2}{|\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2|} |v_j(M_{p,n-1})^* X|^2 \gg n \log^{-O(1)} n$$

with overwhelming probability for some $1 \leq T_- < T_+ \ll K^2 \log^{O(1)} n$, which is equivalent to proving that

$$(4.2) \qquad \sum_{T_- \leq j \leq T_+} \frac{1}{n} \frac{\sigma_j(M_{p,n-1})^2 |v_j(M_{p,n-1})^* X|^2}{|\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2|} \gg \log^{-O(1)} n$$

with overwhelming probability for some $1 \leq T_- < T_+ \ll K^2 \log^{O(1)} n$.

In the interlacing identity in Lemma 3.5, we have

$$(4.3) \qquad \sum_{j=1}^{\min(p,n-1)} \frac{1}{n} \frac{\sigma_j(M_{p,n-1})^2 |v_j(M_{p,n-1})^* X|^2}{\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2} = \frac{1}{n} ||X||^2 - \lambda_i(W_{p,n}).$$

By Lemma 3.7, one gets $\frac{1}{p} ||X||^2 = 1 + o(1)$ with overwhelming probability. And since $p/n = y + o(1)$, one has

$$(4.4) \qquad \sum_{j=1}^{\min(p,n-1)} \frac{1}{n} \frac{\sigma_j(M_{p,n-1})^2 |v_j(M_{p,n-1})^* X|^2}{\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2} = y + o(1) - \lambda_i(W_{p,n})$$

with overwhelming probability. In order to show (4.2), we will evaluate

$$(4.5) \qquad \begin{aligned} &\sum_{j > i+T_+ \text{ or } j < i-T_-} \frac{1}{n} \frac{\sigma_j(M_{p,n-1})^2 |v_j(M_{p,n-1})^* X|^2}{\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2} \\ &= \sum_{j > i+T_+ \text{ or } j < i-T_-} \frac{1}{n} \frac{|v_j(M_{p,n-1})^* X|^2 \lambda_j(W_{p,n-1})}{\lambda_j(W_{p,n-1}) - \lambda_i(W_{p,n})} \end{aligned}$$

for some $T_-, T_+ = K^2 \log^{O(1)} n$, where $W_{p,n-1} = \frac{1}{n-1} M_{p,n-1}^* M_{p,n-1}$.

The Machenko-Pastur law implies $\lambda_j(W_{p,n-1}) = \Theta(1)$ for every $1 \le j \le \min(p, n-1)$ with overwhelming probability.

Let A be a large constant to be chosen later. From Theorem 4.1, we have that (by taking $\delta = \log^{-A/20} n$)

$$(4.6) \qquad\qquad N_I = p\alpha_I |I| + O(|I|p \log^{-A/20} n)$$

with overwhelming probability for any interval I of length $|I| = K^2 \log^A n / n$, where $\alpha_I := \frac{1}{|I|} \int_I \rho_{MP,y}(x)\, dx$. For such an interval, we see from Lemma 3.7 that with overwhelming probability

$$\sum_{j:\lambda_j(W_{p,n-1})\in I} \frac{1}{n} |v_j(M_{p,n-1})^* X|^2 = \frac{N_I}{n} + O(\frac{K^2 \log^{A/2} n}{n})$$

and thus by (4.6) (for A large enough),

$$\sum_{j:\lambda_j(W_{p,n-1})\in I} \frac{1}{n} |v_j(M_{p,n-1})^* X|^2 = y\alpha_I |I| + O(|I| \log^{-A/20} n).$$

Set $d_I := \frac{\text{dist}(\lambda_i(W_{p,n}),I)}{|I|}$. If $d_I \ge \log n$(say), then

$$\frac{\lambda_j(W_{p,n-1})}{\lambda_j(W_{p,n-1}) - \lambda_i(W_{p,n})} = 1 + \frac{\lambda_i(W_{p,n})}{d_I|I|} + O(\frac{\lambda_i(W_{p,n})}{d_I^2|I|})$$

for all j in the above sum, and since $\lambda_i(W_{p,n}) = \Theta(1)$, we get

$$(4.7) \quad \sum_{j:\lambda_j(W_{p,n-1})\in I} \frac{1}{n} \frac{|v_j(M_{p,n-1})^* X|^2 \lambda_j(W_{p,n-1})}{\lambda_j(W_{p,n-1}) - \lambda_i(W_{p,n})} = y\alpha_I |I| \left(1 + \frac{\lambda_i(W_{p,n})}{d_I|I|}\right) + O(\frac{\alpha_I}{d_I^2})$$

$$+ O(|I| \log^{-A/20} n) + O(\frac{\log^{-A/20} n}{d_I}).$$

We now partition the real line into intervals I of length $K^2 \log^A n/n$, and sum (4.7) over all intervals I with $d_I \ge \log n$. Bounding α_I crudely by $O(1)$, we see that $\sum_I O(\frac{\alpha_I}{d_I^2}) = O(\frac{1}{\log n}) = o(1)$. Similarly, one has

$$\sum_I O(|I| \log^{-A/20} n) = O(\log^{-A/20} n) = o(1)$$

and

$$\sum_I O(\frac{\log^{-A/20} n}{d_I}) = O(\log^{-A/20} n \log n) = o(1).$$

Finally, Riemann integration of the principal value integral

$$\text{p.v.} \int_a^b \frac{\rho_{MP,y}(x)}{x - \lambda_i(W_{p,n})}\, dx := \lim_{\varepsilon \to 0} \int_{a \le x \le b : |x - \lambda_i(W_{p,n})| > \varepsilon} \frac{\rho_{MP,y}(x)}{x - \lambda_i(W_{p,n})}\, dx$$

shows that

$$\sum_I y\alpha_I |I| \left(1 + \frac{\lambda_i(W_{p,n})}{d_I|I|}\right) = \text{p.v.} \int_a^b y \frac{x\rho_{MP,y}(x)}{x - \lambda_i(W_{p,n})}\, dx + o(1).$$

If $|\lambda_i(W_{p,n}) - a| \leq o(1)$, using the formula for the Stieltjes transform, one obtains from residue calculus that

$$
\begin{aligned}
\text{p.v.} \int_a^b y \frac{x\rho_{MP,y}(x)}{x - \lambda_i(W_{p,n})} \, dx &= y \left(1 + \text{p.v.} \lambda_i(W_{p,n}) \int_a^b \frac{\rho_{MP,y}(x)}{x - \lambda_i(W_{p,n})} \, dx \right) \\
&= y \left(1 + (1 - \sqrt{y})^2 \frac{1}{\sqrt{y} - y} \right) + o(1) \\
&= \sqrt{y} + o(1),
\end{aligned}
$$

thus

$$
\text{(4.8)} \qquad \sum_{j > i + T_+ \text{ or } j < i - T_-} \frac{1}{n} \frac{\sigma_j(M_{p,n-1})^2 |v_j(M_{p,n-1})^* X|^2}{\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2} = \sqrt{y} + o(1),
$$

and in (4.4),

$$
\text{(4.9)} \qquad \sum_{j=1}^{\min(p,n-1)} \frac{1}{n} \frac{\sigma_j(M_{p,n-1})^2 |v_j(M_{p,n-1})^* X|^2}{\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2} = y + o(1) - \lambda_i(W_{p,n}) = 2\sqrt{y} - 1 + o(1).
$$

When $0 < y < 1$, $\sqrt{y} > 2\sqrt{y} - 1$. (4.2) follows by comparing (4.8) and (4.9).

If $|\lambda_i(W_{p,n}) - b| \leq o(1)$, we have

$$
\begin{aligned}
\text{p.v.} \int_a^b y \frac{x\rho_{MP,y}(x)}{x - \lambda_i(W_{p,n})} \, dx &= y \left(1 + \text{p.v.} \lambda_i(W_{p,n}) \int_a^b \frac{\rho_{MP,y}(x)}{x - \lambda_i(W_{p,n})} \, dx \right) \\
&= y \left(1 - (1 + \sqrt{y})^2 \frac{1}{\sqrt{y} + y} \right) + o(1) \\
&= -\sqrt{y} + o(1),
\end{aligned}
$$

thus

$$
\text{(4.10)} \qquad \sum_{j > i + T_+ \text{ or } j < i - T_-} \frac{1}{n} \frac{\sigma_j(M_{p,n-1})^2 |v_j(M_{p,n-1})^* X|^2}{\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2} = -\sqrt{y} + o(1),
$$

and in (4.4),

$$
\text{(4.11)} \qquad \sum_{j=1}^{\min(p,n-1)} \frac{1}{n} \frac{\sigma_j(M_{p,n-1})^2 |v_j(M_{p,n-1})^* X|^2}{\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2} = y + o(1) - \lambda_i(W_{p,n}) = -2\sqrt{y} - 1 + o(1).
$$

When $0 < y \leq 1$, $-\sqrt{y} > -2\sqrt{y} - 1$. Then (4.2) follows by comparing (4.10) and (4.11).

By the concentration theorem 4.1 and the Cauchy interlacing law, the interval I with $d_I < \log n$ will contribute at most $K^2 \log^{O(1)} n$ eigenvalues and we can set T_-, T_+ accordingly. The proof is now complete.

4.2. **Proof of Theorem 4.1:** We first have a crude upper bound on the number of eigenvalues of W on an interval. The proof can be found in Section 5.2, [11].

Proposition 4.4. *(Upper bound on ESD) Assume the hypotheses in Theorem 4.1, then for any interval $I \subset \mathbb{R}$ with length $|I| \geq K \log^2 n/n$, one has*

$$
N_I \ll n|I|
$$

with overwhelming probability, where N_I is the number of eigenvalues in the interval I.

The strategy is to 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*

$$s_{MP,y}(z) := \int_{\mathcal{R}} \frac{1}{x - z} \rho_{MP,y}(x) \, dx = \int_{a}^{b} \frac{1}{2\pi x y (x - z)} \sqrt{(b - x)(x - a)} \, dx.$$

And thanks to the next proposition, one gets control on ESD through control on the *Stieltjes transforms*.

Proposition 4.5. *(Lemma 29, [11]) Let $1/10 \geq \eta \geq 1/n$, and $a, b, \varepsilon, \delta > 0$. Suppose that one has the bound*

$$|s(z) - s_{MP,y}(z)| \leq \delta$$

with (uniformly) overwhelming probability for all z with $a \leq Re(z) \leq b$ and $Im(z) \geq \eta$. Then for any interval I in $[a - \varepsilon, b + \varepsilon]$ with $|I| \geq max(2\eta, \frac{\eta}{\delta} \log \frac{1}{\delta})$, one has

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

with overwhelming probability.

By Proposition 4.5, our objective is to show

(4.12) $$|s(z) - s_{MP,y}(z)| = o(\delta)$$

with (uniformly) overwhelming probability for all z with $a \leq \text{Re}(z) \leq b$ and $\text{Im}(z) \geq \eta := \frac{K^2 \log^6 n}{n \delta^8}$.

Since $s_{MP,y}(z)$ is the unique solution to the equation

$$s_{MP,y}(z) + \frac{1}{y + z - 1 + y z s_{MP,y}(z)} = 0$$

in the upper half plane (see [2]), we investigate a similar equation for $s(z)$.

From Lemma 3.6, we have

(4.13) $$s(z) = \frac{1}{p} \sum_{k=1}^{p} \frac{1}{\xi_{kk} - z - Y_k}$$

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

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

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 4.6. *For $1 \leq k \leq n$, $Y_k = \mathbf{E}(Y_k|W_k) + o(\delta^2)$ holds with (uniformly) overwhelming probability for all z with $a \leq Re(z) \leq b$ and $Im(z) \geq \eta$.*

Proof. Decompose

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

Let $T \subset \{1, \ldots, n-1\}$. Let H be the space spanned by $\{u_j(W_k)\}$ for $j \in T$ and P_H be the orthogonal projection onto H. Thus $\sum_{j \in T} R_j = ||P_H(X_k)||^2 - \dim(H)$.

By Lemma 3.7, we conclude with overwhelming probability

$$(4.14) \qquad |\sum_{j \in T} R_j| \ll \sqrt{|T|} K \log n + K^2 \log^2 n.$$

Using the triangle inequality,

$$(4.15) \qquad \sum_{j \in T} |R_j| \ll |T| + K^2 \log^2 n.$$

Let $z = x + \sqrt{-1}\eta$, where $\eta = \frac{K^2 \log^6 n}{n\delta^8}$ and $a \leq x \leq b$. We will use two auxiliary parameters $\alpha = \delta^2 \log^{-1.1} n$, $\delta' = \delta^2 \log^{-0.1} n$ in later estimation.

First, for those $j \in T$ such that $|\lambda_j(W_k) - x| \leq \delta'\eta$, the function $\frac{\lambda_j(W_k)}{\lambda_j(W_k) - x - \sqrt{-1}\eta}$ has magnitude $O(\frac{1}{\eta})$. From Proposition 4.4, $|T| \ll n\delta'\eta$, the contribution for these $j \in T$,

$$|\frac{1}{n} \sum_{j \in T} \frac{\lambda_j(W_k)}{\lambda_j(W_k) - z} R_j| \ll \frac{1}{n\eta} \sum_{j \in J} |R_j| \ll \frac{n\delta'\eta + K^2 \log^2 n}{n\eta} = o(\delta^2).$$

For the contribution of the remaining indices, we subdivide them as

$$(1 + \alpha)^l \delta'\eta \leq |\lambda_j(W_k) - x| \leq (1 + \alpha)^{l+1} \delta'\eta$$

for $0 \leq l \ll \log n/\alpha$, and then sum over l.

For each such interval, the function $\frac{\lambda_j(W_k)}{\lambda_j(W_k)-x-\sqrt{-1}\eta}$ has magnitude $O(\frac{1}{(1+\alpha)^l\delta'\eta})$ and fluctuates by at most $O(\frac{\alpha}{(1+\alpha)^l\delta'\eta})$. Say $T(l)$ is the set of all j's in this interval, by Proposition 4.4, $|T(l)| \ll n\alpha(1+\alpha)^l\delta'\eta$. Together with bounds (4.14), (4.15), the contribution for these j on such an interval,

$$|\frac{1}{n}\sum_{j\in T(l)}\frac{\lambda_j(W_k)}{\lambda_j(W_k)-z}R_j| \ll \frac{1}{\alpha(1+\alpha)^l\delta'\eta}\frac{\sqrt{\alpha(1+\alpha)^l\delta'\eta n}K\log n+K^2\log^2 n}{n}$$

$$+\frac{\alpha}{(1+\alpha)^l\delta'\eta}\frac{\alpha(1+\alpha)^l\delta'\eta+K^2\log^2 n}{n}$$

$$\ll \frac{\sqrt{\alpha}}{\sqrt{\delta'\eta n}}K\log n+\frac{K^2\log^2 n}{\delta'\eta n}+\alpha^2.$$

Summing over l (taking into account that $l \ll \log n/\alpha$), we will get

$$|\frac{1}{n}\sum_l\sum_{j\in T(l)}\frac{\lambda_j(W_k)}{\lambda_j(W_k)-z}R_j| \ll \frac{K\log^2 n}{\sqrt{\alpha\delta'\eta n}}+\frac{K^2\log^3 n}{\alpha\delta'\eta n}+\alpha\log n=o(\delta^2).$$

$\square$

Recall $s_{MP,y}(z)$ has an explicit expression

$$s_{MP,y}(z) = -\frac{y+z-1-\sqrt{(y+z-1)^2-4yz}}{2yz},$$

where we take the branch of $\sqrt{(y+z-1)^2-4yz}$ with cut at $[a,b]$ that is asymptotically $y+z-1$ as $z\to\infty$.

From (4.13) and Proposition 4.6, we have with overwhelming probability that

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

where we used Lemma 3.7 to obtain that $\xi_{kk}=||X_k||^2/n=1+o(\delta^2)$ with overwhelming probability.

By assumption $p/n\to y$, when n is large enough,

$$(4.16)\qquad\qquad s(z)+\frac{1}{y+z-1+yzs(z)+o(\delta^2)}=0$$

holds with overwhelming probability.

In (4.16), for the error term $o(\delta^2)$, one has either $\frac{o(\delta^2)}{y+z-1+yzs(z)}=o(\delta^2)$ or $y+z-1+yzs(z)=o(1)$. In the latter case, we get $s(z)=-\frac{y+z-1}{yz}+o(1)$. In the first case, we impose a Taylor expansion on (4.16),

$$s(z)(y+z-1+yzs(z))+1+o(\delta^2)=0.$$

Completing a perfect square for $s(z)$ in the above identity, one can solve the equation for $s(z)$,

$$(4.17)\qquad\qquad \sqrt{yz}(s(z)+\frac{y+z-1}{2yz})=\pm\sqrt{\frac{(y+z-1)^2}{4yz}-1+o(\delta^2)}.$$

If $\frac{o(\delta^2)}{\sqrt{\frac{(y+z-1)^2}{4yz}-1}}=o(\delta)$, by a Taylor expansion on the right hand side of (4.17), we have $\sqrt{yz}(s(z)+$

$\frac{y+z-1}{2yz})=\pm\sqrt{\frac{(y+z-1)^2}{4yz}-1}+o(\delta)$. Therefore, $s(z)=s_{MP,y}(z)+o(\delta)$ or $s(z)=s_{MP,y}(z)-$

$\frac{\sqrt{(y+z-1)^2 - 4yz}}{yz} + o(\delta) = -s_{MP,y}(z) - \frac{y+z-1}{yz} + o(\delta)$. If $\frac{(y+z-1)^2}{4yz} - 1 = o(\delta^2)$, from (4.17) and the explicit formula for $s_{MP,y}(z)$, we still have $s(z) = s_{MP,y}(z) + o(\delta)$.

To summarize the above discussion, one has, with overwhelming probability, either

$$(4.18) \qquad\qquad s(z) = s_{MP,y}(z) + o(\delta),$$

or

$$(4.19) \qquad s(z) = s_{MP,y}(z) - \frac{\sqrt{(y+z-1)^2 - 4yz}}{yz} + o(\delta) = -s_{MP,y}(z) - \frac{y+z-1}{yz} + o(\delta),$$

or

$$(4.20) \qquad\qquad s(z) = -\frac{y+z-1}{yz} + o(1).$$

We may assume the above trichotomy holds for all $z = x + \sqrt{-1}\eta$ with $a \le x \le b$ and $\eta_0 \le \eta \le n^{10}/\delta$ where $\eta_0 = \frac{K^2 \log^6 n}{n\delta^8}$.

When $\eta = n^{10}/\delta$, from $|s(z)| \le 1/\eta$ and $|s_{MP,y}(z)| \le 1/\eta$, we have $s(z)$ and $s_{MP,y}(z)$ are both $o(\delta)$ and therefore (4.18) holds in this case. By continuity, we conclude that either (4.18) holds in the domain of interest or there exists some z in the domain such that (4.18) and (4.19) or (4.18) and (4.20) hold together.

On the other hand, (4.18) or (4.20) cannot hold at the same time. Otherwise, $s_{MP,y}(z) + \frac{y+z-1}{yz} = o(1)$. However, from $s_{MP,y}(z)(s_{MP,y}(z) + \frac{y+z-1}{yz}) = -\frac{1}{yz}$ and $|s_{MP,y}(z)| \le \frac{\sqrt{2}}{\sqrt{y}(1 - \sqrt{y} + \sqrt{\eta_0})}$, one can see that $|s_{MP,y}(z) + \frac{y+z-1}{yz}|$ is bounded from below, which implies a contradiction.

Similarly, (4.18) or (4.19) cannot both hold except when $(y + z - 1)^2 - 4yz = o(\delta^2)$. Otherwise, we can conclude that $2s_{MP,y}(z) + \frac{y+z-1}{yz} = o(\delta)$. From the explicit formula of $s_{MP,y}$,

$$s_{MP,y}(z) + \frac{y+z-1}{yz} = \frac{\sqrt{(y+z-1)^2 - 4yz}}{yz}.$$

One can conclude $|2s_{MP,y}(z) + \frac{y+z-1}{yz}| \ge C\delta$, which contradicts our assertion. Actually, if $(y + z - 1)^2 - 4yz = o(\delta^2)$, (4.18) and (4.19) are equivalent.

In conclusion, (4.18) holds with overwhelming probability in the domain of interest.

5. Gap theorem and Four Moment theorem

In this section, we complete the proofs of the main results, Theorem 1.5 and Theorem 1.7. The proofs follow closely from those in [11] (as well as in [14], [13]), so we shall focus on the changes needed to that argument. We assume substantial familiarity with the materials in [14], [13], and will cite from them repeatedly.

It is convenient to use the *augmented matrix*

$$(5.1) \qquad\qquad \mathbf{M} := \begin{pmatrix} 0 & M^* \\ M & 0 \end{pmatrix}$$

which is a $(p + n) \times (p + n)$ Hermitian matrix with eigenvalues $\pm\sigma_1(M), \ldots, \pm\sigma_p(M)$ and $n - p$ zeros. In this way, we can import the results obtained in [14], [13] and [11] to the model discussed in this paper.

As mentioned in the beginning of Section 5, one can assume that

$$|\zeta_{ij}|, |\zeta'_{ij}| \leq n^{10/C_0}$$

almost surely for all i, j. We also assume that the distributions of M, M' are continuous to ensure the singular values are almost surely simple.

Let us first state a weaker version the Four Moment Theorem as we assume gap properties for the matrices considered:

Theorem 5.1 (Four Moment theorem with Gap assumption). *For sufficiently small $c_0 > 0$ and sufficiently large $C_0 > 0$ ($C_0 = 10^4$ will suffice) the following holds for every $k \geq 1$. Let $M = (\zeta_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ and $M' = (\zeta'_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ be two random matrices satisfying condition **C1** with the indicated constant C_0, and assume that for each i, j that ζ_{ij} and ζ'_{ij} match to order 4. Let W, W' be the associated covariance matrices. Assume also that M, M' obey the gap property and $p/n \to y$ for some $0 < y \leq 1$.*

Let $G : \mathbb{R}^k \to \mathbb{R}$ be a smooth function obeying the derivative bounds

(5.2) $$|\nabla^j G(x)| \leq n^{c_0}$$

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

Then for any $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, and for n sufficiently large depending on k, c_0, we have

(5.3) $$|\mathbf{E}(G(n\lambda_{i_1}(W), \ldots, n\lambda_{i_k}(W))) - \mathbf{E}(G(n\lambda_{i_1}(W'), \ldots, n\lambda_{i_k}(W')))| \leq n^{-c_0}.$$

If ζ_{ij} and ζ'_{ij} only match to order 3 rather 4, then the conclusion (5.3) still holds provided that one strengthens (5.2) to

$$|\nabla^j G(x)| \leq n^{-jc_1}$$

for all $0 \leq j \leq 5$ and $x \in \mathbb{R}^k$ and any $c_1 > 0$, provided that c_0 is sufficiently small depending on c_1.

The Four Moment theorem follows directly from Theorem 1.7 and Theorem 5.1. The next two sections are devoted to the proofs of Theorem 1.7 and Theorem 5.1.

5.1. **Proof of Theorem 5.1.** The key technical step (also used in proving Theorem 1.7) is the truncated Four Moment Theorem, which follows by applying [13, Proposition 6.1 and Proposition 6.2] (or see [11, Proposition 35]) to the *argumented matrix* **M**. The proof is omitted here.

Theorem 5.2 (Truncated four moment theorem). *For sufficiently small $c_0 > 0$ and sufficiently large $C_0 > 0$ ($C_0 = 10^4$ will suffice) the following holds for every $k \geq 1$. Let $M = (\zeta_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ and $M' = (\zeta'_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ be two random matrices satisfying condition **C1** with the indicated constant C_0, and assume that for each i, j that ζ_{ij} and ζ'_{ij} match to order 4. Assume also that $|\zeta_{ij}|, |\zeta'_{ij}| \leq n^{10/C_0}$ and $p/n \to y$ for some $0 < y \leq 1$.*

Let $G : \mathbb{R}^k \times \mathbb{R}^k_+ \to \mathbb{R}$ be a smooth function obeying the derivative bounds

(5.4) $$|\nabla^j G(x_1, \ldots, x_k, q_1, \ldots, q_k)| \leq n^{c_0}$$

for all $0 \leq j \leq 5$ and $x_1, \ldots, x_k \in \mathbb{R}$, $q_1, \ldots, q_k \in \mathbb{R}_+$, and such that G is supported on the region $q_1, \ldots, q_k \leq n^{c_0}$, and the gradient ∇ is in all $2k$ variables.

Then for any $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, and for n sufficiently large depending on ε, k, c_0, we have

$$
\begin{aligned}
(5.5) \quad & |\mathbf{E}(G(\sqrt{n}\sigma_{i_1}(M), \ldots, \sqrt{n}\sigma_{i_k}(M), Q_{i_1}(M), \ldots, Q_{i_k}(M))) \\
& - \mathbf{E}(G(\sqrt{n}\sigma_{i_1}(M'), \ldots, \sqrt{n}\sigma_{i_k}(M'), Q_{i_1}(M'), \ldots, Q_{i_k}(M')))| \leq n^{-c_0}.
\end{aligned}
$$

If ζ_{ij} and ζ'_{ij} only match to order 3 rather 4, then the conclusion (5.5) still holds provided that one strengthens (5.4) to

$$|\nabla^j G(x_1, \ldots, x_k, q_1, \ldots, q_k)| \leq n^{-jc_1}$$

for any $c_1 > 0$, provided that c_0 is sufficiently small depending on c_1.

As in the arguments in Section 6 in [11], we use the qualities for $1 \leq i \leq p$,

$$
\begin{aligned}
Q_i(\mathbf{M}) &:= \sum_{\lambda \neq \sigma_i(M)} \frac{1}{|\sqrt{n}(\lambda - \sigma_i(M))|^2} \\
&= \frac{1}{n}\left(\sum_{1 \leq j \leq p: j \neq i} \frac{1}{|\sigma_j(M) - \sigma_i(M)|^2} + \frac{n-p}{\sigma_i(M)^2} + \sum_{j=1}^{p} \frac{1}{|\sigma_j(M) + \sigma_i(M)|^2} \right).
\end{aligned}
$$

The gap property (up to the edge) on M ensures an upper bound on $Q_i(\mathbf{M})$. The proof repeats exactly the proof of Lemma 32 in [11].

Proposition 5.3. *If M satisfies the gap property, then for any $c_0 > 0$(independent of n), and any $1 \leq i \leq p$, one has $Q_i(\mathbf{M}) \leq n^{c_0}$ with high probability.*

Now Theorem 5.1 follows by defining $\tilde{G} : \mathbb{R}^k \times \mathbb{R}^k_+ \to \mathbb{R}$ to be

$$\tilde{G}(\sqrt{n}\sigma_{i_1}, \ldots, \sqrt{n}\sigma_{i_k}, Q_{i_1}, \ldots, Q_{i_k}) := G(\sqrt{n}\sigma_{i_1}, \ldots, \sqrt{n}\sigma_{i_k}) \prod_{j=1}^{k} \eta(Q_{i_j})$$

where $\eta(x)$ is a smooth cutoff to the region $x \leq n^{c_0}$ which equals 1 on $x \leq n^{c_0}/2$. From Propositon 5.3, we have

$$|\mathbf{E}(G(\sqrt{n}\sigma_{i_1}(M), \ldots, \sqrt{n}\sigma_{i_k}(M))) - \mathbf{E}(\tilde{G}(\sqrt{n}\sigma_{i_1}(M), \ldots, \sqrt{n}\sigma_{i_k}(M), Q_{i_1}(\mathbf{M}), \ldots, Q_{i_k}(\mathbf{M})))| \ll \frac{1}{n^c},$$

for some $c > 0$, and a similar relation holds for M'. The proof is complete by using the above relations and Theorem 5.2.

5.2. **Proof of the Gap Theorem.** We first have a gap theorem under additional exponential decay hypothesis on the ensembles of M. The proof is presented in Section 5.3.

Theorem 5.4 (Gap theorem up to the edge). *Let $M = (\zeta_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ be a random matrix obeying condition* **C1**, *and the entries ζ_{ij} satisfy exponential decay in the sense that $\mathbf{P}(|\zeta_{ij}| \geq t^C) \leq \exp(-t)$ for all $t \geq C'$ for all i, j and some constants $C, C' > 0$. Then M obeys the gap property.*

The next observation is the following matching lemma (See Lemma 33 in [11]), which together with Theorem 5.2, ensures us to remove the exponential decay hypothesis in Theorem 5.4.

Lemma 5.5 (Matching lemma,[11])**.** *Let ζ be a complex random variable with mean zero, unit variance, and third moment bounded by some constant a. Then there exists a complex random variable $\tilde{\zeta}$ with support bounded by the ball of radius $O_a(1)$ centered at the origin (and in particular, obeying the exponential decay hypothesis uniformly in ζ for fixed a) which matches ζ to third order.*

Now consider the matrix $M = (\zeta_{ij})_{1 \le i \le p, 1 \le j \le n}$ in Theorem 1.7. By the matching lemma, we can find a random matrix $M' = (\zeta'_{ij})_{1 \le i \le p, 1 \le j \le n}$ such that ζ'_{ij} satisfies the exponential decay hypothesis and ζ'_{ij} matches ζ_{ij} to third order for each i, j. By Theorem 5.4, the matrix M' obeys the gap property. Similarly as in Section 6, [11], let $\eta(x)$ be a smooth cutoff to the region $x \le n^{c_0}$. Then by Proposition 5.3, $\mathbf{E}\eta(Q_i(M')) = 1 - O(n^{-c_1})$, which, by using Theorem 5.2, implies that $\mathbf{E}\eta(Q_i(M')) = 1 - O(n^{-c_2})$ for some c_2 independent of n. Hence, M also satisfies the gap property.

5.3. **Proof of Theorem 5.4:** The proof follows closely to that discussed in Section 5, [13]. We shall mainly mention the corresponding changes. Interested readers can find the detailed proofs in [14]. First, in order to operate an induction argument, we need to treat the edge case $i = 1, p$ separately.

Proposition 5.6 (Extreme cases)**.** *Theorem 5.4 is true when $i = 1$ or $i = p$.*

Proof. By symmetry, it suffices to show for $i = p$. In the interlacing identity (Lemma 3.5),

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

On the right-hand side of the above identity, $Y \in \mathbb{C}^n$, by Lemma 3.7, $||Y||^2/n = 1 + o(1)$ with overwhelming probability. And $\sigma_p(M_{p,n})^2/n = \lambda_p(W) = b + o(1)$ with overwhelming probability. All the terms in the left-hand side is are negative signs, thus

$$\frac{\frac{1}{n}\sigma_j(M_{p-1,n})^2 |u_j(M_{p-1,n})^* Y|^2}{|\sigma_j(M_{p-1,n})^2 - \sigma_p(M_{p,n})^2|} \ll 1.$$

From Theorem 4.2 and Lemma 3.8, one can conclude that $|u_{p-1}(M_{p-1,n})^* Y|^2 \ge n^{-c/10}$ with high probability. Therefore, $|\sigma_j(M_{p-1,n})^2 - \sigma_p(M_{p,n})^2| \ge n^{-c}$ with high probability. The conclusion follows from the Cauchy interlacing law. $\qquad\square$

For the general case for the gap theorem, we write i_0, p_0 instead of i, p and define $N_0 := p_0 + n$, as in [14], we introduce the *regularized gap*

$$(5.6) \qquad g_{i,l,p} := \inf_{1 \le i_- \le i - l < i \le i_+ \le p} \frac{\sqrt{N_0}\sigma_{i_+}(M_{p,n}) - \sqrt{N_0}\sigma_{i_-}(M_{p,n})}{\min(i_+ - i_-, \log^{C_1} N_0)^{\log^{0.9} N_0}},$$

where $C_1 > 1$ is a large constant to be determined later. To show Theorem 5.4, it is enough to show that

$$g_{i_0,1,p_0} \le n^{-c_0},$$

for $1 < i_0 < p_0$.

By repeating the arguments in Section 3.5, [14], the proof relies on the following two key propositions. The idea is to propagate a narrow gap for $M_{p,n}$ backwards in p until one can use Theorem 4.1 to control the occurrence of the gap.

Proposition 5.7 (Backwards propagation of gap). *. Suppose $p_0/2 \le p < p_0$ and $l \le p/10$ is such that*

$$(5.7) \qquad\qquad g_{i_0,l,p+1} \le \delta$$

for some $1 < \delta \le 1$(which can depend on n), and that

$$(5.8) \qquad\qquad \sqrt{n}\sigma_{p+1}(M_{p+1,n}) - \sqrt{n}\sigma_p(M_{p+1,n}) \ge \delta \exp(\log^{0.91} n_0).$$

Then $i_0 \le p$. Suppose further that

$$g_{i_0,l+1,p} \ge 2^m g_{i_0,l,p+1}$$

for some $m \ge 0$ with

$$2^m \le \delta^{-1/2}.$$

Let X_{p+1} be the $p+1^{th}$ row of $M_{p_0,n}$, and let $u_1(M_{p,n}), \ldots, u_p(M_{p,n})$ be an orthonormal system of right singular vectors of $M_{p,n}$ associated to $\sigma_1(M_{p,n}), \ldots, \sigma_p(M_{p,n})$. Then one of the following statement holds:

* (i) *(Macroscopic spectral concentration) There exists $1 \le i_- < i_+ \le p+1$ with $i_+ - i_- \ge \log^{C_1/2} n$ such that $|\sqrt{n}\sigma_{i_+}(M_{p+1,n}) - \sqrt{n}\sigma_{i_-}(M_{p+1,n})| \le \delta^{1/4} \exp(\log^{0.95})(i_+ - i_-).$*
* (ii) *(Small inner products) There exists $1 \le i_- \le i_0 - l < i_0 \le i_+ \le p$ with $i_+ - i_- \le \log^{C_1/2} n$ such that*
$$\sum_{i_- \le j < i_+} |X_{p+1}^* u_j(M_{p,n})|^2 \le \frac{i_+ - i_-}{2^{m/2} \log^{0.01} n}.$$
* (iii) *(Large singular value) For some $1 \le i \le p+1$ one has*
$$|\sigma_i(M_{p+1,n})| \ge \frac{\sqrt{n} \exp(-\log^{0.96} n)}{\delta^{1/2}}.$$
* (iv) *(Large inner product) There exists $1 \le i \le p$ such that*
$$|X_{p+1}^* u_i(M_{p,n})|^2 \ge \frac{\exp(-\log^{0.96} n)}{\delta^{1/2}}.$$
* (v) *(Large row) We have*
$$||X_{p+1}||^2 \ge \frac{n \exp(-\log^{0.96} n)}{\delta^{1/2}}.$$
* (vi) *(Large inner product near i_0) There exists $1 \le i \le p$ with $|i - i_0| \le \log^{C_1} n$ such that*
$$|X_{p+1}^* u_i(M_{p,n})|^2 \ge 2^{m/2} \log^{0.8} n.$$

Proof. Apply Lemma 5.3 in [13] to the *augmented matrix*

$$A_{p+n+1} := \sqrt{n} \begin{pmatrix} 0 & M_{p+1,n}^* \\ M_{p+1,n} & 0 \end{pmatrix}$$

and

$$A_{p+n} := \sqrt{n} \begin{pmatrix} 0 & M_{p,n}^* \\ M_{p,n} & 0 \end{pmatrix}.$$

Noticed A_{p+n} is A_{p+n+1} with the rightmost column and bottom column(which is X_{p+1} and $p+1$ zeros) removed. The eigenvalues of A_{p+n} are $\pm\sqrt{n}\sigma_1(M_{p,n}), \ldots, \pm\sqrt{n}\sigma_p(M_{p,n})$ and 0, and an orthonormal eigenbasis includes the vectors $\begin{pmatrix} u_j(M_{p,n}) \\ v_j(M_{p,n}) \end{pmatrix}$ for $1 \le j \le p$. (The "Large coefficient" event in Lemma 5.3, [13]) cannot occur as A_{p+n+1} has zero diagonals.) $\qquad\square$

The next proposition claims that the events (i)-(vi) occurs with small probability.

Proposition 5.8 (Bad events are rare)**.** *Suppose that $p_0/2 \leq p < p_0$ and $l \leq p/10$ and set $\delta := n_0^{-\kappa}$ for some sufficiently small fixed $\kappa > 0$. Then*

(a) *The events (i), (iii), (iv), (v) in Proposition 5.7 all fail with high probability.*

(b) *There is a constant C' such that all the coefficients of the right singular vectors $u_j(M_{p,n})$ for $1 \leq j \leq p$ are of magnitude at most $n^{-1/2} \log^{C'} n$ with overwhelming probability. Conditioning $M_{p,n}$ to be a matrix with this property, the events (ii) and (vi) occur with a conditional probability of at most $2^{-\kappa m} + n^{-\kappa}$.*

(c) *Furthermore, there is a constant C_2 (depending on C', κ, C_1) such that if $l \geq C_2$ and $M_{p,n}$ is conditioned as in (b), then (ii) and (vi) in fact occur with a conditional probability of at most $2^{-\kappa m} \log^{-2C_1} n + n^{-\kappa}$.*

The proof of the above proposition repeats the proof of Proposition 53 in [14] with the major difference being that Theorem 4.1 and Theorem 4.2 are applied instead of Theorem 60 and Proposition 62 in [14].

References

[1] 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.

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

[3] A. Borodin and P.J. Forrester. Increasing subsequences and the hard-to-soft edge transition in matrix ensembles. *Journal of Physics A: Mathematical and General*, 36:2963, 2003.

[4] A. Edelman. *Eigenvalues and condition numbers of random matrices*. PhD thesis, Massachusetts Institute of Technology, 1989.

[5] 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.

[6] L. Erdös, B. Schlein, and H.T. Yau. Universality of random matrices and local relaxation flow. *Submitted to Inv. Math. arxiv. org/abs/0907.5605*, 2009.

[7] L. Erdös, B. Schlein, and H.T. Yau. Local semicircle law and complete delocalization for Wigner random matrices. Accepted in Comm. Math. Phys. *Communications in Mathematical Physics*, 287(2):641–655, 2010.

[8] L. Erdös, B. Schlein, H.T. Yau, and J. Yin. The local relaxation flow approach to universality of the local statistics for random matrices. *To appear in Annales Inst. H. Poincaré (B), Probability and Statistics. Arxiv preprint arXiv:0911.3687*, 2009.

[9] O.N. Feldheim and S. Sodin. A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geometric And Functional Analysis*, 20(1):88–123, 2010.

[10] VA Marvcenko and LA Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1:457, 1967.

[11] T. Tao and V. Vu. Random covariance matrices: Universality of local statistics of eigenvalues. *Arxiv preprint arXiv:0912.0966*, 2009.

[12] T. Tao and V. Vu. Random matrices: The distribution of the smallest singular values. *Geometric And Functional Analysis*, 20(1):260–297, 2010.

[13] T. Tao and V. Vu. Random matrices: Universality of local eigenvalue statistics up to the edge. *Communications in Mathematical Physics*, pages 1–24, 2010.

[14] T. Tao and V. Vu. Random matrices: Universality of local eigenvalue statistics. *Acta Mathematica*, pages 1–78, 2011.

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, US

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