OPTIMAL UPPER BOUND FOR THE INFINITY NORM OF EIGENVECTORS OF RANDOM MATRICES

BY KE WANG

A dissertation submitted to the
Graduate School—New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
Graduate Program in Mathematics

Written under the direction of
Professor Van Vu
and approved by

________________________________________

________________________________________

________________________________________

________________________________________

New Brunswick, New Jersey
May, 2013
Let $M_n$ be a random Hermitian (or symmetric) matrix whose upper diagonal and diagonal entries are independent random variables with mean zero and variance one. It is well known that the empirical spectral distribution (ESD) converges in probability to the semicircle law supported on $[-2, 2]$. In this thesis we study the local convergence of ESD to the semicircle law. One main result is that if the entries of $M_n$ are bounded, then the semicircle law holds on intervals of scale $\log n/n$. As a consequence, we obtain the delocalization result for the eigenvectors, i.e., the upper bound for the infinity norm of unit eigenvectors corresponding to eigenvalues in the bulk of spectrum, is $O(\sqrt{\log n/n})$. The bound is the same as the infinity norm of a vector chosen uniformly on the unit sphere in $\mathbb{R}^n$. We also study the local version of Marchenko-Pastur law for random covariance matrices and obtain the optimal upper bound for the infinity norm of singular vectors. This is joint work with V. Vu.

In the last chapter, we discuss the delocalization properties for the adjacency matrices of Erdős-Rényi random graph. This is part of some earlier results joint with L. Tran and V. Vu.
Acknowledgements

First and foremost, I would like to thank my advisor, Professor Van Vu, for his valuable guidance and essential support over the years. He introduced me to the fascinating world of random matrix theory and opened the door to research for me. His enthusiasm for research, immense knowledge and logical way of thinking have influenced and inspired me greatly. I feel extremely fortunate to have him as my advisor.

I would also like to express my gratitude to the other members of my thesis committee: Professor Michael Kiessling, Professor Swastik Kopparty, and Professor Alexander Soshnikov. Their valuable time, feedback and support are greatly appreciated.

It is a pleasure for me to thank the Department of Mathematics of Rutgers University. I was fortunate to encounter many faculty members here who have motivated me, shaped my thinking and deeply influenced my future.

Furthermore, I would like to acknowledge Hoi Nguyen, Sean O’Rourke, Linh Tran and Gabriel Tucci for many useful conversations, suggestions and collaboration. Each of them deserves more thanks that I can ever express.

I also want to thank my friends at Rutgers for their encouragement and support. Their care and friendship helped me adjust to a new country.

Last but not the least, I express my gratitude towards my parents for constant love and support. The most special thanks goes to my dear husband, Tianling Jin, for enlightening my life with his presence. He is one of the best things that’s ever happened to me.
Dedication

To my parents, Zhiyuan Wang and Yanming Zhao.

To my husband, Tianling Jin.
# Table of Contents

Abstract ................................................................. ii

Acknowledgements ..................................................... iii

Dedication ................................................................. iv

Terminology ............................................................... vii

1. Preliminaries ......................................................... 1
   1.1. Random matrices ............................................... 1
   1.2. Some concentration inequalities ............................... 2
       1.2.1. Chernoff bound ......................................... 3
       1.2.2. Azuma’s inequality .................................... 3
       1.2.3. Talagrand’s inequality ................................. 4
       1.2.4. Hanson-Wright inequality ............................. 6

2. Random Hermitian matrices ....................................... 10
   2.1. Semicircle law ................................................. 10
       2.1.1. Moment method ........................................ 12
       2.1.2. Stieltjes transform method ........................... 13
   2.2. Local semicircle law and the new result .................... 14
       2.2.1. Proof of Theorem 19 .................................. 16
       2.2.2. Proof of Lemma 23 .................................... 24
   2.3. Optimal upper bound for the infinity norm of eigenvectors . . 27
       2.3.1. Proof of the bulk case ................................. 28
       2.3.2. Proof of the edge case ............................... 30

3. Random covariance matrices ...................................... 34
Terminology

Asymptotic notation is used under the assumption that \( n \to \infty \). For functions \( f \) and \( g \) of parameter \( n \), we use the following notation as \( n \to \infty \):

- \( f = O(g) \) if \( |f|/|g| \) is bounded from above;
- \( f = o(g) \) if \( f/g \to 0 \);
- \( f = \omega(g) \) if \( |f|/|g| \to \infty \), or equivalently, \( g = o(f) \);
- \( f = \Omega(g) \) if \( g = O(f) \);
- \( f = \Theta(g) \) if \( f = O(g) \) and \( g = O(f) \).

The expectation of a random variable \( X \) is denoted by \( \mathbf{E}(X) \) and \( \text{Var}(X) \) denotes its variance.

We use \( 1_A \) for the characteristic function of a set \( A \) and \( |A| \) for its cardinality. For a vector \( x = (x_1, \ldots, x_n) \in \mathbb{C}^n \), the 2-norm is

\[
\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}
\]

and the infinity norm is

\[
\|x\|_{\infty} = \max_i |x_i|.
\]

For an \( n \times n \) matrix \( M = (M_{ij})_{1 \leq i,j \leq n} \), we denote the trace

\[
\text{trace}(M) = \sum_{i=1}^{n} M_{ii},
\]

the spectral norm

\[
\|M\|_2 = \sup_{x \in \mathbb{C}^n, \|x\|_2 = 1} |Mx|,
\]

and the Frobenius norm

\[
\|M\|_F = \sqrt{\sum_{i,j=1}^{n} |M_{ij}|^2}.
\]
Chapter 1

Preliminaries

1.1 Random matrices

Random matrices was introduced by Wishart [72] in 1928 in mathematical statistics and starts to gain more attention after Wigner [70] used them as a prominent tool in studying level spacing distributions of heavy nuclei in a complex nuclear system in the fifties. A series of beautiful work have been established by Wigner, Mehta [47] and Dyson [23, 24, 25, 22, 26] shortly after. Since then the subject of random matrix theory has been developing deeper and more far reaching, not only because it is connected to systems such as nuclear physics, quantum chaos [14], zeros of Riemann $\zeta$ functions (see [13] and the reference therein) and etc but also finds many applications in areas as varied as multivariate statistics and component analysis [42, 43], wireless communication [68] and numerical analysis [27].

A major topic in random matrix theory is the universality conjecture, which asserts under certain conditions on the entries, the local-scale distribution of eigenvalues of random matrices obeys the same asymptotic laws regardless of the distribution of entries.

The celebrated Wigner’s semicircle law [71] is a universal result in the sense that the eigenvalue distribution of Hermitian matrices with iid entries is independent of the underlying distribution of the entries. The goal is to study the limiting spectral behavior of random matrices as the matrix size tends to infinity. Consider the empirical spectral distribution (ESD) function of an $n \times n$ Hermitian matrix $W_n$, which is a one-dimensional function

$$F^{W_n}(x) = \frac{1}{n} \left| \{1 \leq j \leq n : \lambda_j(W) \leq x \} \right|.$$
**Theorem 1** (Wigner’s semicircle law, [71]). Let $M_n = (\xi_{ij})_{1 \leq i, j \leq n}$ be an $n \times n$ random symmetric matrix whose entries satisfy the conditions:

- The distribution law for each $\xi_{ij}$ is symmetric;
- The entries $\xi_{ij}$ with $i \leq j$ are independent;
- The variance of each $\xi_{ij}$ is 1;
- For every $k \geq 2$, there is a uniform bound $C_k$ on the $k^{th}$ moment of each $\xi_{ij}$.

Then the ESD of $W_n = \frac{1}{\sqrt{n}} M_n$ converges in probability to the semicircle law with density function

$$\frac{1}{2\pi} \sqrt{4 - x^2}$$

that is supported on $[-2, 2]$.

Other work regarding the universality of spectral properties includes those regarding the edge spectral distributions for a large class of random matrices, see [19], [55, 57, 50], [49] and [41] for instance. There are also universality type of results for the random covariance matrices (will be defined in Chapter 3), for example, [11], [56] and [7].

More recently, major breakthroughs on Wigner matrices have been made by Erdős, Schlein, Yau, Yin [29, 30, 31, 28] and Tao, Vu [66, 64]. The conclusion, roughly speaking, asserts that the general local spectral statistics (say the largest eigenvalue, the spectral gap etc) are universal, i.e. it follows the statistics of the corresponding Gaussian ensembles, depending on the symmetry type of the matrix. The methods are also developed to handle covariance matrices [62, 32, 51, 69].

In particular, the local semicircle law lies in the heart of understanding the individual eigenvalue position and deriving the universality results. Our results refine the previous ones obtained in the references mentioned above, and the proof strategies are adapted from those.

### 1.2 Some concentration inequalities

Concentration inequalities estimate the probability that a random variable deviates from some value (usually its expectation) and play an important role in the random matrix theory. The most basic example is the law of large numbers, which states...
that under mild condition, the sum of independent random variables are around the expectation with large probability. In the following, we collect a few concentration inequalities that are used in this paper or frequently used in related references.

1.2.1 Chernoff bound

Chernoff bound gives exponentially decreasing bounds on tail distribution for the sum of iid bounded random variables.

**Theorem 2** (Theorem 2.3, [17]). Let $X_1,\ldots,X_n$ be iid random variables with $\mathbb{E}(X_i) = 0$ and $\text{Var}(X_i) = \sigma^2$. Assume $|X_i| \leq 1$. Let $X = \sum_{i=1}^{n} X_i$, then

$$
P(|X| \geq \epsilon \sigma) \leq 2e^{-\epsilon^2/4},$$

for any $0 \leq \epsilon \leq 2\sigma$.

A more generalized version is the following

**Theorem 3** (Theorem 2.10 and Theorem 2.13, [17]). Let $X_1,\ldots,X_n$ be independent random variables. And let $X = \sum_{i=1}^{n} X_i$.

- If $X_i \leq \mathbb{E}(X_i) + a_i + M$ for $1 \leq i \leq n$, then one has the upper tail

  $$
P(X \geq \mathbb{E}(X) + \lambda) \leq e^{-\frac{\lambda^2}{2(\text{Var}(X) + \sum_{i=1}^{n} a_i^2 + M\lambda/3)}}.$$

- If $X_i \geq \mathbb{E}(X_i) - a_i - M$ for $1 \leq i \leq n$, then one has the lower tail

  $$
P(X \leq \mathbb{E}(X) - \lambda) \leq e^{-\frac{\lambda^2}{2(\text{Var}(X) + \sum_{i=1}^{n} a_i^2 + M\lambda/3)}}.$$

1.2.2 Azuma’s inequality

If the random variables $X_i$ are not jointly independent, one may refer to the Azuma’s inequality if $\{X_i\}$ is a $c$-Lipschitz martingale introduced in Chapter 2, [17].

A martingale is a sequence of random variables $\{X_1, X_2, X_3, \ldots\}$ that satisfies $\mathbb{E}(|X_i|) \leq \infty$ and the conditional expectation

$$
\mathbb{E}(X_{n+1}|X_1,\ldots,X_n) = X_n.
$$
For a vector of positive entries $c = (c_1, \ldots, c_n)$, a martingale is said to be $c$-Lipschitz if

$$|X_i - X_{i-1}| \leq c_i,$$

for $1 \leq i \leq n$.

**Theorem 4** (Theorem 2.19, [17]). If a martingale $\{X_1, X_2, X_3, \ldots, X_n\}$ is $c$-Lipschitz for $c = (c_1, \ldots, c_n)$. Let $X = \sum_{i=1}^{n} X_i$, then

$$\mathbb{P}(|X - \mathbb{E}(X) | \geq \lambda) \leq 2 e^{-\frac{\lambda^2}{2 \sum_{i=1}^{n} c_i^2}}.$$

In particular, for independent random variables $X_i$, one has the following from Azuma’s inequality.

**Theorem 5** (Theorem 2.20, [17]). Let $X_1, \ldots, X_n$ be independent random variables that satisfy

$$|X_i - \mathbb{E}(X_i)| \leq c_i,$$

for $1 \leq i \leq n$. Let $X = \sum_{i=1}^{n} X_i$. Then

$$\mathbb{P}(|X - \mathbb{E}(X) | > \lambda) \leq 2 e^{-\frac{\lambda^2}{2 \sum_{i=1}^{n} c_i^2}}.$$

### 1.2.3 Talagrand’s inequality

Let $\Omega = \Omega_1 \times \ldots \times \Omega_n$ be a product space equipped with product probability measure $\mu = \mu_1 \times \ldots \times \mu_n$. For any vector $w = (w_1, \ldots, w_n)$ with non-negative entries, the weighted Hamming distance between two points $x, y \in \Omega$ is defined as

$$d_w(x, y) = \sum_{i=1}^{n} w_i 1\{x_i \neq y_i\}.$$

For any subset $A \subset \Omega$, the distances are defined as

$$d_w(x, A) = \inf_{y \in A} d_w(x, y)$$

and

$$D(x, A) = \sup_{w \in W} d_w(x, A),$$

where

$$W := \{w = (w_1, \ldots, w_n) | w_i \geq 0, \|w\| \leq 1\}.$$
Talagrand investigated the concentration of measure phenomena in product space: for any measurable set $A \subset \Omega$ with $\mu(A) > 1/2$ (say), almost all points are concentrated within a small neighborhood of $A$.

**Theorem 6 ([60]).** For any subset $A \subset \Omega$, one has

$$\mu(\{x \in \Omega | D(x, A) \geq t\}) \leq \frac{e^{-t^2/4}}{\mu(A)},$$

for any $t > 0$.

Talagrand’s inequality turns out to be rather powerful in combinatorial optimizations and many other areas. See [60], [46] and [58] for more examples. One striking consequence is the following version for independent uniformly bounded random variables.

**Theorem 7** (Talagrand’s inequality,[60]). *Let $D$ be the unit disk $\{z \in \mathbb{C}, |z| \leq 1\}$. For every product probability $\mu$ supported on a dilate $K \cdot D^n$ of the unit disk for some $K > 0$, every convex 1-Lipschitz function $F : \mathbb{C}^n \to \mathbb{R}$ and every $t \geq 0$,

$$\mu(|F - M(F)| \geq t) \leq 4 \exp(-t^2/16K^2),$$

where $M(F)$ denotes the median of $F$.*

One important application of Talagrand’s inequality in random matrix theory is a result by Guionnet and Zeitouni in [39]. Consider a random Hermitian matrix $W_n$ with independent entries $w_{ij}$ with support in a compact region $S$, say $|w_{ij}| \leq K$. Let $f$ be a real convex $L$-Lipschitz function and define

$$Z := \sum_{i=1}^{n} f(\lambda_i),$$

where $\lambda_i$’s are the eigenvalues of $\frac{1}{\sqrt{n}}W_n$. We are going to view $Z$ as the function of the variables $w_{ij}$.

The next concentration inequality is an extension of Theorem 1.1 in [39] (see also Theorem F.5 [63]).

**Lemma 8.** *Let $W_n, f, Z$ be as above. Then there is a constant $c > 0$ such that for any $T > 0$

$$\mathbb{P}(|Z - \mathbb{E}(Z)| \geq T) \leq 4 \exp(-c \frac{T^2}{K^2L^2}).$$
1.2.4 Hanson-Wright inequality

The Hanson-Wright inequality [40] controls the quadratic forms in random variables and appears to be quite useful in studying random matrices. A random variable $X$ with mean $\lambda$ is said to be sub-gaussian if there exists constants $\alpha, \gamma > 0$ such that

$$P(|X - \lambda| \geq t) \leq \alpha e^{-\gamma t^2}. \quad (1.1)$$

For random variables with heavier tails than the gaussian, like the exponential distribution, we can define sub-exponential random variable $X$ with mean $\lambda$ if there exists constants $\alpha, \gamma > 0$ such that

$$P(|X - \lambda| \geq t) \leq \alpha e^{-\gamma t}. \quad (1.2)$$

A random variable $X$ is sub-gaussian if and only if $X^2$ is sub-exponential.

**Theorem 9** (Hanson-Wright inequality). If $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ is symmetric and $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is a random vector with $x_i$ independent with mean zero, variance one and sub-gaussian with constants $\alpha, \gamma$ as in (1.1). Let $B = (|a_{ij}|)$, then there exist constants $C, C' > 0$ such that

$$P(|x^T Ax - \text{trace}(A)| \geq t) \leq C e^{-\min\{C' t^2/\|A\|^2_F, C' t/\|B\|_2\}} \quad (1.3)$$

for any $t > 0$.

In Hanson and Wright’s paper [40], the random variables are assumed to be symmetric. Later, Wright [73] extends the result to non-symmetric random variables. We record a proof for the sake of completeness.

**Proof.** First, we can assume that $a_{ii} = 0$ for every $i$. Otherwise, if $a_{ii} \neq 0$ for some $i$, consider the diagonal matrix $D = \text{diag}(a_{11}, \ldots, a_{nn})$ and the matrix $A_1 = A - D$. Thus

$$x^T Ax - \text{trace}(A) = x^T A_1 x + \sum_{i=1}^{n} a_{ii}(x_i^2 - 1).$$

Since $x_i$ are sub-gaussian random variables, $x_i^2 - 1$ are independent mean-zero sub-exponential random variables. By Bernstein’s inequality, there exists constant $C_1$ depending on $\alpha, \gamma$ such that

$$P\left(\sum_{i=1}^{n} a_{ii}(x_i^2 - 1) > t\right) \leq 2 \exp\left(-C_1 \min\left(\frac{t^2}{\sum_i a_{ii}^2}, \frac{t}{\max_i |a_{ii}|}\right)\right).$$
On the other hand, $\|A\|_F^2 \geq \sum_i a^2_{ii}$ and $\|B\|_2 \geq \max_i |a_{ii}|$. Notice also that $\|A\|_F \geq \|A_1\|_F$ and $\|B\|_2 \geq \|B_1\|_2$ where $B_1 = B - \text{diag}(|a_{11}|, \ldots, |a_{nn}|)$. Thus it is enough to show that (1.3) holds for the matrix $A_1$, a matrix with zero diagonal entries.

Now, under our assumption, $E(x^T Ax) = \text{trace}(A) = 0$. Let us first consider the case that $x_i$'s have symmetric distribution. By Markov's inequality, for $\lambda > 0$, we have

$$P(x^T Ax > t) \leq e^{-\lambda t}E(\exp(\lambda x^T Ax)).$$

Let $y = (y_1, \ldots, y_n)^T$ be a vector of independent standard normal random variables. Assume $y$ and $x$ are independent. The idea is to show there exists a constant $C_1$ that depends only on $\alpha, \gamma$ as in (1.1) such that

$$E(\exp(\lambda x^T Ax)) \leq E(\exp(C_1 \lambda y^T By)). \quad (1.4)$$

This can be proved by observing

$$E(\exp(\lambda x^T Ax)) = \sum_{k=0}^{\infty} \frac{\lambda^k E(x^T Ax)^k}{k!} = \sum_{k=0}^{\infty} \frac{\lambda^k E(\sum_{i,j=1}^n a_{ij} x_i x_j)^k}{k!}.$$ 

Since the $x_i$'s have symmetric distribution, in the expansion of $E(\sum_{i,j=1}^n a_{ij} x_i x_j)^k$, only the terms contain (the product of) $E(x_i^{2s})$ for some integer $s \geq 1$ are nonzero. We can use a change of variables to bound

$$E(x_i^{2s}) \leq \int_0^\infty x^{2s}e^{-\gamma x^2}dx = \frac{\alpha \sqrt{\pi}}{2 \sqrt{\gamma} (2\gamma)^s} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} y^{2s}e^{-y^2/2}dy \leq C_2^s E(y_i^{2s}),$$

for some $C_1$ depending on $\alpha$ and $\gamma$. By triangle inequality, (1.4) holds.

Since the matrix $B$ is symmetric, it can be decomposed as $B = U^T \Lambda U$ where $U$ is an $n \times n$ orthogonal matrix and $\Lambda$ is diagonal matrix with $\mu_1, \ldots, \mu_n$, the eigenvalues of $B$ in the diagonal entries. And

$$\sum_{i=1}^n a_{ii} = \sum_{i=1}^n \mu_i = 0, \quad \sum_{i=1}^n \mu_i^2 = \text{trace}(B)^2 = \|A\|_F^2.$$

Let $z = Uy = (z_1, \ldots, z_n)$. Then $z_i$'s are iid standard normal. And $y^T By = z^T \Lambda z = \sum_{i=1}^n \mu_i z_i^2 = \sum_{i=1}^n \mu_i (z_i^2 - 1)$, where $z_i^2 - 1$ are independent mean-zero $\chi^2$-random variables of freedom one. By direct computation, there exists a constant $C'$
such that $\mathbb{E}[\exp(C_1 \lambda \mu_i(z_i^2 - 1))] \leq \exp(C \lambda^2 \mu_i^2)$ for sufficiently small $\lambda$. Thus

$$\mathbb{E}[\exp(\lambda x^T Ax)] \leq \mathbb{E}[\exp(C_1 \lambda y^T By)] = \prod_{i=1}^n \mathbb{E}[\exp(C_1 \lambda \mu_i(z_i^2 - 1))] \leq \prod_{i=1}^n \exp(C \lambda^2 \mu_i^2) = \exp(C \lambda^2 \|A\|_F^2).$$

Therefore,

$$\mathbb{P}(x^T Ax > t) \leq e^{-\lambda t} \mathbb{E}(\exp(\lambda x^T Ax)) \leq e^{-\lambda t + C \lambda^2 \|A\|_F^2}.$$

Choose $\lambda_0 = \min\left(\frac{tC'}{\|A\|_F^2}, \frac{C'}{\|B\|_F^2}\right)$ for some constant $C'$ such that

$$e^{-\lambda_0 t + C \lambda_0^2 \|A\|_F^2} \leq e^{-\lambda_0 t/2}.$$

This completes the proof for the case that the random variables have symmetric distribution.

For the general case when the distributions of $x_i$'s are not necessarily symmetric. We use a coupling technique. Take independent random vector copies $x^k = (x^k_1, \ldots, x^k_n)$ for $k = 1, 2, 3, 4$ that have the same distribution as $x$. Then $X = (x_i - x_i^1)^n_{i=1}$ is a vector of independent symmetric, sub-gaussian random variables. Thus (1.3) holds for random vector $X$. And

$$x^T Ax + x^{1T} Ax^{1} = X^T AX + 2x^T Ax^{1}. \quad (1.5)$$

For the term $x^T Ax^{1}$,

$$\mathbb{P}(x^T Ax^{1} > t) \leq e^{-\lambda t} \mathbb{E}[\exp(\lambda x^T Ax^{1})].$$

Let $\mathbb{E}_x(\cdot)$ denote the expectation conditioning on $x$. By Jensen’s inequality,

$$\mathbb{E}[\exp(-\lambda x^T Ax^{k})] \geq \exp(\mathbb{E}[\lambda x^T Ax^{k}]) = 1,$$

thus

$$\mathbb{E}[\exp(\lambda x^T Ax^{1})] = \mathbb{E}\left(\mathbb{E}_x[\exp(\lambda x^T Ax^{1})]\right) \leq \mathbb{E}\left(\mathbb{E}_x[\exp(\lambda x^T A(x_1^1 - x^2))]\right)$$

$$= \mathbb{E}[\exp(\lambda x^T A(x_1^1 - x^2))] = \mathbb{E}\left(\mathbb{E}_{x^1,x^2}[\exp(\lambda x^T A(x_1^1 - x^2))]\right)$$

$$\leq \mathbb{E}\left(\mathbb{E}_{x^1,x^2}[\exp(\lambda (x - x^3)^T A(x_1^1 - x^2))]\right) = \mathbb{E}[\exp(C \lambda y^T By')],$$
for some sufficiently large $C$ depending on $\alpha, \gamma$. The $y, y'$ in the last inequality are independent vectors of independent standard normal random variables. And the last inequality follows similar to the proof of (1.4) by a Taylor expansion since now the vectors $x - x^3, x^1 - x^2$ are symmetric and sub-gaussian.

Factor $B = U^T \Lambda U$. Then $y^T By' = (Uy)^T \Lambda U y' := z^T Az' = \sum_{i=1}^n \mu_i z_i z'_i$, where $z, z'$ are independent random vectors and the entries are standard normal. By direct computation or use Bernstein’s inequality (notice that $z_i z'_i$ are mean-zero sub-exponential), we can prove that

$$
P(x^T A x^1 > t) \leq e^{-C_1 \min\{t^2/\|A\|_F^2, t/\|B\|_2\}}.$$

Therefore, from (1.5),

$$
P(x^T A x > t) = \sqrt{\p(x^T A x > t, x^1 T A x^1 > t)} \leq \sqrt{\p(x^T A x + x^1 T A x^1 > 2t)}$$

$$= \sqrt{\p(X^T A X + 2x^T A x^1 > 2t)} \leq \p(X^T A X > t)^{1/2} \p(2x^T A x^1 > t)^{1/2}$$

$$\leq C \exp\left(-C' \min\left(\frac{t^2}{\|A\|_F^2}, \frac{t}{\|B\|_2}\right)\right),$$

for some constants $C$ and $C'$.

For the upper bound $\p(x^T A x < -t) = \p(x^T (-A)x > t)$, apply (1.6) with $-A$. 

$\square$
Chapter 2
Random Hermitian matrices

2.1 Semicircle law

A Wigner matrix is a Hermitian (or symmetric in the real case) matrix that the upper diagonal and diagonal entries are independent random variables. In this context, we consider the Wigner matrix $M_n = (\zeta_{ij})_{1 \leq i, j \leq n}$ has the upper diagonal entries as iid complex (or real) random variables with zero mean and unit variance, and the diagonal entries as iid real random variables with bounded mean and variance.

A cornerstone of random matrix theory is the semicircle law that dates back to Wigner [71] in the fifties. Denote by $\rho_{sc}$ the semi-circle density function with support on $[-2, 2]$,

$$\rho_{sc}(x) := \begin{cases} \frac{1}{\pi} \sqrt{4 - x^2}, & |x| \leq 2 \\ 0, & |x| > 2. \end{cases} \quad (2.1)$$

Theorem 10 (Semicircular law). Let $M_n$ be a Wigner matrix and let $W_n = \frac{1}{\sqrt{n}} M_n$. Then for any real number $x$,

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

in the sense of probability (and also in the almost sure sense, if the $M_n$ are all minors of the same infinite Wigner Hermitian matrix), where we use $|I|$ to denote the cardinality of a finite set $I$.

The semicircle law can be proved by using both the moment method and the Stieltjes transform method (see [1, 8, 61] for details). We will mention the frameworks of both method in the next subsections.
Figure 2.1: Plotted above is the distribution of the (normalized) eigenvalues of a random symmetric Bernoulli matrix with matrix size $n = 5000$. The red curve is the semicircle law with density function $\rho_{sc}(x)$.

**Remark 11.** Wigner [71] proved this theorem for special ensembles, i.e. for $1 \leq i \leq j \leq n$, $\zeta_{ij}$ are real iid random variables that have symmetric distributions, variance one and $E(|\zeta_{ij}|^{2m}) \leq B_m$ for all $m \geq 1$. Many extensions have been developed later. For example, a more general version was proved by Pastur [48], where $\zeta_{ij}(i \leq j)$ are assumed to be iid real random variables that have mean zero, variance one and satisfy Linderberg condition. Thus it is sufficient to assume the $2 + \epsilon$ ($\epsilon > 0$) moment of $\zeta_{ij}$ are bounded. On the other hand, the semicircle law was first proved in the sense of convergence in probability and later improved to the sense of almost sure convergence by Arnold [2, 3] (see [1, 8] for a detailed discussion).

**Remark 12.** One consequence of Theorem 10 is that we expect most of the eigenvalues of $W_n$ to lie in the interval $(-2 + \epsilon, 2 + \epsilon)$ for $\epsilon > 0$ small; we shall thus refer to this region as the *bulk* of the spectrum. And the region $(-2 - \epsilon, -2 + \epsilon) \cup (2 - \epsilon, 2 + \epsilon)$ is referred as the *edge* of the spectrum.
2.1.1 Moment method

The most direct proof of semicircle law is the moment method given in Wigner’s original proof. It is also called the trace method as it invokes the trace formula: for a positive integer $k$, the $k$-th moment of the ESD $F_{W_n}(x)$ is given by

$$m_k = \int x^k F_{W_n}(dx) = \frac{1}{n} \text{trace}(W_n^k).$$

The starting point of moment method is the moment convergence theorem.

**Theorem 13** (Moment convergence theorem). Let $X$ is a random variable that all the moments exist and assume the probability distribution of $X$ is completely determined by its moments. If

$$\lim_{n \to \infty} E(X_n^k) = E(X^k),$$

then the sequence $\{X_n\}$ converges to $X$ in distribution.

Specially, if the distribution of $X$ is supported on a bounded interval, then the convergence of moments is equivalent to the convergence in distribution.

For the semi-circle distribution, the moments are given by

**Lemma 14.** For odd moments $k = 2m + 1$,

$$m_{2m+1,\text{sc}} = \int_{-2}^{2} x^{2k+1} \rho_{\text{sc}}(x) dx = 0.$$

For even moments $k = 2m$,

$$m_{2m,\text{sc}} = \int_{-2}^{2} x^{k} \rho_{\text{sc}}(x) dx = \frac{1}{m+1} \binom{2m}{m}.$$

**Proof.** For $k = 2m + 1$, by symmetry,

$$\int_{-2}^{2} x^{k} \rho_{\text{sc}}(x) dx = 0.$$

For $k = 2m$, recall that Beta function

$$B(x, y) = 2 \int_{0}^{\pi/2} \sin^{2x-1} \theta \cos^{2y-1} \theta d\theta = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$
Thus
\[
m_{2m,sc} = \int_{-2}^{2} x^k \rho_{sc}(x) dx = \frac{1}{\pi} \int_{0}^{2} x^k \sqrt{4 - x^2} dx
\]
\[
= \frac{2^{k+2}}{\pi} \int_{0}^{\pi/2} \sin^k \theta \cos^2 \theta d\theta
\]
\[
= \frac{2^{k+2}}{\pi} \frac{1}{2} B\left(\frac{k+1}{2}, \frac{3}{2}\right) = \frac{2^{k+1} \Gamma\left(\frac{k+1}{2}\right) \Gamma\left(\frac{3}{2}\right)}{\pi \Gamma\left(\frac{k+4}{2}\right)}
\]
\[
= \frac{4^{m+1} (2m)! (\sqrt{\pi})^2}{\pi 4^m m!} \frac{1}{2 (m+1)!} = \frac{1}{m+1} \binom{2m}{m}.
\]
\[
\square
\]

Notice that from the trace formula,
\[
E(m_k) = \frac{1}{n} E(\text{trace}(W_n^k)) = \frac{1}{n} \sum_{1 \leq i_1, \ldots, i_k \leq n} E\xi_{i_1} \xi_{i_2} \xi_{i_3} \cdots \xi_{i_k} i_1.
\]
The problem of showing the convergence of moments is reduced to a combinatorial counting problem. And the semicircle law can be proved by showing that

**Lemma 15.** For \( k = 2m + 1 \),
\[
\frac{1}{n} E(\text{trace}(W_n^k)) = O\left(\frac{1}{\sqrt{n}}\right);
\]
For \( k = 2m \),
\[
\frac{1}{n} E(\text{trace}(W_n^k)) = \frac{1}{m+1} \binom{2m}{m} + O\left(\frac{1}{n}\right).
\]
And for each fixed \( k \),
\[
\text{Var}\left(\frac{1}{n} \text{trace}(W_n^k)\right) = O\left(\frac{1}{n^2}\right).
\]

We are going to illustrate the calculation of Lemma 15 in section 4.5 for discrete ensembles, similar to Wigner’s original proof. It is remarkable that the proof can be applied, with essentially no modifications, for a more general class of matrices.

### 2.1.2 Stieltjes transform method

The Stieltjes transform \( s_n(z) \) of a Hermitian matrix \( W_n \) is defined for any complex number \( z \) not in the support of \( F^{W_n}(x) \),
\[
s_n(z) = \int_{\mathbb{R}} \frac{1}{x - z} dF^{W_n}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i(W_n) - z}.
\]
The Stieltjes transform can be thought of as the generating function of the moments from the observation: for $z$ large enough,

$$s_n(z) = \frac{1}{n} \text{trace}(W_n - z)^{-1} = -\frac{1}{n} \sum_{k=0}^{n} \frac{\text{trace}(W_n^k)}{z^{k+1}} = -\frac{1}{n} \sum_{k=0}^{n} m_k z^{k+1}.$$ 

Since $(W_n - z)^{-1}$ is called the resolvent of matrix $W_n$, this method is also known as the resolvent method.

By a contour integral, the Stieltjes transform $s(z)$ of the semi-circle distribution is given by

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

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

The semicircle law follows from the criterion of convergence:

**Proposition 16** (Section 2.4, [61]). Let $\mu_n$ be a sequence of probability measure defined on the real line and $\mu$ be a deterministic probability measure. Then $\mu_n$ converges to $\mu$ in probability if and only if $s_{\mu_n}(z)$ converges to $s_\mu(z)$ in probability for every $z$ in the upper half plane.

A more careful analysis of the Stieltjes transform $s_n(z)$ gives more accurate and powerful control on the ESD of $W_n$. We can going to use the Stieltjes transform method frequently in this paper to prove the local version of semicircle law, which subsumes the semicircle law as a special case.

### 2.2 Local semicircle law and the new result

From the semicircle law, we can expect the number of eigenvalues of $W_n = \frac{1}{\sqrt{n}} M_n$ on any fixed interval $I \subset (-2, 2)$ to be of order $n|I|$. It is natural to ask how many eigenvalues of $W_n$ lie on the interval $I$ if the length $|I|$ shrinks with $n$? The eigenvalue density on the smaller scale still follows the semicircle distribution and this is usually called the local semicircle law (LSCL). This problem lies in the heart of proving universality of the local eigenvalue statistics, see [30, 29, 32, 28] and [66, 64].
The leading idea is that we expect that the semi-circle law holds 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 \( W_n \) on the interval \( I \). Of course, the reader can easily see that \( I \) cannot be arbitrarily short (since \( N_I \) is an integer). 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 some similarity between this definition and the definition of threshold functions for random graphs.)

We would like to raise the following problem.

**Problem 17.** Determine the threshold scale (if exists).

A recent result [10] shows that the maximum gap between two consecutive (bulk) eigenvalues of GUE is of order \( \Theta(\sqrt{\log n}/n) \). Thus, if we partition the bulk into intervals of length \( \alpha \sqrt{\log n}/n \) for some small \( \alpha \), one of these intervals contains at most one eigenvalue with high probability. Thus, giving the universality phenomenon, one has reasons to expect that the LSCL do not hold below the \( \sqrt{\log n}/n \) scale, at least for a large class of random matrices.

**Question 18.** Under which condition (for the atom variables of \( M_n \)) the local semi-circle law holds for \( M_n \) at scale \( \log n/n \)?

There have been a number of partial results concerning this question. In [5], Bai et al. proved that the rate of convergence to the SCL is \( O(n^{-1/2}) \) (under a sixth moment assumption). Recently, the rate of convergence is improved to be \( O(n^{-1} \log^b n) \) for some constant \( b > 3 \) by Götze and Tikhomirov [38], assuming the entries of \( M_n \) have a uniform sub-exponential decay. In [30], Erdős, Schlein and Yau proved the LSCL for
scale $n^{-2/3}$ (under some technical assumption on the entries). At two later papers, they strengthened this result significantly. In particular, in [31], they proved scale $\log^2 n/n$ for random matrices with subgaussian entries (this is a consequence of [31, Theorem 3.1]). In [66], Tao and Vu showed that if the entries are bounded by $K$ (which may depend on $n$), then the LSCL holds with scale $K^2 \log^{20} n/n$. The constant 20 was reduced to 2 in a recent paper [67] by Tran, Vu and Wang.

The first main result of this paper is the following.

**Theorem 19.** 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 random matrix with entries bounded by $K$ where $K$ may depend on $n$. 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 $C_2 K^2 \log n/n$.

This provides an affirmative answer for Question 18 in the case when $K = O(1)$ (the matrix has bounded entries).

**Theorem 20.** Let $M_n$ be a random matrix with bounded entries. Then the LSCL holds for $M_n$ at scale $\log n/n$.

By Theorem 19, 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 $\delta$ and $\delta'$. 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 $\delta$ and $\delta'$?

### 2.2.1 Proof of Theorem 19

Let $s_n(z)$ be the Stieltjes transform of $W_n = \frac{1}{\sqrt{n}} M_n$ and $s(z)$ be that of the semicircle distribution. It is well known that if $s_n(z)$ is close to $s(z)$, then the spectral distribution
of $M_n$ is close to the semi-circle distribution (see for instance [8, Chapter 11], [30]). In order to show that $s_n(z)$ is close to $s(z)$, the key observation that the equation

\[ s(z) = -\frac{1}{z + s(z)} \quad (2.2) \]

which defines the Stieltjes transform is stable. This observation was used by Bai et. al. to prove the $n^{-1/2}$ rate of convergence and also served as the starting point of Erdős et. al. approach [30].

We are going to follow this approach whose first step is the following lemma. The proof is a minor modification of the proof of Lemma 64 in [66]. See also the proof of Corollary 4.2 from [30].

**Lemma 21.** 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

\[ |s_n(z) - s(z)| \leq \delta \]

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

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

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

We are going to show (by taking $L = 4, \varepsilon = 1$)

\[ |s_n(z) - s(z)| \leq \delta \quad (2.3) \]

with probability at least $1 - n^{-C}$ ($C$ sufficiently large depending on $C_1$, say $C = C_1 + 10^4$ would suffice) for all $z$ in the region $\{z \in \mathbb{C} : |\text{Re}(z)| \leq 4, |\text{Im}(z)| \geq \eta\}$, where

\[ \eta = \frac{K^2 C^2 \log n}{n \delta^6}. \]

In fact, it suffices to prove (2.3) for any fixed $z$ in the related region. Indeed, notice that $s_n(z)$ is Lipschitz continuous with the Lipschitz constant $O(n^2)$ in the region of interest and equation (2.3) follows by a standard $\varepsilon$-net argument. See also the proof of Theorem 1.1 in [29].
By Schur’s complement, \( s_n(z) \) can be written as
\[
s_n(z) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\sqrt{n} - z - Y_k}
\]
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
\[
E(Y_k|W_{n,k}) = \frac{1}{n} \text{Trace}(W_{n,k} - zI)^{-1} = (1 - \frac{1}{n}) s_{n,k}(z)
\]
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
\[
E(Y_k|W_{n,k}) = s_n(z) + o(\delta^2).
\]

On the other hand, we have the following concentration of measure result.

**Proposition 22.** For \( 1 \leq k \leq n \), \(|Y_k - E(Y_k|W_{n,k})| \leq \delta^2/\sqrt{C} \) holds with probability at least \( 1 - 20n^{-C} \) uniformly for all \( z \) with \( |\text{Re}(z)| \leq 4 \) and \( \text{Im}(z) \geq \eta \).

The proof of this Proposition in [30, 29, 31] relies on Hanson-Wright inequality. In [66], Tao and Vu introduced a new argument based on the so-called projection lemma, which is a consequence of Talagrand inequality.

We will try to follow this argument here. However, the projection lemma is not sufficiently strong for our purpose. The key new ingredient is a generalization called weighted projection lemma. With this lemma, we are able to obtain better estimate on \( Y_k \) (which is a sum of many terms) by breaking its terms into the real and imaginary part (the earlier argument in [66] only considered absolute values of the terms). The details now follow.
Lemma 23 (Weighted projection lemma). Let $X = (\xi_1, \ldots, \xi_n) \in \mathbb{C}^n$ be a random vector whose entries are independent with mean 0 and variance 1. Assume for each $i$, $|\xi_i| \leq K$ almost surely for some $K$, where $K \geq \sup_i |\xi_i|^4 + 1$. Let $H$ be a subspace of dimension $d$ with an orthonormal basis $\{u_1, \ldots, u_d\}$. Assume $c_1, \ldots, c_d$ are constants that $0 < c_j \leq 1$ for every $j$. Then

$$P \left( \left| \sqrt{\sum_{j=1}^{d} c_j u_j^* X} - \sqrt{\sum_{j=1}^{d} |c_j|} \right| \geq t \right) \leq 10 \exp(-t^2/20K^2).$$

In particular,

$$\left| \sum_{j=1}^{d} c_j (|u_j^* X|^2 - 1) \right| \leq 2t \sqrt{\sum_{j=1}^{d} c_j + t^2} \quad (2.5)$$

with probability at least $1 - 10 \exp(-t^2/20K^2)$.

The proof will be deferred to section 2.2.2.

First, we record a lemma that provides a crude upper bound on the number of eigenvalues in short intervals. The proof is a minor modification of existing arguments as Theorem 5.1 in [31] or Proposition 66 in [66].

Lemma 24. For any constant $C_1 > 0$, there exists a constant $C_2 > 0$ ($C_2$ depending on $C_1$, say $C_2 > 10K(C_1 + 10)$ suffices) such that for any interval $I \subset (-4,4)$ with $|I| \geq \frac{C_2 K^2 \log n}{n}$,

$$N_I \ll n |I|$$

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

Proof. By union bounds, it suffices to show for $|I| = \frac{C_2 K^2 \log n}{n}$. Suppose the interval $I = (x, x + \eta) \subset (-4,4)$ with $\eta = |I|$. Let $z = x + \sqrt{-1}\eta$.

$$N_I = \sum_{i=1}^{n} \mathbf{1}_{(\lambda_i(W_n) \in I)} \leq \sum_{\lambda_i(W_n) \in I} \frac{\eta^2}{(\lambda_i(W_n) - x)^2 + \eta^2} \leq 2 \sum_{i=1}^{n} \frac{\eta^2}{(\lambda_i(W_n) - x)^2 + \eta^2} = 2n\eta \Im \frac{1}{\lambda_i(W_n) - x - \sqrt{-1}\eta} = 2n\eta \Im s_n(z)$$

Recall the expression of $s_n(z)$ in (2.4),

$$s_n(z) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{z - a_k^* (W_n - zI)^{-1} a_k}$$
where $W_{n,k}$ 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. Thus $a_k = \frac{1}{\sqrt{n}}X_k$ where the entries of $X_k$ are independent random variable with mean 0 and variance 1. Applying the inequality $|\text{Im} \frac{1}{z}| \leq 1/|\text{Im} z|$, we have

$$N_I \leq 2\eta \sum_{k=1}^{n} \frac{1}{\eta + \text{Im} a_k^*(W_{n,k} - zI)^{-1}a_k}.$$ 

On the other hand,

$$a_k^*(W_{n,k} - zI)^{-1}a_k = \sum_{j=1}^{n-1} \frac{|a_k^*u_j(W_{n,k})|^2}{\lambda_j(W_{n,k}) - \sqrt{-1}\eta},$$

and

$$\text{Im} a_k^*(W_{n,k} - zI)^{-1}a_k = \frac{n}{\eta} \sum_{j=1}^{n-1} \frac{|X_k^*u_j(W_{n,k})|^2}{n^2 + (\lambda_j(W_{n,k}) - x)^2} \geq \frac{1}{2n\eta} \sum_{\lambda_j(W_{n,k}) \in I} |X_k^*u_j(W_{n,k})|^2.$$ 

Thus

$$N_I \leq 4n^2\eta^2 \sum_{k=1}^{n} \frac{1}{\eta} \sum_{\lambda_j(W_{n,k}) \in I} |X_k^*u_j(W_{n,k})|^2.$$ 

Now we prove by contradiction. If $N_I \geq Cn\eta$ for some constant $C > 100$, then there exists $k \in \{1, 2, \ldots, n\}$ such that

$$\sum_{\lambda_j(W_{n,k}) \in I} |X_k^*u_j(W_{n,k})|^2 \geq Cn\eta,$$

thus

$$\sum_{\lambda_j(W_{n,k}) \in I} |X_k^*u_j(W_{n,k})|^2 \leq \frac{4n\eta}{C}.$$ 

By Cauchy interlacing law, $|\{\lambda_j(W_{n,k}) \in I\}| \geq N_I - 2 \geq N_I/2$. By Lemma 23, one concludes that

$$\sum_{\lambda_j(W_{n,k}) \in I} |X_k^*u_j(W_{n,k})|^2 \geq \frac{N_I}{4} \geq \frac{Cn\eta}{4}$$

with probability at least $1 - n^{-(C_1 + 10)}$, assuming $C_2 \geq 10K(C_1 + 10)$. Thus $4n\eta/C \geq Cn\eta/4$ contradicts $C > 100$. This completes the proof. \qed
Now we prove Proposition 22. Notice that
\[ Y_k = a_k^*(W_{n,k} - zI)^{-1}a_k = \sum_{j=1}^{n-1} \frac{|u_j(W_{n,k})^*a_k|^2}{\lambda_j(W_{n,k}) - z}. \]

We evaluate
\[
|Y_k - \mathbf{E}(Y_k|W_{n,k})| = |Y_k - (1 - \frac{1}{n}s_{n,k}(z))| = \sum_{j=1}^{n-1} \frac{|u_j(W_{n,k})^*a_k|^2 - \frac{1}{n}}{\lambda_j(W_{n,k}) - z}
\]
\[= \frac{1}{n} \left| \sum_{j=1}^{n-1} \frac{|u_j(W_{n,k})^*X_k|^2 - 1}{\lambda_j(W_{n,k}) - z} \right| := \frac{1}{n} \left| \sum_{j=1}^{n-1} \frac{R_j}{\lambda_j(W_{n,k}) - x - \sqrt{-1}\eta} \right|. \]

Without loss of generality, we just consider the case \(\lambda_j(W_{n,k}) - x \geq 0\). First, for the set \(J\) of eigenvalues \(\lambda_j(W_{n,k})\) such that \(0 \leq \lambda_j(W_{n,k}) - x \leq \eta\), from Lemma 24 one has 
\[ |J| \ll n\eta \text{ and in Lemma 23, by taking } t = 4K\sqrt{C\log n}, \]

\[ \frac{1}{n} \left| \sum_{j \in J} \frac{R_j}{\lambda_j(W_{n,k}) - x - \sqrt{-1}\eta} \right| \]
\[ \leq \frac{1}{n} \left| \sum_{j \in J} \frac{\lambda_j(W_{n,k}) - x}{(\lambda_j(W_{n,k}) - x)^2 + \eta^2} R_j \right| + \frac{1}{n} \left| \sum_{j \in J} \frac{\eta}{(\lambda_j(W_{n,k}) - x)^2 + \eta^2} R_j \right| \]
\[ \leq \frac{1}{n\eta} \left| \sum_{j \in J} \frac{(\lambda_j(W_{n,k}) - x)\eta}{(\lambda_j(W_{n,k}) - x)^2 + \eta^2} R_j \right| + \frac{1}{n\eta} \left| \sum_{j \in J} \frac{\eta^2}{(\lambda_j(W_{n,k}) - x)^2 + \eta^2} R_j \right| \]
\[ \leq \frac{10}{n\eta} (K\sqrt{C\log n}\sqrt{|J|} + K^2C\log n) \]
\[ \leq \frac{20\delta^3}{\sqrt{C}} \]
with probability at least \(1 - 10n^{-c}\).

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 one has 24, \(|J_l| \ll n\alpha a\). Again by Lemma 23 (take \(t = K\sqrt{C(l+1)\log n}\),

\[
\frac{1}{n} \left| \sum_{j \in J_l} \lambda_j(W_{n,k}) - x - \sqrt{-1\eta} \right|
\leq \frac{1}{n} \left| \sum_{j \in J_l} \lambda_j - x \right| + \frac{1}{n} \sum_{j \in J_l} (\lambda_j - x)^2 + \eta^2 R_j
\leq \frac{1 + \alpha}{na} \left| \sum_{j \in J_l} a(\lambda_j - x) \right| (1 + \alpha) + \eta^2 R_j
\leq \frac{(1 + \alpha) + \eta^2}{na^2} (K \sqrt{C(l + 1)} \sqrt{\log n} \sqrt{naa} + K^2 C(l + 1) \log n)
\leq 20\delta^3 \frac{l + 1}{\sqrt{C} (1 + \alpha)^{1/2}},
\]
with probability at least \(1 - 10n^{-C(l+1)}\).

Summing over \(l\), we have
\[
\frac{1}{n} \left| \sum_{l = 1}^\infty \sum_{j \in J_l} \lambda_j(W_{n,k}) - x - \sqrt{-1\eta} \right| \leq \frac{40\delta^3}{\sqrt{C}},
\]
with probability at least \(1 - 10n^{-C}\). This completes the proof of Proposition 22.

Let \(|Y_k - E(Y_k|W_{n,k})| = \delta_C^2 \leq 200\delta^3/\sqrt{C}\). Inserting the bounds into (2.4), one has
\[
s_n(z) + \frac{1}{n} \sum_{k=1}^n \frac{1}{s_n(z) + z + \delta_C^2} = 0
\]
with probability at least \(1 - 10n^{-C}\). The term \(|\zeta_{kk}/\sqrt{n}| = o(\delta^2)\) as \(|\zeta_{kk}| \leq K\) by assumption. For the error term \(\delta_C\), we can consider that either \(|s_n(z) + z| = o(1)\) or \(|s_n(z) + z| \geq C_1 > 0\) for some constant \(C_1\). In the former case, we have
\[
s_n(z) = -z + o(1).
\]
In the later case, by choosing \(C\) large enough, we can operate a Taylor expansion to get
\[
s_n(z) + \frac{1}{z + s_n(z)} \left( 1 + O\left( \frac{\delta_C^2}{z + s_n(z)} \right) \right) = 0.
\]
And thus
\[
s_n(z) + \frac{1}{z + s_n(z)} = O(\delta_C^2),
\]
with probability at least \(1 - 10n^{-C}\). Multiplying \(z + s_n(z)\) on both sides and completing the perfect square, we have
\[
s_n(z) = -\frac{z}{2} \pm \sqrt{1 - \frac{z^2}{4} + O(\delta_C^2)}. \quad (2.7)
\]
Now we consider the cases \( O(\delta_C^2)/\sqrt{4-z^2} = O(\delta_C) \) and \( 4-z^2 = o(1) \). In the first case, after a Taylor expansion, we can conclude

\[
\sin(z) = -\frac{z}{2} \pm \sqrt{1 - \frac{z^2}{4}} + O(\delta_C).
\]

In the second case, from (2.7), one has

\[
\sin(z) = -\frac{z}{2} + O(\delta_C) = s(z) + O(\delta_C).
\]

Recall that \( s(z) \) is the unique solution to the quadratic equation \( s(z) + \frac{1}{s(z)+z} = 0 \) with positive imaginary part and has the explicit form

\[
s(z) = -z + \frac{\sqrt{z^2 - 4}}{2},
\]

where \( \sqrt{z^2 - 4} \) is the branch of square root with a branch cut in \([-2, 2]\) and asymptotically equals \( z \) at infinity. In conclusion, we have in the region

\[
|z| \leq 10, \ |\text{Re}(z)| \leq 4, \ \text{Im}(z) \geq \eta,
\]

either

\[
s_n(z) = -z + o(1), \quad (2.8)
\]

or

\[
s_n(z) = s(z) - \sqrt{z^2 - 4} + O(\delta_C), \quad (2.9)
\]

or

\[
s_n(z) = s(z) + O(\delta_C), \quad (2.10)
\]

with probability at least \( 1 - 10n^{-(C+100)} \). By choosing \( C \) sufficiently large, it is not hard to say that (2.8) and (2.9) or (2.8) and (2.10) do not hold at the same time.

Since otherwise, one has \( s(z) = O(\delta_C) \) or \( s(z) + z = O(\delta_C) \), which contradicts the fact that \( |s(z) + z| \) and \( |s(z)| \) have positive lower bounds. And (2.9) and (2.10) are disconnected from each other except \( |z^2 - 4| = O(\delta^2) \). The possibility (2.8) or (2.9) holds only when \( \text{Im}(z) = o(1) \) since \( s_n(z) \) and \( z \) both have positive imaginary parts.

By a continuity argument, we can show that (2.10) must hold throughout the region except that \( |z^2 - 4| = O(\delta^2) \). In that case, (2.9) and (2.10) are actually equivalent. Thus we always have (2.10) holds with probability at least \( 1 - 10n^{-(C+100)} \).

Applying Lemma 21, we have
Theorem 25. 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 $C_2 K^2 \log n/\delta^8$, 

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

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

In particular, Theorem 19 follows.

2.2.2 Proof of Lemma 23

Denote $f(X) = \sqrt{\sum_{j=1}^d c_j |u_j^* X|^2}$, which is a function defined on $\mathbb{C}^n$.

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

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

Second, $f(X)$ is $1$-Lipschitz. Noticed that

$$f(X) \leq \|\pi_H(X)\| = \sqrt{\sum_{j=1}^d |u_j^* X|^2} \leq \|X\|.$$

Since $f(X)$ is convex, one has

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

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

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

Now we can apply the following Talagrand’s inequality (see Theorem 69, [66]):

Thus

$$P(|f(X) - Mf(X)| \geq t) \leq 4 \exp(-t^2/16K^2). \quad (2.11)$$

To conclude the proof, it suffices to show

$$|Mf(X) - \sqrt{\sum_{j=1}^d c_j}| \leq 2K.$$
It is enough to prove that $\mathbf{P}(f(X) \geq 2K + \sqrt{\sum_{j=1}^{d} c_j}) \leq 1/4$ and $\mathbf{P}(f(X) \leq -2K + \sqrt{\sum_{j=1}^{d} c_j}) \leq 1/4$. And this can be done by applying the Chebyshev inequality.

Denote $u_j = (u_j^1, \ldots, u_j^n) \in \mathbb{C}^n$. Then

$$c_j |u_j^* X|^2 = c_j \sum_{i=1}^{n} |u_j^i|^2 |\xi_i|^2 + c_j \sum_{i \neq k} u_j^i u_j^k \xi_i \xi_k.$$ 

$$\mathbf{P} \left( f(X) \geq 2K + \sqrt{\sum_{j=1}^{d} c_j} \right) \leq \mathbf{P} \left( \sum_{j=1}^{d} c_j |u_j^* X|^2 - \sum_{j=1}^{d} c_j \geq 4K \sqrt{\sum_{j=1}^{d} c_j} \right)$$

$$= \mathbf{P} \left( \sum_{j=1}^{d} \left( \sum_{i=1}^{n} |u_j^i|^2 |\xi_i|^2 - 1 \right) + \sum_{j=1}^{d} c_j \sum_{i \neq k} u_j^i u_j^k \xi_i \xi_k \geq 4K \sqrt{\sum_{j=1}^{d} c_j} \right)$$

$$\leq \frac{\mathbf{E} \left( \sum_{j=1}^{d} \left( \sum_{i=1}^{n} |u_j^i|^2 |\xi_i|^2 - 1 \right) + \sum_{j=1}^{d} c_j \sum_{i \neq k} u_j^i u_j^k \xi_i \xi_k \right)^2}{16K^2 (\sum_{j=1}^{d} c_j)}$$

Now we evaluate

$$S_1 := \mathbf{E} \left( \sum_{j=1}^{d} c_j \left( \sum_{i=1}^{n} |u_j^i|^2 |\xi_i|^2 - 1 \right) \right)^2$$

and

$$S_2 := \mathbf{E} \left( \sum_{j=1}^{d} c_j \sum_{i \neq k} u_j^i u_j^k \xi_i \xi_k \right)^2.$$ 

$$S_1 = \mathbf{E} \left( \sum_{j=1}^{d} c_j \left( \sum_{i=1}^{n} |u_j^i|^2 |\xi_i|^2 - 1 \right) \right)^2$$

$$= \mathbf{E} \left( \sum_{j=1}^{d} c_j \left( \sum_{i=1}^{n} |u_j^i|^2 |\xi_i|^2 - 1 \right) \right) \left( \sum_{k=1}^{d} c_k \left( \sum_{s=1}^{n} |u_k^s|^2 |\xi_s|^2 - 1 \right) \right)$$

$$= \mathbf{E} \sum_{j,k=1}^{d} c_j c_k \left( \sum_{i=1}^{n} |u_j^i|^2 |\xi_i|^2 - 1 \right) \left( \sum_{s=1}^{n} |u_k^s|^2 |\xi_s|^2 - 1 \right)$$

$$= \mathbf{E} \sum_{j,k=1}^{d} c_j c_k \sum_{i=1}^{n} |u_j^i|^2 |u_k^i|^2 |\xi_i|^4 + \mathbf{E} \sum_{j,k=1}^{d} c_j c_k \sum_{i \neq s} \left| u_j^i \right|^2 \left| u_k^s \right|^2 \left| \xi_i \right|^2 \left| \xi_s \right|^2 - \sum_{j,k=1}^{d} c_j c_k.$$
Therefore,

\[ S_1 \leq (K + 1) \sum_{i=1}^{n} \left( \sum_{j=1}^{d} c_j |u_{ij}|^2 \right)^2 + \sum_{i \neq s} \left( \sum_{j=1}^{d} c_j |u_{ij}|^2 \right) \left( \sum_{k=1}^{d} c_k |u_{ik}|^2 \right) - \sum_{i,s=1}^{n} \left( \sum_{j=1}^{d} c_j |u_{ij}|^2 \right) \left( \sum_{k=1}^{d} c_k |u_{ik}|^2 \right) \]

\[ = (K + 1) \sum_{i=1}^{n} \left( \sum_{j=1}^{d} c_j |u_{ij}|^2 \right)^2 - \sum_{i=1}^{n} \left( \sum_{j=1}^{d} c_j |u_{ij}|^2 \right)^2 = K \sum_{i=1}^{n} \left( \sum_{j=1}^{d} c_j |u_{ij}|^2 \right)^2 \]

\[ \leq K \sum_{i=1}^{n} \left( \sum_{j=1}^{d} c_j |u_{ij}|^2 \right) = K \left( \sum_{j=1}^{d} c_j \right) \]

and

\[ S_2 = E \left( \sum_{j,l=1}^{d} c_j c_l \sum_{i \neq k} u_{ij}^l u_{ik}^l \bar{\xi}_i \bar{\xi}_k \right) \left( \sum_{l=1}^{d} c_l \sum_{s \neq t} u_{il}^s \bar{\xi}_s \bar{\xi}_t \right) \]

\[ = \sum_{j,l=1}^{d} c_j c_l \sum_{i \neq k} u_{ij}^l u_{ik}^l \]

\[ \leq \sum_{j=1}^{d} c_j \sum_{i,k=1}^{n} |u_{ij}|^2 |u_{ik}|^2 = \sum_{j=1}^{d} c_j, \]

here we used the fact \( \sum_{j \neq l} c_j c_l \sum_{i \neq k} (u_{ij}^l u_{ik}^l)(u_{ij}^k u_{ik}^k) \leq 0 \), since for \( j \neq l \),

\[ 0 = \left( \sum_{i=1}^{n} u_{ij}^l u_{ik}^l \right) \left( \sum_{k=1}^{n} u_{ij}^k u_{ik}^k \right) = \sum_{i=1}^{n} |u_{ij}^l|^2 |u_{ij}^k|^2 + \sum_{i \neq k} (u_{ij}^l u_{ik}^l)(u_{ij}^k u_{ik}^k). \]

Therefore,

\[ P \left( f(X) \geq 2K + \sqrt{\sum_{j=1}^{d} c_j} \right) \leq (K + 1)/8K^2 \leq 1/4. \]

With the same argument, one can show

\[ P \left( f(X) \leq -2K + \sqrt{\sum_{j=1}^{d} c_j} \right) \leq 1/4. \]

This completes the proof.
Figure 2.2: Plotted above are the empirical cumulative distribution functions of the distribution of $\sqrt{n} \times \|v\|_{\infty}$ for $n = 1000$, evaluated from 500 samples. In the blue curve, $v$ is a unit eigenvector for GOE. And $v$ is a unit eigenvector for symmetric random sign matrix in the red curve. The green curve is generated for $v$ to have a uniform distribution on the unit sphere $S_n$.

2.3 Optimal upper bound for the infinity norm of eigenvectors

It has been long conjectured that $u_i$ must look like a uniformly chosen vector from the unit sphere. Indeed, for one special random matrix model, the GOE, one can identify a random eigenvector with a random vector from the sphere, using the rotational invariance property (see [47] for more details). For other models of random matrices, this invariance is lost and only very recently we have some theoretical support for the conjecture [65, 44]. In particular, it is proved in [65] that under certain assumption, the inner product $u \cdot a$ satisfies s central limit theorem, for any fixed vector $a \in S_n$. For numerical simulation in Figure 2.2, we plot the cumulative distribution functions of the (normalized) infinity norm of eigenvector $v$ for GOE and random symmetric Bernoulli matrix separately, and compare them with the vector chosen uniformly from the unit sphere.

One important property of a random unit vector is that it has small infinity norm. It is well-known and easy to prove that if $w$ is chosen randomly (uniformly) from $S_n$ (the unit sphere in $\mathbb{R}^n$), then with high probability $\|w\|_{\infty} = O(\sqrt{\log n}/n)$ and this
bound is optimal up to the hidden constant in $O$. We are going to show

**Theorem 26 (Delocalization 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 $1 \leq i \leq n$ with $\lambda_i(W_n) \in [-2 + \epsilon, 2 - \epsilon]$, let $u_i(W_n)$ denote the corresponding unit eigenvector, then
  \[
  \|u_i(W_n)\|_\infty \leq \frac{C_2 K \log^{1/2} n}{\sqrt{n}}
  \]
  with probability at least $1 - n^{-C_1}$.

- **(Edge case)** For any $\epsilon > 0$ and any $1 \leq i \leq n$ 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
  \[
  \|u_i(W_n)\|_\infty \leq \frac{C_2 K^2 \log n}{\sqrt{n}}
  \]
  with probability at least $1 - n^{-C_1}$.

### 2.3.1 Proof of the bulk case

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

**Lemma 27 (Eq (4.3), [29] or Lemma 41, [66]).** 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 a unit 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 $\lambda_j(B_{n-1})$. 

Proof. From the equation
\[
\begin{pmatrix}
a & X^* \\
X & B_{n-1}
\end{pmatrix}
\begin{pmatrix}
x \\
v
\end{pmatrix}
= \lambda_i(B_n)
\begin{pmatrix}
x \\
v
\end{pmatrix},
\]
one has
\[xX + B_{n-1}v = \lambda_i(B_n)v.\]
Since none of eigenvalues of \(B_{n-1}\) are equal to \(\lambda_i(B_n)\), the matrix \(\lambda_i(B_n)I - B_{n-1}\) is invertible. Thus \(v = x(\lambda_i(B_n)I - B_{n-1})^{-1}X\). Inserting the expression of \(v\) into the
\[|x|^2 + \|v\|^2 = 1\]
and decomposing
\[(\lambda_i(B_n)I - B_{n-1})^{-1} = \sum_{j=1}^{n-1} (\lambda_j(B_{n-1}) - \lambda_i(B_n))^{-1} u_j(B_{n-1}),\]
we prove that
\[|x|^2 = \frac{1}{1 + \sum_{j=1}^{n-1} (\lambda_j(B_{n-1}) - \lambda_i(B_n))^2 |u_j(B_{n-1})|^2}.\]

First, for the bulk case, for any \(\lambda_i(W_n) \in (-2 + \varepsilon, 2 - \varepsilon)\), by Theorem 19, one can find an interval \(I \subset (-2 + \varepsilon, 2 - \varepsilon)\), centered at \(\lambda_i(W_n)\) and \(|I| = K^2C \log n/n\), 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 27, we have
\[
|x|^2 \leq \frac{1}{1 + n^{-1} |I|^{-2} \sum_{j \in J} |u_j(W_{n-1})|^2} \leq \frac{1}{1 + 100^{-1} n^{-1} |I|^{-2} |J|} \leq \frac{K^2 C_2^2 \log n}{n}.
\]
for some constant $C_2$ with probability at least $1 - n^{-C_1 - 10}$. The third inequality follows from Lemma 23 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 $1 - n^{-C_1}$.

### 2.3.2 Proof of the edge case

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

**Lemma 28** (Interlacing identity, Lemma 37, [64]). If none of the eigenvalues of $W_{n-1}$ is equal to $\lambda_i(W_n)$, then

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

(2.13)

**Proof.** Let $u_i(W_n)$ be the eigenvector corresponding to the eigenvalue $\lambda_i(W_n)$. Let $u_i^* = (v^*, x)$ where $v \in \mathbb{R}^{n-1}$ and $x \in \mathbb{R}$.

From the equation

$$\begin{pmatrix} W_{n-1} - \lambda_i(W_n)I_{n-1} & Y \\ Y^* & \xi_{nn} - \lambda_i(W_n) \end{pmatrix} \begin{pmatrix} v \\ x \end{pmatrix} = 0$$

one has

$$(W_{n-1} - \lambda_i(W_n)I_{n-1})v + x Y = 0$$

and

$$Y^* v + x (\xi_{nn} / \sqrt{n} - \lambda_i(W_n)) = 0.$$

Since none of the eigenvalues of $W_{n-1}$ is equal to $\lambda_i(W_n)$, one can solve $v = -x (W_{n-1} - \lambda_i(W_n)I_{n-1})^{-1} Y$ from the first identity. Plugging into the second identity, we have

$$\frac{1}{\sqrt{n}} \xi_{nn} - \lambda_i(W_n) = Y^* (W_{n-1} - \lambda_i(W_n)I_{n-1})^{-1} Y.$$

The conclusion follows by composing

$$(W_{n-1} - \lambda_i(W_n)I_{n-1})^{-1} = \sum_{j=1}^{n-1} \frac{u_j(W_{n-1})^* u_j(W_{n-1})}{\lambda_j(W_{n-1}) - \lambda_i(W_n)}.$$
By symmetry, it suffices to consider the case $\lambda_i(W_n) \in [2 - \epsilon, 2 + \epsilon]$ for $\epsilon > 0$ small. By Lemma 27, 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, $|u_j(W_{n-1})^* X| \leq K \sqrt{\log n}$ 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-Schwarz 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 28, we are going to show for $T_+ - T_- = O(\log n)$ (the choice of $T_+, T_-$ will be given later) that
\[
|\sum_{j \geq T_+, \omega j \leq T_-} \frac{|u_j(W_{n-1})^* Y|^2}{\lambda_j(W_{n-1}) - \lambda_i(W_n)}| \leq 2 - \epsilon - \frac{\sqrt{T_+ - T_-}}{C^{1.5} K \sqrt{\log n}} + o(1), \tag{2.14}
\]
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 $|I| = \frac{K^2 C \log n}{n \delta^s}$ 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 19, the number of eigenvalues
\[
|J_k| = N_{I_k} \leq n \alpha_k |I_k| + \delta^k n |I_k|
\]
with probability at least $1 - n^{-C_1 - 100}$, where

$$\alpha_{I_k} = \int_{I_k} \rho_{sc}(x) \, dx / |I_k|.$$ 

By Lemma 23, for the $k$th interval,

$$\frac{1}{n} \sum_{j \in J_k} \frac{|u_j(W_{n-1}^*)X|^2}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|} \leq \frac{1}{n \text{dist}(\lambda_i(W_n), I_k)} \sum_{j \in J_k} |u_j(W_{n-1}^*)X|^2$$

$$\leq \frac{1}{n \text{dist}(\lambda_i(W_n), I_k)} (|J_k| + K \sqrt{|J_k| \sqrt{C \log n} +CK^2 \log n})$$

$$\leq \frac{\alpha_{I_k}|I_k|}{\text{dist}(\lambda_i(W_n), I_k)} + \frac{\delta^k|I_k|}{\text{dist}(\lambda_i(W_n), I_k)} + \frac{CK^2 \log n}{n \text{dist}(\lambda_i(W_n), I_k)}$$

$$+ \frac{K \sqrt{n\alpha_{I_k} + n\delta^k |I_k| \sqrt{C \log n}}}{n \text{dist}(\lambda_i(W_n), I_k)}$$

$$\leq \frac{\alpha_{I_k}|I_k|}{\text{dist}(\lambda_i(W_n), I_k)} + 2\delta^{k-16} + \delta^{8k-8} + \delta^{4k-15},$$

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

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

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

The contribution of such intervals can be computed similarly by

$$\frac{1}{n} \sum_{j \in J_k} \frac{|u_j(W_{n-1})X|^2}{|\lambda_j(W_{n-1}) - \lambda_i(W_n)|} \leq \frac{1}{n \text{dist}(\lambda_i(W_n), I_k)} \sum_{j \in J_k} |u_j(W_{n-1}^*)X|^2$$

$$\leq \frac{1}{n \text{dist}(\lambda_i(W_n), I_k)} (|J_k| + K \sqrt{|J_k| \sqrt{C \log n} +CK^2 \log n})$$

$$\leq \frac{\alpha_{I_k}|I_k|}{\text{dist}(\lambda_i(W_n), I_k)} + \frac{\delta^k|I_k|}{\text{dist}(\lambda_i(W_n), I_k)} + \frac{CK^2 \log n}{n \text{dist}(\lambda_i(W_n), I_k)}$$

$$+ \frac{K \sqrt{n\alpha_{I_k} + n\delta^k |I_k| \sqrt{C \log n}}}{n \text{dist}(\lambda_i(W_n), I_k)}$$

$$\leq \frac{\alpha_{I_k}|I_k|}{\text{dist}(\lambda_i(W_n), I_k)} + \frac{100\delta^{k_0}}{k - k_0},$$

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

Summing over all intervals for $k \geq 18$ (say), we have

$$\left| \sum_{j \geq T_{+} \text{or } j \leq T_{-}} \frac{|u_j(W_{n-1})Y|}{\lambda_j(W_{n-1}) - \lambda_i(W_n)} \right| \leq \left| \sum_{I_k} \frac{\alpha_{I_k}|I_k|}{\text{dist}(\lambda_i(W_n), I_k)} \right| + \delta. \quad (2.15)$$
Using Riemann integration of the principal value integral,

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

where

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

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

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

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

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

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 $1 - n^{-C_1 - 10}$,

$$
|x| \leq \frac{C_2 K^2 \log n}{\sqrt{n}}.
$$
Chapter 3
Random covariance matrices

3.1 Marchenko-Pastur law

The sample covariance matrix plays an important role in fields as diverse as multivariate statistics, wireless communications, signal processing and principal component analysis. In this chapter, we extend the results obtained for random Hermitian matrices discussed in the previous chapter to random covariance matrices, focusing on the changes needed for the proofs.

Let $X$ be a random vector $X = (X_1, \ldots, X_p)^T \in \mathbb{C}^{p \times 1}$ and assume for simplicity that $X$ is centered. Then the true covariance matrix is given by

$$\mathbb{E}(XX^*) = (\text{cov}(X_i, X_j))_{1 \leq i, j \leq p}.$$

Consider $n$ independent samples or realizations $x_1, \ldots, x_n \in \mathbb{C}^p$ and form the $p \times n$ data matrix $M = (x_1, \ldots, x_n)$. Then the (sample) covariance matrix is an $n \times n$ non-negative definite matrix defined as

$$W_{n,p} = \frac{1}{n} M^* M.$$

If $n \to +\infty$ and $p$ is fixed, then the (sample) covariance matrix converges (entrywise) to the true covariance matrix almost surely. Now we focus on the case that $p$ and $n$ tend to infinity as the same time.

Let $M_{n,p} = (\zeta_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$ be a random $p \times n$ matrix, where $p = p(n)$ is an integer such that $p \leq n$ and $\lim_{n \to \infty} p/n = y$ for some $y \in (0, 1]$. The matrix ensemble $M$ is said to obey condition C1 with constant $C_0$ if the random variables $\zeta_{ij}$ are jointly independent, have mean zero and variance one, and obey the moment condition

$$\sup_{i,j} \mathbb{E}|\zeta_{ij}|^{C_0} \leq C$$

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 matrix has at most \( p \) non-zero eigenvalues which are ordered as

\[
0 \leq \lambda_1(W) \leq \lambda_2(W) \leq \ldots \leq \lambda_p(W).
\]

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 \( M u_i = \sigma_i v_i \) 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 [45] and [6]).

**Theorem 29.** (Marchenko-Pastur Law) Assume a \( p \times n \) random matrix \( M \) obeys condition \( C_1 \) 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

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

where

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

**Remark 30.** When \( y = 1 \), the density function \( \rho_{MP,y}(x) \) is supported on the interval \([0,4]\) and

\[
\frac{d\mu}{dx} = \rho_{MP,y}(x) = \frac{1}{2\pi} \sqrt{\frac{4-x}{x}}.
\]

Actually, by a change of variable \( x \to x^2 \), the distribution \( \mu \) is the image of the semicircle law.

In the following, we discuss briefly the frameworks for two approaches: the moment method and Stieltjes transform method, the latter of which is the core of the proof for our new results.
Figure 3.1: Plotted above are the density functions $\rho_{MP,y}(x)$ of Marchenko-Pastur law for $y = 0.4$, $y = 0.6$, $y = 0.8$ and $y = 1$. Notice that for $y = 1$, the density function has a singularity at $x = 0$.

3.1.1 Moment method

Similar as the proof of semicircle law, we use the trace formula: for a positive integer $k$, the $k$-th moment of the ESD $F_W(x)$ is given by

$$m_k = \int x^k F_W(dx) = \frac{1}{p} \text{trace}(W^k) = \frac{1}{n} \text{trace}((M^*M/n)^k).$$

For the Marchenko-Pastur distribution, the moments are given by

**Lemma 31.** For $k \geq 0$,

$$m_{k,MP} = \int_a^b x^k \rho_{MP,y}(x)dx = \sum_{i=0}^{k-1} \frac{1}{i+1} \binom{k}{i} \binom{k-1}{i} y^i.$$
Figure 3.2: Plotted above is the distribution of the eigenvalues of $\frac{1}{n}M^*M$ where $M$ is a $p \times n$ random Bernoulli matrix with $n = 5000$ and $y = p/n = 0.8$. The red curve is the Marchenko-Pastur law with density function $\rho_{MP,y}(x)$.

Proof. We have $a = (1 - \sqrt{y})^2$, $b = (1 + \sqrt{y})^2$, $a + b = 2(1 + y)$ and $b - a = 4\sqrt{y}$.

$$m_{k,MP} = \int_a^b x^k \rho_{MP,y}(x)dx = \int_a^b x^k \frac{\sqrt{(b-x)(x-a)}}{2\pi xy} dx$$

$$= \int_{t=x-(a+b)/2}^{t=x+(a+b)/2} \frac{1}{2\pi y} \int_{(b-a)/2}^{(b-a)/2} (t + a + b) t^{-1} \sqrt{\frac{(b-a)^2}{4} - t^2} dt$$

$$= \frac{2}{2\pi y} \int_0^{2\sqrt{y}} \sqrt{4y - t^2} \sum_{i=0}^{[k-1]/2} \binom{k-1}{2i} t^{2i} (y + 1)^{k-1-2i} dt$$

$$= \frac{s^{[k-1]/2}}{4y} \sum_{i=0}^{[k-1]/2} \frac{2}{\pi} \binom{k-1}{2i} (4y)^i (y + 1)^{k-1-2i} \int_0^1 s^{i-1/2}(1 - s)^{1/2} ds.$$

Recall that Beta function

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$
Thus

\[ m_k = \frac{2}{\pi} \sum_{i=0}^{\left\lfloor \frac{k-1}{2} \right\rfloor} \binom{k-1}{2i} (4y)^i (y+1)^{k-1-2i} B(i+1/2, 3/2) \]

\[ = \frac{2}{\pi} \sum_{i=0}^{\left\lfloor \frac{k-1}{2} \right\rfloor} \binom{k-1}{2i} (4y)^i (y+1)^{k-1-2i} \frac{(2i)! \pi}{4^i i! (2i+1)!} \]

\[ = \sum_{i=0}^{\left\lfloor \frac{k-1}{2} \right\rfloor} \binom{k-1}{2i} (2i)! \frac{(2i)!}{i!(i+1)!} \sum_{j=0}^{k-2i} \binom{k-1-2i}{j} y^j \]

\[ = \frac{k-1}{k} \sum_{r=0}^{\min(r,k-1-r)} \sum_{i=0}^{\left\lfloor \frac{k-1}{2} \right\rfloor} \frac{k}{i!} \frac{k}{(i+1)!} \frac{k}{(k-1-i-r)!} y^r \]

\[ = \sum_{i=0}^{k-1} \frac{1}{i+1} \binom{k}{i} \binom{k-1}{i} y^i. \]

From the trace formula, one has

\[ \mathbb{E}(m_k) = \mathbb{E}\left( \frac{1}{p} \text{trace}(W^k) \right) = \mathbb{E}\left( \frac{1}{n} \text{trace}\left( \left( \frac{M^* M}{n} \right)^k \right) \right) \]

\[ = \frac{1}{pn^k} \sum_{\{i_1, \ldots, i_k\}} \sum_{\{j_1, \ldots, j_k\}} \mathbb{E}(\zeta_{i_1j_1} \bar{\zeta}_{i_2j_1} \zeta_{i_2j_2} \bar{\zeta}_{i_3j_2} \cdots \zeta_{i_kj_k} \bar{\zeta}_{i_1j_k}). \]

The Marčenko-Pastur law follows by showing

**Lemma 32.** For each fixed integer \( k \),

\[ \mathbb{E}\left( \frac{1}{p} \text{trace}(W^k) \right) = \sum_{r=0}^{k-1} \frac{p}{r+1} \binom{k}{r} \binom{k-1}{r} + O(n^{-1}), \quad (3.1) \]

and

\[ \text{Var}\left( \frac{1}{p} \text{trace}(W^k) \right) = O(n^{-2}). \quad (3.2) \]

The detail of the proof, which is a combinatorial counting argument, can be found in Chapter 3 of the book by Bai and Silverstein [8]. It is straightforward to show (deterministically)

\[ \mathbb{E}\left( \frac{1}{p} \text{trace}(W^k) \right) \rightarrow m_k, \]

for each fixed integer \( k \). And (3.2) together with Borel–Cantelli lemma imply that, in fact, \( p^{-1} \text{trace}(W^k) \) is close to its expectation \( \mathbb{E}(p^{-1} \text{trace}(W^k)) \).
3.1.2 Stieltjes transform method

The Stieltjes transform of Marchenko-Pastur law is given by

\[ s_{MP,y}(z) := \int_{\mathbb{R}} \frac{1}{x-z} \rho_{MP,y}(x) \, dx = \int_{a}^{b} \frac{1}{2\pi xy(x-z)} \sqrt{(b-x)(x-a)} \, dx, \]

which is the unique solution to the equation

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

in the upper half plane.

More explicitly,

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

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

By Proposition 16, the Marchenko-Pastur law follows by showing that \( s(z) \to s_{MP,y}(z) \) in probability for every \( z \) in the upper half plane. Similar to the Hermitian case, we are going to prove the local version of Marchenko-Pastur law, which is much stronger.

3.2 Local Marchenko-Pastur law and the new result

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 [32], [51],[62],[69] for the soft edge and [63], [16] 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 quantitative local Marchenko-Pastur law (LMPL).

**Theorem 33.** 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

$$\left| N_I(W_{n,p}) - p \int_I \rho_{MP,y}(x) \, dx \right| \leq 6p|I|.$$ 

for any interval $I \subset (a + \epsilon, b - \epsilon)$ of length at least $C_2 K^2 \log n/n$.

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

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

which is the unique solution to the equation

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

in the upper half plane. We will show that $s(z)$ satisfies a similar equation.

The analogue of Lemma 21 is the following:

**Proposition 34.** (Lemma 29, [62]) Let $1/10 \geq \eta \geq 1/n$, and $L_1, L_2, \epsilon, \delta > 0$. For any constant $C_1 > 0$, there exists a constant $C > 0$ such that if one has the bound

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

with (uniformly) probability at least $1 - n^{-C}$ for all $z$ with $L_1 \leq \text{Re}(z) \leq L_2$ and $\text{Im}(z) \geq \eta$. Then for any interval $I$ in $[L_1 - \epsilon, L_2 + \epsilon]$ with $|I| \geq \max(2\eta, \frac{\eta}{3} \log \frac{1}{\delta})$, one has

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

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

Our objective is to show

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

with probability at least $1 - n^{-C}$ for all $z$ in the region $R_y$, where

$$R_y = \{ z \in \mathbb{C} : |z| \leq 10, a - \epsilon \leq \text{Re}(z) \leq b + \epsilon, \text{Im}(z) \geq \eta \}$$
if $y \neq 1$, and

$$R_y = \{ z \in \mathbb{C} : |z| \leq 10, \epsilon \leq \text{Re}(z) \leq 4 + \epsilon, \text{Im}(z) \geq \eta \}$$

if $y = 1$. We use the parameter

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

In the defined region $R_y$,

$$|s_{MP,y}(z)| = O(1).$$

The next lemma is an analogue of Lemma 24.

**Lemma 35.** For any constant $C_1 > 0$, there exists a constant $C_2 > 0$ ($C_2$ depending on $C_1$, say $C_2 > 10K(C_1 + 10)$ suffices) such that for any interval $I = (x, x + \eta) \subset (a, b)$ with $|I| \geq \frac{C_2K^2\log n}{n}$ and $x > 0$ constant,

$$N_I \ll n|I|$$

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

**Proof.** By union bounds, it suffices to show for $|I| = \frac{C_2K^2\log n}{n}$. Suppose the interval $I = (x, x + \eta) \subset (a, b)$ with $\eta = |I|$. Let $z = x + \sqrt{-1}\eta$.

$$N_I = \sum_{i=1}^{n} 1_{\lambda_i(W_n) \in I} \leq 2 \sum_{\lambda_i(W_n) \in I} \frac{\eta^2}{(\lambda_i(W_n) - x)^2 + \eta^2} = 2n\eta\text{Im}s(z).$$

Recall the expression of $s(z)$ in (2.4), we have

$$s(z) = \frac{1}{p} \sum_{k=1}^{p} \frac{1}{\xi_{kk} - z - a_k^*(W_k - zI)^{-1}a_k}, \quad (3.4)$$

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 $X_i^* \in \mathbb{C}^n$ ($1 \leq i \leq p$) be the rows of $M$. Thus

$$\xi_{kk} = X_k^*X_k/n = ||X_k||^2/n, a_k = \frac{1}{n}M_kX_k, W_k = \frac{1}{n}M_kM_k^*.$$
Applying the inequality $|\text{Im} \frac{1}{z}| \leq 1/|\text{Im} z|$, we have

$$N_I \leq \frac{2n\eta}{p} \sum_{k=1}^{p} \frac{1}{\eta + \text{Im} \alpha_k^*(W_k - z I)^{-1}a_k}.$$ 

On the other hand,

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

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

$$\sigma_j(M_k)^2 = n\lambda_j(W_k).$$

And by hypothesis, for each $k$,

$$\text{Im} \alpha_k^*(W_k - z I)^{-1}a_k = \frac{\eta}{n^2} \sum_{j=1}^{n-1} \frac{\sigma_j(M_k)^2|X_k^*u_j(M_k)|^2}{(\lambda_j(W_k) - z)^2 + \eta^2} \geq \frac{\eta}{n^2} \sum_{j=1}^{n-1} \frac{\sigma_j(M_k)^2|X_k^*u_j(M_k)|^2}{2\eta^2} \geq \frac{1}{2n\eta} \sum_{\lambda_j(W_k) \in I} \frac{\sigma_j(M_k)^2}{n} |X_k^*u_j(M_k)|^2 \gg \frac{1}{n\eta} \sum_{\lambda_j(W_k) \in I} |X_k^*u_j(M_k)|^2.$$

The last inequality is obtained by the fact $\sigma_j(M_k)/\sqrt{n} = \Theta(1)$ in the region considered.

Thus

$$N_I \ll \frac{n^2\eta^2}{p} \sum_{k=1}^{p} \frac{1}{n \sum_{\lambda_j(W_k) \in I} |X_k^*u_j(M_k)|^2}.$$ 

Now we prove by contradiction. If $N_I \geq Cn\eta$ for some constant $C > 100$, then there exists $k \in \{1, 2, \ldots, n\}$ such that

$$\sum_{\lambda_j(W_k) \in I} |X_k^*u_j(M_k)|^2 \geq Cn\eta.$$

Therefore

$$\sum_{\lambda_j(W_k) \in I} |X_k^*u_j(M_k)|^2 \leq \frac{n\eta}{C}.$$
By Cauchy interlacing law, $|\{\lambda_j(W_k) \in I\}| \geq N_I - 2 \geq N_I/2$. By Lemma 23, one concludes that
\[
\sum_{\lambda_j(W_k) \in I} |X_k^* u_j(M_k)|^2 \geq \frac{N_I}{4} \geq \frac{Cn\eta}{4}
\]
with probability at least $1 - n^{-(C_1 + 10)}$, assuming $C_2 \geq 10K(C_1 + 10)$. Thus we get contradiction $n\eta/C \geq Cn\eta/4$ by choosing $C$ large. This completes the proof.

Now we prove (3.3). First, by Schur’s complement, one can rewrite
\[
s(z) = \frac{1}{p} \text{Tr}(W - zI)^{-1} = \frac{1}{p} \sum_{k=1}^{p} \frac{1}{\lambda_j(W_k) - z} X_k^* u_j(M_k)
\]
(3.5)
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 $X_i^* \in \mathbb{C}^n$ ($1 \leq i \leq p$) 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
\[
\mathbb{E}(Y_k|W_k) = \sum_{j=1}^{p-1} \frac{1}{n} \frac{\lambda_j(W_k)}{\lambda_j(W_k) - z} = \frac{p-1}{n} + \frac{z}{n} \sum_{j=1}^{p-1} \frac{1}{\lambda_j(W_k) - z} = \frac{p-1}{n} (1 + z s_k(z))
\]
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

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

and thus

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

In fact a similar estimate holds for $Y_k$ itself:

**Proposition 36.** For $1 \leq k \leq n$, $|Y_k - E(Y_k|W_k)| \leq \delta^2 / \sqrt{C}$ holds with probability at least $1 - 20n^{-C}$ uniformly for all $z$ in the region $R_y$.

To prove Proposition 36, we decompose

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

$$:= \sum_{j=1}^{p-1} \frac{\lambda_j(W_k)}{n} \left( \frac{1}{\lambda_j(W_k) - x} - \frac{1}{\lambda_j(W_k) - \sqrt{-1}\eta} \right) R_j. \tag{3.6}$$

The estimation of (3.6) is a repetition of the calculation in (2.6). Without loss of generality, we may just consider the case $\lambda_j(W_k) - x \geq 0$. By assumption, we have $|\lambda_j(W_k)| = \Theta(1)$ and $|\lambda_j(W_k)| \leq 4$. First, for the set $J$ of eigenvalues $\lambda_j(W_k)$ such that $0 \leq \lambda_j(W_k) - x \leq \eta$, one has $|J| \leq n\eta$ and thus by Lemma 23, by taking $t = 4K\sqrt{C\log n}$,

$$\frac{1}{n} \left| \sum_{j \in J} \frac{\lambda_j(W_k)}{\lambda_j(W_k) - x - \sqrt{-1}\eta} R_j \right| \leq \frac{1}{n} \left| \sum_{j \in J} (\lambda_j(W_k) - x) \right| R_j + \frac{1}{n} \left| \sum_{j \in J} \frac{\eta \lambda_j(W_k)}{\lambda_j(W_k) - x} R_j \right| \leq \frac{10}{n\eta} (K\sqrt{C\log n} \sqrt{|J|} + K^2 C \log n) \leq \frac{20\delta^3}{\sqrt{C}}$$

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

For the other eigenvalues, we divide the real line into small intervals. For the set $J_t$ of eigenvalues $\lambda_j(W_k)$ such that $a \leq \lambda_j(W_k) - x \leq (1 + \alpha)a$, where we use the
parameters \( a = (1 + \alpha)^l \eta \) and \( \alpha = 10 \) (say), one has \(|J_1| \leq n a \alpha \). Thus by Lemma 23 (taking \( t = K \sqrt{C(l + 1) \log n} \)),

\[
\frac{1}{n} \left| \sum_{j \in J} \frac{\lambda_j(W_k) - x - \sqrt{-1} \eta}{\eta} R_j \right| \\
\leq \frac{1}{n} \left| \sum_{j \in J} \frac{(\lambda_j(W_k) - x) \lambda_j(W_k)}{\left( \lambda_j(W_k) - x \right)^2 + \eta^2} R_j \right| + \frac{1}{n} \left| \sum_{j \in J} \frac{\eta \lambda_j(W_k)}{\left( \lambda_j(W_k) - x \right)^2 + \eta^2} R_j \right| \\
\leq \frac{1 + \alpha}{n a} \left| \sum_{j \in J} \frac{a(\lambda_j(W_k) - x) \lambda_j(W_k)}{(1 + \alpha) \left( \lambda_j(W_k) - x \right)^2 + \eta^2} R_j \right| + \frac{\eta}{n \alpha^2} \left| \sum_{j \in J} \frac{a^2 \lambda_j(W_k)}{(\lambda_j(W_k) - x)^2 + \eta^2} R_j \right| \\
\leq \left( \frac{1 + \alpha}{n a} + \frac{\eta}{n \alpha^2} \right) \left( K \sqrt{C(l + 1) \log n} \sqrt{n \alpha a} + K^2 C(l + 1) \log n \right) \\
\leq \frac{20 \delta^3}{\sqrt{C}} \frac{l + 1}{(1 + \alpha)^{l/2}},
\]

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

Summing over \( l \), we have

\[
\frac{1}{n} \left| \sum_{l} \sum_{j \in J_l} \frac{\lambda_j(W_k) - x - \sqrt{-1} \eta}{\eta} R_j \right| \leq \frac{40 \delta^3}{\sqrt{C}}
\]

with probability at least \( 1 - 10n^{-C} \).

Thus \(|Y_k - \mathbf{E}(Y_k | W_k)| := \delta_C^2 \leq 200 \delta^2 / \sqrt{C} \). Therefore, inserting the bounds to (3.5), we have

\[
s(z) + \frac{1}{y + z - 1 + yz s(z) + \delta_C^2} = 0
\]

with probability at least \( 1 - 10n^{-C} \).

Recall the explicit expression of \( s_{MP,y}(z) \)

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

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 (3.5) and Proposition 36, we have with high probability that

\[
s(z) + \frac{1}{\frac{y}{n} + z - 1 + \frac{yz}{n} s(z) + \delta_C^2} = 0,
\]

where we used Lemma 23 to obtain that \( \xi_{kk} = ||X_k||^2/n = 1 + o(1) \) with probability at least \( 1 - n^{-C} \).
By assumption $p/n \to y$, when $n$ is large enough,

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

(3.7)

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

In (3.7), for the error term $\delta_C^2$, one has either $\frac{\delta_C^2}{y + z - 1 + yzs(z)} = O(\delta_C^2)$ 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 (3.7) (by choosing $C$ sufficiently large),

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

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

$$\sqrt{yz}(s(z) + \frac{y + z - 1}{2yz}) = \pm \sqrt{\frac{(y + z - 1)^2}{4yz} - 1 + O(\delta_C^2)}.$$

(3.8)

If $\frac{O(\delta_C^2)}{\sqrt{\frac{(y + z - 1)^2}{4yz} - 1}} = O(\delta_C)$, by a Taylor expansion on the right hand side of (3.8), we have

$$\sqrt{yz}(s(z) + \frac{y + z - 1}{2yz}) = \pm \sqrt{\frac{(y + z - 1)^2}{4yz} - 1 + O(\delta_C)}.$$

Therefore,

$$s(z) = s_{MP,y}(z) + O(\delta_C)$$

or

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

If $\frac{(y + z - 1)^2}{4yz} - 1 = o(1)$, from (3.8) and the explicit formula for $s_{MP,y}(z)$, we still have $s(z) = s_{MP,y}(z) + o(1)$.

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

$$s(z) = s_{MP,y}(z) + O(\delta_C),$$

(3.9)

or

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

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

(3.10)
or

\[ s(z) = -\frac{y + z - 1}{yz} + o(1). \] (3.11)

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

When \( \eta = n^{10}/\delta \), from \(|s(z)| \leq 1/\eta\) and \(|s_{MP,y}(z)| \leq 1/\eta\), we have \( s(z) \) and \( s_{MP,y}(z) \) are both \( O(\delta_C) \) and therefore (3.9) holds in this case. By continuity, we conclude that either (3.9) holds in the domain of interest or there exists some \( z \) in the domain such that (3.9) and (3.10) or (3.9) and (3.11) hold together.

On the other hand, (3.9) and (3.11) cannot hold at the same time. Otherwise, \( s_{MP,y}(z) + \frac{y + z - 1}{yz} = O(\delta_C) \). However, from \( s_{MP,y}(z)(s_{MP,y}(z) + \frac{y + z - 1}{yz}) = -\frac{1}{yz} \) and \(|s_{MP,y}(z)| \leq \frac{\sqrt{2}}{\sqrt{\eta}(1 - \sqrt{\eta} + \sqrt{\eta_0})}\), one can see that \(|s_{MP,y}(z) + \frac{y + z - 1}{yz}|\) is bounded from below, which implies a contradiction (by choosing \( C \) large enough).

Similarly, (3.9) or (3.10) cannot both hold except when \((y + z - 1)^2 - 4yz = o(1)\).

Otherwise, we can conclude that \( 2s_{MP,y}(z) + \frac{y + z - 1}{yz} = O(\delta_C) \). From the explicit formula of \( s_{MP,y} \),

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

One can conclude \(|2s_{MP,y}(z) + \frac{y + z - 1}{yz}|\) is bounded from below, which is a contradiction (by choosing \( C \) sufficiently large). Actually, if \((y + z - 1)^2 - 4yz = o(1)\), (3.9) and (3.10) are equivalent.

In conclusion, (3.9) holds with probability at least \( 1 - n^{-C} \) in the domain of interest. By Proposition 34, one can derive the following LMPL for random covariance matrices.

**Theorem 37.** 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 \( C_2 K^2 \log n/n^{\delta^8} \).

In particular, Theorem 33 follows.
3.3 Optimal upper bound for the infinity norm of singular vectors

**Theorem 38 (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 $1 \leq i \leq p$ 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

  $\|u_i\|_{\infty} \leq \frac{C_2 K \log^{1/2} n}{\sqrt{n}}$

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

- **(Edge case)** For any $\epsilon > 0$ and any $1 \leq i \leq p$ 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

  $\|u_i\|_{\infty} \leq \frac{C_2 K^2 \log n}{\sqrt{n}}$

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

3.3.1 Proof of the bulk case

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. By symmetry, it is enough to prove the delocalization for the right unit singular vectors.

**Lemma 39 (Corollary 25, [62]).** 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
\[
|y|^2 = \frac{1}{1 + \sum_{j=1}^{\min(p-1,n)} \frac{\sigma_j(M_{p-1,n})^2}{(\sigma_j(M_{p-1,n})^2 - \sigma_i(M_{p,n})^2)^2} |u_j(M_{p-1,n})^\ast Y|^2},
\]
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 \( M_{p-1,n} \).

**Proof.** First consider the right singular vectors. Since
\[
M_{p,n}^\ast M_{p,n} \begin{pmatrix} u \\ x \end{pmatrix} = \sigma_i(M_{p,n})^2 \begin{pmatrix} u \\ x \end{pmatrix},
\]
we have
\[
\begin{pmatrix} M_{p,n-1}^\ast M_{p,n-1} & M_{p,n-1}^\ast X \\ X^\ast M_{p,n-1} & X^\ast X \end{pmatrix} \begin{pmatrix} u \\ x \end{pmatrix} = \sigma_i(M_{p,n})^2 \begin{pmatrix} u \\ x \end{pmatrix}.
\]
Thus
\[
M_{p,n-1}^\ast M_{p,n-1} u + x M_{p,n-1}^\ast X = \sigma_i(M_{p,n})^2 u.
\]
By assumption, none of the singular values of \( M_{p,n-1} \) are equal to \( \sigma_i(M_{p,n}) \), we can solve
\[
u = x(M_{p,n-1}^\ast M_{p,n-1} - \sigma_i(M_{p,n})^2)^{-1} M_{p,n-1}^\ast X
\]
\[
= x \sum_{j=1}^{\min(p,n-1)} \frac{u_j(M_{p,n-1}) u_j(M_{p,n-1})^\ast}{(\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2)^2} M_{p,n-1}^\ast X
\]
\[
= x \sum_{j=1}^{\min(p,n-1)} \frac{\sigma_j(M_{p,n-1}) u_j(M_{p,n-1}) v_j(M_{p,n-1})^\ast X}{(\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2)^2}.
\]
The second equality follows from

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

On the other hand, from \(|x|^2 + \|u\|^2 = 1\), one has

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

The formula for the left singular vectors can be proved similarly.

If \(\lambda_i(W_{p,n})\) lies within the bulk of spectrum, by Theorem 37, 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 27, one has

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

with probability at least \(1 - n^{-C_1-10}\).

Thus, by the union bound and Lemma 39, \(\|u_i(M_{p,n})\|_\infty \leq \frac{C_2K \log^{1/2} n}{\sqrt{n}}\) holds with probability at least \(1 - n^{-C_1}\).

3.3.2 Proof of the edge case

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 28.

Lemma 40 (Interlacing identity for singular values, Lemma 3.5 [69]). Assume the notations in Lemma 39, then for every \(i\),

\[
\sum_{j=1}^{\min(p,n-1)} \frac{\sigma_j(M_{p,n-1})^2}{\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2} |v_j(M_{p,n-1})^*X|^2 = \|X\|^2 - \sigma_i(M_{p,n})^2. \tag{3.12}
\]
Similarly, we have
\[
\min(p-1,n) \sum_{j=1}^{\min(p-1,n)} \frac{\sigma_j(M_{p-1,n})^2 |u_j(M_{p-1,n})^* Y|^2}{\sigma_j(M_{p-1,n})^2 - \sigma_i(M_{p,n})^2} = \|Y\|^2 - \sigma_i(M_{p,n})^2. \tag{3.13}
\]

**Proof.** Apply Lemma 28 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
\[
u_j(M_{p,n-1}^* M_{p,n-1}) = \sigma_j(M_{p,n-1}) \nu_j(M_{p,n-1})^*,
\]
(3.12) follows.

Similarly, to show (3.13), apply Lemma 40 to the matrix
\[
M_{p,n}^* = \begin{pmatrix}
M_{p-1,n}^* M_{p-1,n} & M_{p-1,n}^* Y \\
Y^* M_{p-1,n} & \|Y\|^2
\end{pmatrix}.
\]

By the union bound and Lemma 27, 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
\[
\min(p,n-1) \sum_{j=1}^{\min(p,n-1)} \frac{\sigma_j(M_{p,n-1})^2}{(\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2)^2} |v_j(M_{p,n-1})^* X|^2 \geq \frac{n}{C^4 K^4 \log^2 n}.
\]

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 \(1 - n^{-C_1 - 100}\),
\[
\min(p,n-1) \sum_{j=1}^{\min(p,n-1)} \frac{\sigma_j(M_{p,n-1})^2}{(\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2)^2} |v_j(M_{p,n-1})^* X|^4 \geq \frac{n}{C^3 K^2 \log n}.
\]

By Cauchy-Schwarz 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),
\[
\sum_{T_- \leq j \leq T_+} \frac{\frac{1}{n} \sigma_j(M_{p,n-1})^2}{(\sigma_j(M_{p,n-1})^2 - \sigma_i(M_{p,n})^2)} |v_j(M_{p,n-1})^* X|^2 \geq \frac{\sqrt{T_+ - T_-}}{C^{3.5} K \sqrt{\log n}}.
\]
On the other hand, by the projection lemma, with probability at least $1 - n^{-C_1-100}$, $\|X\|^2/n = y + o(1)$. By (3.12) in Lemma 40,

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

(3.14)

It is enough to evaluate

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

(3.15)

Now we divide the real line into disjoint intervals $I_k$ for $k \geq 0$. Let $|I| = \frac{K^2 C \log n}{n \delta s}$ 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 19, the number of eigenvalues

$$
|J_k| = N_{I_k} \leq n\alpha_{I_k}|I_k| + \delta^kn|I_k|
$$

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 23, for the $k$th interval, with probability at least $1 - n^{-C_1-100}$,

$$
\frac{1}{n} \sum_{j \in I_k} \frac{|\lambda_j(W_{p,n-1})||v_j(M_{p,n-1})X|^2}{|\lambda_j(W_{p,n-1}) - \lambda_i(W_{p,n})|} \leq \frac{1}{n} \left(1 + \frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_{p,n}), I_k)}\right) \sum_{j \in I_k} |v_j(W_{p,n-1})X|^2
$$

$$
\leq \frac{1}{n} \left(1 + \frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_{p,n}), I_k)}\right) (|J_k| + K\sqrt{|I_k|\sqrt{\log n} + CK^2\log n})
$$

$$
\leq \frac{1}{n} \left(1 + \frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_{p,n}), I_k)}\right) (p\alpha_{I_k}|I_k| + \delta^k|I_k| + 4K\sqrt{C\log n \sqrt{n}\sqrt{|I_k|} + CK^2\log n})
$$

$$
\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|. \]
For $k_0 + 1 \leq k \leq N$, let the interval $I_k$’s have the same length of $|I_{k_0}| = 2\delta^{-8k_0}|I|$. 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
\[
\frac{1}{n} \sum_{j \in J_k} \frac{\lambda_i(W_{p,n-1})|v_j(M_{p,n-1})^*X|^2}{|\lambda_i(W_{p,n-1}) - \lambda_i(W_{p,n})|} \leq \frac{1}{n} (1 + \frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_{p,n}), I_k)}) \sum_{j \in J_k} |v_j(M_{p,n-1})^*X|^2 \leq \frac{1}{n} (1 + \frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_{p,n}), I_k)}) (|J_k| + K \sqrt{|J_k|} \sqrt{C \log n} + CK^2 \log n) \leq y(1 + \frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_{p,n}), I_k)}) \alpha_k |I_k| + \frac{100\delta^{k_0-8}}{k - k_0},
\]
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,
\[
y \sum_{I_k} (1 + \frac{\lambda_i(W_{p,n})}{\text{dist}(\lambda_i(W_{p,n}), I_k)}) \alpha_k |I_k| = |p.v. \int_{a}^{b} y \frac{x p_{MP,y}(x)}{x - \lambda_i(W_{p,n})} dx| + o(1)
\]
\[
= y \left(1 + p.v. \lambda_i(W_{p,n}) \int_{a}^{b} \frac{\rho_{MP,y}(x)}{x - \lambda_i(W_{p,n})} dx\right),
\]
where
\[
p.v. \int_{a}^{b} y \frac{x p_{MP,y}(x)}{x - \lambda_i(W_{p,n})} dx = \begin{cases} 
\sqrt{y} + o(1), & \text{if } |\lambda_i(W_{p,n}) - a| = o(1), \\
\sqrt{y} - o(1), & \text{if } |\lambda_i(W_{p,n}) - b| = o(1).
\end{cases}
\]
by using the explicit formula for the Stieltjes transform and from residue calculus (see below).

If $|\lambda_i(W_{p,n}) - a| = o(1)$, using the formula for the Stieltjes transform, one obtains from residue calculus that
\[
p.v. \int_{a}^{b} y \frac{x p_{MP,y}(x)}{x - \lambda_i(W_{p,n})} dx = y \left(1 + p.v. \lambda_i(W_{p,n}) \int_{a}^{b} \frac{\rho_{MP,y}(x)}{x - \lambda_i(W_{p,n})} dx\right)
\]
\[
= y \left(1 + (1 - \sqrt{y})^2 \frac{1}{\sqrt{y} - y} \right) + o(1)
\]
\[
= \sqrt{y} + o(1).
\]
If $|\lambda_i(W_{p,n}) - b| = o(1)$, we have

\[
\text{p.v.} \int_a^b \frac{y \rho_{MP,y}(x)}{x - \lambda_i(W_{p,n})} \, dx = y \left( 1 + \text{p.v.} \frac{\rho_{MP,y}(x)}{x - \lambda_i(W_{p,n})} \right)
\]

\[
= y \left( 1 - \left(1 + \sqrt{y}\right)^2 \frac{1}{\sqrt{y} + y} \right) + o(1)
\]

\[
= -\sqrt{y} + o(1).
\]

Now for the rest of eigenvalues such that

\[
|\lambda_i(W_{p,n}) - \lambda_j(W_{p,n-1})| \leq |J_0| + |J_1| + \ldots + |J_{20}| \leq |I|/\delta^{60},
\]

the number of eigenvalues is given by $T_+ - T_- \ll n|I|/\delta^{60} = CK^2 \log n/\delta^{68}$. Thus

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

by choosing $C$ sufficiently large. By comparing (3.14), (3.15) and (3.17), one can conclude with probability at least $1 - n^{-C_1 - 10}$,

\[
|x| \leq \frac{C_2 K^2 \log n}{\sqrt{n}}.
\]
Chapter 4
Adjacency matrices of random graphs

4.1 Introduction

In this chapter, we consider the Erdős-Rényi random graph $G(n, p)$. Given a real number $p = p(n), 0 \leq p \leq 1$, the Erdős-Rényi graph on a vertex set of size $n$ is obtained by drawing an edge between each pair of vertices, randomly and independently, with probability $p$.

Given a graph $G$ on $n$ vertices, the adjacency matrix $A$ of $G$ is an $n \times n$ matrix whose entry $a_{ij}$ equals one if there is an edge between the vertices $i$ and $j$ and zero otherwise. All diagonal entries $a_{ii}$ are defined to be zero. The eigenvalues and eigenvectors of $A$ carry valuable information about the structure of the graph and have been studied by many researchers for quite some time, with both theoretical and practical motivations (see, for example, [9], [12], [33], [53] [37], [34], [36], [35], [66], [30], [54], [52]).

Let $A_n$ be the adjacency matrix of $G(n, p)$. Thus $A_n$ is a random symmetric $n \times n$ matrix whose upper triangular entries are independent identical distributed (iid) copies of a real random variable $\xi$ and diagonal entries are 0. $\xi$ is a Bernoulli random variable that takes values 1 with probability $p$ and 0 with probability $1 - p$.

$$E\xi = p, Var\xi = p(1 - p) = \sigma^2.$$  

Usually it is more convenient to study the normalized matrix

$$M_n = \frac{1}{\sigma}(A_n - pJ_n),$$

where $J_n$ is the $n \times n$ matrix all of whose entries are 1. $M_n$ has entries with mean zero and variance one. The global properties of the eigenvalues of $A_n$ and $M_n$ are essentially the same (after proper scaling), thanks to the following lemma.
Figure 4.1: Plotted above is the probability density function of the ESD of $G(2000, 0.2)$.

**Lemma 41.** (Lemma 36, [66]) Let $A, B$ be symmetric matrices of the same size where $B$ has rank one. Then for any interval $I$,

$$|N_I(A + B) - N_I(A)| \leq 1,$$

where $N_I(M)$ is the number of eigenvalues of $M$ in $I$.

**Theorem 42.** For $p = \omega(\frac{1}{n^2})$, the empirical spectral distribution (ESD) of the matrix $\frac{1}{\sqrt{n}\sigma}A_n$ converges in distribution to the semicircle distribution which has a density $\rho_{sc}(x)$ with support on $[-2, 2]$,

$$\rho_{sc}(x) := \frac{1}{2\pi\sigma} \sqrt{4 - x^2}.$$

If $np = O(1)$, the semicircle law no longer holds. In this case, the graph almost surely has $\Theta(n)$ isolated vertices, so in the limiting distribution, the point 0 will have positive constant mass.

In [20], Dekel, Lee and Linial, motivated by the study of nodal domains, raised the following question.

**Question 43.** Is it true that almost surely every eigenvector $u$ of $G(n, p)$ has $\|u\|_\infty = o(1)$?
Later, in their journal paper [21], the authors added one sharper question.

**Question 44.** Is it true that almost surely every eigenvector $u$ of $G(n, p)$ has $\|u\|_{\infty} = n^{-1/2+o(1)}$?

The bound $n^{-1/2+o(1)}$ was also conjectured by V. Vu of this paper in an NSF proposal (submitted Oct 2008). He and Tao [66] proved this bound for eigenvectors corresponding to the eigenvalues in the bulk of the spectrum for the case $p = 1/2$. If one defines the adjacency matrix by writing $-1$ for non-edges, then this bound holds for all eigenvectors [66, 64].

The above two questions were raised under the assumption that $p$ is a constant in the interval $(0, 1)$. For $p$ depending on $n$, the statements may fail. If $p \leq \frac{(1-\epsilon)\log n}{n}$, then the graph has (with high probability) isolated vertices and so one cannot expect that $\|u\|_{\infty} = o(1)$ for every eigenvector $u$. We raise the following questions:

**Question 45.** Assume $p \geq \frac{(1+\epsilon)\log n}{n}$ for some constant $\epsilon > 0$. Is it true that almost surely every eigenvector $u$ of $G(n, p)$ has $\|u\|_{\infty} = o(1)$?

**Question 46.** Assume $p \geq \frac{(1+\epsilon)\log n}{n}$ for some constant $\epsilon > 0$. Is it true that almost surely every eigenvector $u$ of $G(n, p)$ has $\|u\|_{\infty} = \frac{n^{-1/2+o(1)}}{n}$?

Our main result settles (positively) Question 44 and almost Question 45. This result follows from Corollary 52 obtained in Section 2.

**Theorem 47.** (Infinity norm of eigenvectors) Let $p = \omega(\log n/n)$ and let $A_n$ be the adjacency matrix of $G(n, p)$. Then there exists an orthonormal basis of eigenvectors of $A_n$, $\{u_1, \ldots, u_n\}$, such that for every $1 \leq i \leq n$, $\|u_i\|_{\infty} = o(1)$ almost surely.

For Questions 43 and 46, we obtain a good quantitative bound for those eigenvectors which correspond to eigenvalues bounded away from the edge of the spectrum.

**Definition 48.** Let $E$ be an event depending on $n$. Then $E$ holds with overwhelming probability if $P(E) \geq 1 - \exp(-\omega(\log n))$.

For convenience, in the case when $p = \omega(\log n/n) \in (0, 1)$, we write

$$p = \frac{g(n) \log n}{n},$$
where \( g(n) \) is a positive function such that \( g(n) \rightarrow \infty \) as \( n \rightarrow \infty \) (\( g(n) \) can tend to \( \infty \) arbitrarily slow).

**Theorem 49.** Assume \( p = g(n) \log n / n \in (0, 1) \), where \( g(n) \) is defined as above. Let \( B_n = \frac{1}{\sqrt{n \sigma}} A_n \). For any \( \kappa > 0 \), and any \( 1 \leq i \leq n \) with \( \lambda_i(B_n) \in [-2 + \kappa, 2 - \kappa] \), there exists a corresponding eigenvector \( u_i \) such that

\[
\|u_i\|_{\infty} = O\left( \sqrt{\log(2) g(n) \log n} \right) np
\]

with overwhelming probability.

### 4.2 A small perturbation lemma

\( A_n \) is the adjacency matrix of \( G(n, p) \). In the proofs of Theorem 47 and Theorem 49, we actually work with the eigenvectors of a perturbed matrix

\[
A_n + \epsilon N_n,
\]

where \( \epsilon = \epsilon(n) > 0 \) can be arbitrarily small and \( N_n \) is a symmetric random matrix whose upper triangular elements are independent with a standard Gaussian distribution.

The entries of \( A_n + \epsilon N_n \) are continuous and thus with probability 1, the eigenvalues of \( A_n + \epsilon N_n \) are simple. Let

\[
\mu_1 < \ldots < \mu_n
\]

be the ordered eigenvalues of \( A_n + \epsilon N_n \), which have a unique orthonormal system of eigenvectors \( \{w_1, \ldots, w_n\} \). By the Cauchy interlacing principle, the eigenvalues of \( A_n + \epsilon N_n \) are different from those of its principal minors, which satisfies a condition of Lemma 27.

Let \( \lambda_i \)'s be the eigenvalue of \( A_n \) with multiplicity \( k_i \) that are defined in increasing order:

\[
\ldots \lambda_{i-1} < \lambda_i = \lambda_{i+1} = \ldots = \lambda_{i+k_i} < \lambda_{i+k_i+1} \ldots
\]

By Weyl’s theorem, one has for every \( 1 \leq j \leq n \),

\[
|\lambda_j - \mu_j| \leq \epsilon \|N_n\|_{op} = O(\epsilon \sqrt{n}). \tag{4.1}
\]

Thus the behaviors of eigenvalues of \( A_n \) and \( A_n + \epsilon N_n \) are essentially the same by choosing \( \epsilon \) sufficiently small. And everything (except Lemma 27) we used in the proofs
Theorem 47 and Theorem 49 for $A_n$ also applies for $A_n + \epsilon N_n$ by a continuity argument. We will not distinguish $A_n$ from $A_n + \epsilon N_n$ in the proofs.

The following lemma will allow us to transfer the eigenvector delocalization results of $A_n + \epsilon N_n$ to those of $A_n$ at some expense.

**Lemma 50.** With the notation above, there exists an orthonormal basis of eigenvectors of $A_n$, denoted by $\{u_1, \ldots, u_n\}$, such that for every $1 \leq j \leq n$,

$$||u_j||_\infty \leq ||w_j||_\infty + \alpha(n),$$

where $\alpha(n)$ can be arbitrarily small provided $\epsilon(n)$ is small enough.

**Proof.** First, since the coefficients of the characteristic polynomial of $A_n$ are integers, there exists a positive function $l(n)$ such that either $|\lambda_s - \lambda_t| = 0$ or $|\lambda_s - \lambda_t| \geq l(n)$ for any $1 \leq s, t \leq n$.

By (4.1) and choosing $\epsilon$ sufficiently small, one can get

$$|\mu_i - \lambda_{i-1}| > l(n) \quad \text{and} \quad |\mu_{i+k_i} - \lambda_{i+k_i+1}| > l(n)$$

For a fixed index $i$, let $E$ be the eigenspace corresponding to the eigenvalue $\lambda_i$ and $F$ be the subspace spanned by $\{w_i, \ldots, w_{i+k_i}\}$. Both of $E$ and $F$ have dimension $k_i$.

Let $P_E$ and $P_F$ be the orthogonal projection matrices onto $E$ and $F$ separately.

Applying the well-known Davis-Kahan theorem (see [59] Section IV, Theorem 3.6) to $A_n$ and $A_n + \epsilon N_n$, one gets

$$||P_E - P_F||_{\text{op}} \leq \frac{\epsilon ||N_n||_{\text{op}}}{l(n)} := \alpha(n),$$

where $\alpha(n)$ can be arbitrarily small depending on $\epsilon$.

Define $v_j = P_Ew_j \in E$ for $i \leq j \leq i + k_i$, then we have $||v_j - w_j||_2 \leq \alpha(n)$. It is clear that $\{v_i, \ldots, v_{k_i}\}$ are eigenvectors of $A_n$ and

$$||v_j||_\infty \leq ||w_j||_\infty + ||v_j - w_j||_2 \leq ||w_j||_\infty + \alpha(n).$$

By choosing $\epsilon$ small enough such that $n\alpha(n) < 1/2$, $\{v_i, \ldots, v_{k_i}\}$ are linearly independent. Indeed, if $\sum_{j=i}^{k_i} c_j v_j = 0$, one has for every $i \leq s \leq i + k_i, \sum_{j=i}^{k_i} c_j \langle P_E w_j, w_s \rangle = 0$. 

0, which implies \( c_s = -\sum_{j=i}^{k_s} h_j c_j (P_E w_j - w_s, w_s) \). Thus \(|c_s| \leq \alpha(n) \sum_{j=i}^{k_s} |c_j| \), summing over all \( s \), we can get \( \sum_{j=i}^{k_s} |c_j| \leq k\alpha(n) \sum_{j=i}^{k_s} |c_j| \) and therefore \( c_j = 0 \).

Furthermore the set \( \{v_i, \ldots, v_{k_i}\} \) is “almost” an orthonormal basis of \( E \) in the sense that

\[
\left| \left| \left| v_s \right| \right|_2^2 - 1 \right| \leq \left| \left| v_s - w_s \right| \right|_2 \leq \alpha(n) \quad \text{for any } i \leq s \leq i + k_i
\]

\[
|\langle v_s, v_t \rangle| = |\langle P_E w_s, P_E w_t \rangle| = |\langle P_E w_s - w_s, P_E w_t \rangle + \langle w_s, P_E w_t - w_t \rangle| = O(\alpha(n)) \quad \text{for any } i \leq s \neq t \leq i + k_i
\]

We can perform a Gram-Schmidt process on \( \{v_i, \ldots, v_{k_i}\} \) to get an orthonormal system of eigenvectors \( \{u_i, \ldots, u_{k_i}\} \) on \( E \) such that

\[
\left| \left| u_j \right| \right|_\infty \leq \left| \left| w_j \right| \right|_\infty + \alpha(n),
\]

for every \( i \leq j \leq i + k_i \).

We iterate the above argument for every distinct eigenvalue of \( A_n \) to obtain an orthonormal basis of eigenvectors of \( A_n \).

\[\square\]

### 4.3 Proof of Theorem 47

A key ingredient is the following concentration lemma.

**Lemma 51.** Let \( M \) be a \( n \times n \) Hermitian random matrix whose off-diagonal entries \( \xi_{ij} \) are i.i.d. random variables with mean zero, variance 1 and \( |\xi_{ij}| < K \) for some common constant \( K \). Fix \( \delta > 0 \) and assume that the fourth moment \( M_4 := \sup_{i,j} E(|\xi_{ij}|^4) = o(n) \). Then for any interval \( I \subset [-2, 2] \) whose length is at least \( \Omega(\delta^{-2/3} (M_4/n)^{1/3}) \), there is a constant \( c \) such that the number \( N_I \) of the eigenvalues of \( W_n = \frac{1}{\sqrt{n}} M \) which belong to \( I \) satisfies the following concentration inequality

\[
P(|N_I - n \int_I \rho_{sc}(t) dt| > \delta n \int_I \rho_{sc}(t) dt) \leq 4 \exp(-c\frac{\delta^4 n^2 |I|^5}{K^2}).
\]

The proof of Lemma 51 uses the approach of Guionnet and Zeitouni in [39]. Intuitively, one tries to apply Lemma 8 with

\[
Z = N_I = \sum_{i=1}^{n} \mathbf{1}_{\{\lambda_i(W_n) \in I\}}.
\]
However, the characteristic function $1_{\{\lambda_i(W_n) \in I\}}$ is neither convex nor Lipschitz. Thus we construct two auxiliary functions (piecewise linear, convex and Lipschitz) suggested as in [39] to overcome the technical difficulty (see [67] for details).

Apply Lemma 51 for the normalized adjacency matrix $M_n$ of $G(n, p)$ with $K = 1/\sqrt{p}$ we obtain

**Theorem 52.** Consider the model $G(n, p)$ with $np \to \infty$ as $n \to \infty$ and let $\delta > 0$. Then for any interval $I \subset [-2, 2]$ with length at least $\left(\frac{\log(np)}{\sigma^4(np)^{1/2}}\right)^{1/5}$, we have

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

with probability at most $\exp(-cn(np)^{1/2} \log(np))$.

**Lemma 53.** Let $Y = (\zeta_1, \ldots, \zeta_n) \in \mathbb{C}^n$ be a random vector whose entries are i.i.d. copies of the random variable $\zeta = \xi - p$ (with mean 0 and variance $\sigma^2$). Let $H$ be a subspace of dimension $d$ and $\pi_H$ the orthogonal projection onto $H$. Then

$$\mathbf{P}(\| \| \pi_H(Y) \|-\sigma \sqrt{d}\| \geq t) \leq 10 \exp(-\frac{t^2}{4}).$$

In particular,

$$\| \pi_H(Y) \|= \sigma \sqrt{d} + O(\omega(\sqrt{\log n})) \quad (4.2)$$

with overwhelming probability.

**Proof.** The coordinates of $Y$ are bounded in magnitude by 1. Apply Talagrand’s inequality to the map $Y \to \|\pi_H(Y)\|$, which is convex and 1-Lipschitz. We can conclude

$$\mathbf{P}(\| \| \pi_H(Y) \|-M(\| \pi_H(Y) \|)\| \geq t) \leq 4 \exp(-\frac{t^2}{16}) \quad (4.3)$$

where $M(\| \pi_H(Y) \|)$ is the median of $\| \pi_H(Y) \|$.

Let $P = (p_{ij})_{1 \leq i,j \leq n}$ be the orthogonal projection matrix onto $H$.

One has $\text{trace}(P^2) = \text{trace}(P) = \sum_i p_{ii} = d$ and $|p_{ii}| \leq 1$, as well as,

$$\| \pi_H(Y) \|^2 = \sum_{1 \leq i,j \leq n} p_{ij} \zeta_i \zeta_j = \sum_{i=1}^n p_{ii} \zeta_i^2 + \sum_{i \neq j} p_{ij} \zeta_i \zeta_j$$
and
\[ \mathbf{E}\| \pi_H(Y) \|^2 = \mathbf{E}(\sum_{i=1}^n p_{ii} \zeta_i^2) + \mathbf{E}(\sum_{i \neq j} p_{ij} \zeta_i \zeta_j) = \sigma^2 d. \]

Take \( L = 4/\sigma \). To complete the proof, it suffices to show
\[ |M(\| \pi_H(Y) \|) - \sigma \sqrt{d}| \leq L \sigma. \tag{4.4} \]

Consider the event \( \mathcal{E}_+ \) that \( \| \pi_H(Y) \| \geq \sigma L + \sigma \sqrt{d} \), which implies that \( \| \pi_H(Y) \|^2 \geq \sigma^2 (L^2 + 2L\sqrt{d} + d^2). \)

Let \( S_1 = \sum_{i=1}^n p_{ii}(\zeta_i^2 - \sigma^2) \) and \( S_2 = \sum_{i \neq j} p_{ij} \zeta_i \zeta_j \). Now we have
\[ \mathbf{P}(\mathcal{E}_+) \leq \mathbf{P}(\sum_{i=1}^n p_{ii} \zeta_i^2 \geq \sigma^2 d + L\sqrt{d}\sigma^2) + \mathbf{P}(\sum_{i \neq j} p_{ij} \zeta_i \zeta_j \geq \sigma^2 L\sqrt{d}). \]

By Chebyshev’s inequality,
\[ \mathbf{P}(\sum_{i=1}^n p_{ii} \zeta_i^2 \geq \sigma^2 d + L\sqrt{d}\sigma^2) = \mathbf{P}(S_1 \geq L\sqrt{d}\sigma^2) \leq \frac{\mathbf{E}(|S_1|^2)}{L^2 \sigma^4}, \]
where \( \mathbf{E}(|S_1|^2) = \mathbf{E}(\sum_i p_{ii}(\zeta_i^2 - \sigma^2)^2) = \sum_i p_{ii} \mathbf{E}(\zeta_i^4 - \sigma^4) \leq d\sigma^2(1 - 2\sigma^2) \). Therefore,
\[ \mathbf{P}(S_1 \geq L\sqrt{d}\sigma^4) \leq \frac{d\sigma^2(1 - 2\sigma^2)}{L^2 \sigma^4} < \frac{1}{16}. \]

On the other hand, we have \( \mathbf{E}(|S_2|^2) = \mathbf{E}(\sum_{i \neq j} p_{ij} \zeta_i^2 \zeta_j^2) \leq \sigma^4 d \) and
\[ \mathbf{P}(\sum_{i \neq j} p_{ij} \zeta_i \zeta_j \geq \sigma^2 L\sqrt{d}) = \mathbf{P}(S_2 \geq L\sqrt{d}\sigma^2) \leq \frac{\mathbf{E}(|S_2|^2)}{L^2 \sigma^4} < \frac{1}{10}. \]
It follows that \( \mathbf{P}(\mathcal{E}_+) < 1/4 \) and hence \( M(\| \pi_H(Y) \|) \leq L \sigma + \sqrt{d}\sigma. \)

For the lower bound, consider the event \( \mathcal{E}_- \) that \( \| \pi_H(Y) \| \leq \sqrt{d}\sigma - L \sigma \) and notice that
\[ \mathbf{P}(\mathcal{E}_-) \leq \mathbf{P}(S_1 \leq -L\sqrt{d}\sigma^2) + \mathbf{P}(S_2 \leq -L\sqrt{d}\sigma^2). \]

The same argument applies to get \( M(\| \pi_H(Y) \|) \geq \sqrt{d}\sigma - L \sigma \). Now the relations (4.3) and (4.4) together imply (4.2).

Let \( \lambda_n(A_n) \) be the largest eigenvalue of \( A_n \) and \( u = (u_1, \ldots, u_n) \) be the corresponding unit eigenvector. We have the lower bound \( \lambda_n(A_n) \geq np \). And if \( np = \omega(\log n) \), then the maximum degree \( \Delta = (1 + o(1))np \) almost surely (See Corollary 3.14, [15]).
For every $1 \leq i \leq n$,
\[
\lambda_n(A_n) u_i = \sum_{j \in N(i)} u_j,
\]
where $N(i)$ is the neighborhood of vertex $i$. Thus, by Cauchy-Schwarz inequality,
\[
\|u\|_\infty = \max_i \left| \sum_{j \in N(i)} \frac{u_j}{\lambda_n(A_n)} \right| = \frac{1}{\sqrt{\delta n^2}} \sum_{j \in N(i)} \frac{u_j}{\lambda_n(A_n)} = O\left( \frac{1}{\sqrt{n\delta}} \right).
\]

Let $B_n = \frac{1}{\sqrt{n\delta}} A_n$. Since the eigenvalues of $W_n = \frac{1}{\sqrt{n\delta}} (A_n - pJ_n)$ are on the interval $[-2, 2]$, by Lemma 41, $\{\lambda_1(B_n), \ldots, \lambda_{n-1}(B_n)\} \subset [-2, 2]$.

Recall that $np = g(n) \log n$. By Corollary 52, for any interval $I$ with length at least $(\frac{\log(np)}{g^4(np)^4})^{1/5}$, with overwhelming probability, if $I \subset [-2 + \kappa, 2 - \kappa]$ for some positive constant $\kappa$, one has $N_I(B_n) = \Theta(n \int_I \rho_{sc}(x) dx) = \Theta(n|I|)$; if $I$ is at the edge of $[-2, 2]$, with length $o(1)$, one has $N_I(B_n) = \Theta(n \int_I \rho_{sc}(x) dx) = \Theta(n|I|^{3/2})$. Thus we can find a set $J \subset \{1, \ldots, n - 1\}$ with $|J| = \Omega(n|I_0|)$ or $|J| = \Omega(n|I_0|^{3/2})$ such that $|\lambda_j(B_{n-1}) - \lambda_i(B_n)| \leq |I_0|$ for all $j \in J$, where $B_{n-1}$ is the bottom right $(n-1) \times (n-1)$ minor of $B_n$. Here we take $|I_0| = (1/g(n)^{1/20})^{2/3}$. It is easy to check that $|I_0| \geq (\frac{\log(np)}{g^4(np)^4})^{1/5}$.

By the formula in Lemma 27, the entry of the eigenvector of $B_n$ can be expressed as
\[
|x|^2 = \frac{1}{1 + \sum_{j=1}^{n-1} (\lambda_j(B_{n-1}) - \lambda_i(B_n))^{-2} |u_j(B_{n-1})^* \frac{1}{\sqrt{n\sigma}} X|^2} \leq \frac{1}{1 + \sum_{j \in J} (\lambda_j(B_{n-1}) - \lambda_i(B_n))^{-2} |u_j(B_{n-1})^* \frac{1}{\sqrt{n\sigma}} X|^2} \leq \frac{1}{1 + \sum_{j \in J} n^{-1} |I_0|^{-2} |u_j(B_{n-1})^* \frac{1}{\sigma} X|^2} = \frac{1}{1 + n^{-1} |I_0|^{-2} ||\pi_H(\frac{X}{\sigma})||^2} \leq \frac{1}{1 + n^{-1} |I_0|^{-2} |J|}
\]
with overwhelming probability, where $H$ is the span of all the eigenvectors associated to $J$ with dimension $\dim(H) = \Theta(|J|)$, $\pi_H$ is the orthogonal projection onto $H$ and $X \in \mathbb{C}^{n-1}$ has entries that are iid copies of $\xi$. The last inequality in (4.5) follows from Lemma 53 (by taking $t = g(n)^{1/10} \sqrt{\log n}$) and the relations
\[
||\pi_H(X)|| = ||\pi_H(Y + p1_n)|| \geq ||\pi_{H_1}(Y + p1_n)|| \geq ||\pi_{H_1}(Y)||.
\]
Here $Y = X - p1_n$ and $H_1 = H \cap H_2$, where $H_2$ is the space orthogonal to the all $1$ vector $1_n$. For the dimension of $H_1$, $\dim(H_1) \geq \dim(H) - 1$. 


Since either $|J| = \Omega(n|I_0|)$ or $|J| = \Omega(n|I_0|^{3/2})$, we have $n^{-1}|I_0|^{-2}|J| = \Omega(|I_0|^{-1})$ or $n^{-1}|I_0|^{-2}|J| = \Omega(|I_0|^{-1/2})$. Thus $|x|^2 = O(|I_0|)$ or $|x|^2 = O(\sqrt{|I_0|})$. In both cases, since $|I_0| \to 0$, it follows that $|x| = o(1)$. 

4.4 Proof of Theorem 49

The following concentration lemma for $G(n,p)$ will be a key input to prove Theorem 49. The proof is very similar to that of Theorem 25 and is omitted here. Let $B_n = \frac{1}{\sqrt{n\sigma}}A_n$.

Lemma 54 (Concentration for ESD in the bulk). Assume $p = g(n) \log n/n$. For any constants $\varepsilon, \delta > 0$ and any interval $I$ in $[-2+\varepsilon, 2-\varepsilon]$ with $|I| = \Omega((\log^2 g(n) \log n/np)$, the number of eigenvalues $N_I$ of $B_n$ in $I$ obeys the concentration estimate

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

with overwhelming probability.

With the formula in Lemma 27, it suffices to show the following lower bound

$$\sum_{j=1}^{n-1} (\lambda_j(B_{n-1}) - \lambda_i(B_n))^2 u_j(B_{n-1})^* \frac{1}{\sqrt{n\sigma}} X |^2 \geq \frac{np}{\log^2 g(n) \log n}$$

(4.6)

with overwhelming probability, where $B_{n-1}$ is the bottom right $n-1 \times n-1$ minor of $B_n$ and $X \in \mathbb{C}^{n-1}$ has entries that are iid copies of $\xi$. Recall that $\xi$ takes values 1 with probability $p$ and 0 with probability $1-p$, thus $\mathbb{E}\xi = p, \forall\text{var}\xi = p(1-p) = \sigma^2$.

By Lemma 54, we can find a set $J \subset \{1, \ldots, n-1\}$ with $|J| \geq \log^2 g(n) \log n/\sigma$ such that $|\lambda_j(B_{n-1}) - \lambda_i(B_n)| = O(\log^2 g(n) \log n/np)$ for all $j \in J$. Thus in (4.6), it is enough to prove

$$\sum_{j \in J} |u_j(B_{n-1})^T \frac{1}{\sigma} X |^2 = ||\pi_H(\frac{X}{\sigma})||^2 \gg |J|$$

or equivalently

$$||\pi_H(X)||^2 \gg \sigma^2 |J|$$

(4.7)

with overwhelming probability, where $H$ is the span of all the eigenvectors associated to $J$ with dimension $\text{dim}(H) = \Theta(|J|)$. 

□
Let $H_1 = H \cap H_2$, where $H_2$ is the space orthogonal to $1_n$. The dimension of $H_1$ is at least $\dim(H) - 1$. Denote $Y = X - p1_n$. Then the entries of $Y$ are iid copies of $\zeta$. By Lemma 53,

$$||\pi_{H_1}(Y)||^2 \gg \sigma^2|J|$$

with overwhelming probability.

Hence, our claim follows from the relations

$$||\pi_H(X)|| = ||\pi_H(Y + p1_n)|| \geq ||\pi_{H_1}(Y + p1_n)|| = ||\pi_{H_1}(Y)||.$$

4.5 Proof of Theorem 42

We will show that the semicircle law holds for $M_n$. With Lemma 41, it is clear that Theorem 42 follows Lemma 55 directly. The claim actually follows as a special case discussed in the paper [18]. Our proof here uses a standard moment method.

**Lemma 55.** For $p = \omega(\frac{1}{n})$, the empirical spectral distribution (ESD) of the matrix $W_n = \frac{1}{\sqrt{n}}M_n$ converges in distribution to the semicircle law which has a density $\rho_{sc}(x)$ with support on $[-2, 2]$,

$$\rho_{sc}(x) := \frac{1}{2\pi}\sqrt{4 - x^2}.$$

Let $\eta_{ij}$ be the entries of $M_n = \sigma^{-1}(A_n - pJ_n)$. For $i = j$, $\eta_{ij} = -p/\sigma$; and for $i > j$, $\eta_{ij}$ are iid copies of random variable $\eta$, which takes value $(1 - p)/\sigma$ with probability $p$ and takes value $-p/\sigma$ with probability $1 - p$.

$$\mathbb{E}\eta = 0, \mathbb{E}\eta^2 = 1, \mathbb{E}\eta^s = O\left(\frac{1}{(\sqrt{p})^{s-2}}\right) \text{ for } s \geq 2.$$

For a positive integer $k$, the $k^{th}$ moment of ESD of the matrix $W_n$ is

$$\int x^k dF_n^W(x) = \frac{1}{n}\mathbb{E}(\text{Trace}(W_n^k)),$$

and the $k^{th}$ moment of the semicircle distribution is

$$\int_{-2}^2 x^k \rho_{sc}(x) dx.$$
On a compact set, convergence in distribution is the same as convergence of moments. To prove the theorem, we need to show, for every fixed number $k$,

$$\frac{1}{n} E(\text{trace}(W_n^k)) \to \int_{-2}^{2} x^k \rho_{sc}(x) dx, \text{ as } n \to \infty. \tag{4.8}$$

For $k = 2m + 1$, by symmetry,

$$\int_{-2}^{2} x^k \rho_{sc}(x) dx = 0.$$

For $k = 2m$,

$$\int_{-2}^{2} x^k \rho_{sc}(x) dx = \frac{1}{m+1} \binom{2m}{m}.$$

Thus our claim (4.8) follows by showing that

$$\frac{1}{n} E(\text{trace}(W_n^k)) = \begin{cases} O\left(\frac{1}{\sqrt{np}}\right) & \text{if } k = 2m + 1; \\ \frac{1}{m+1} \binom{2m}{m} + O\left(\frac{1}{np}\right) & \text{if } k = 2m. \end{cases} \tag{4.9}$$

We have the expansion for the trace of $W_n^k$,

$$\frac{1}{n} E(\text{trace}(W_n^k)) = \frac{1}{n^{1+k/2}} E(\text{trace}(M_n)^k)$$

$$= \frac{1}{n^{1+k/2}} \sum_{1 \leq i_1, \ldots, i_k \leq n} E\eta_{i_1} \eta_{i_2} \cdots \eta_{i_k} \tag{4.10}$$

Each term in the above sum corresponds to a closed walk of length $k$ on the complete graph $K_n$ on $\{1, 2, \ldots, n\}$. On the other hand, $\eta_{ij}$ are independent with mean 0. Thus the term is nonzero if and only if every edge in this closed walk appears at least twice.

And we call such a walk a *good* walk. Consider a *good* walk that uses $l$ different edges $e_1, \ldots, e_l$ with corresponding multiplicities $m_1, \ldots, m_l$, where $l \leq m$, each $m_h \geq 2$ and $m_1 + \ldots + m_l = k$. Now the corresponding term to this *good* walk has form

$$E\eta_{e_1}^{m_1} \cdots \eta_{e_l}^{m_l}.$$

Since such a walk uses at most $l + 1$ vertices, a naive upper bound for the number of *good* walks of this type is $n^{l+1}$. 
When \( k = 2m + 1 \), recall \( \mathbb{E}\eta^s = \Theta \left( (\sqrt{p})^{2-s} \right) \) for \( s \geq 2 \), and so

\[
\frac{1}{n} \mathbb{E}(\text{Trace}(W_n^k)) = \frac{1}{n^{1+k/2}} \sum_{l=1}^{m} \sum_{\text{good walk of } l \text{ edges}} \mathbb{E}\eta_{e_1}^{m_1} \cdots \eta_{e_l}^{m_l} \\
\leq \frac{1}{n^{m+3/2}} \sum_{l=1}^{m} n^{l+1} \left( \frac{1}{\sqrt{p}} \right)^{m_1-2} \cdots \left( \frac{1}{\sqrt{p}} \right)^{m_l-2} \\
= O\left( \frac{1}{\sqrt{np}} \right).
\]

When \( k = 2m \), we classify the \textit{good} walks into two types. The first kind uses \( l \leq m - 1 \) different edges. The contribution of these terms will be

\[
\frac{1}{n^{1+k/2}} \sum_{l=1}^{m-1} \sum_{\text{1st kind of good walk of } l \text{ edges}} \mathbb{E}\eta_{e_1}^{m_1} \cdots \eta_{e_l}^{m_l} \\
\leq \frac{1}{n^{1+m}} \sum_{l=1}^{m-1} n^{l+1} \left( \frac{1}{\sqrt{p}} \right)^{m_1-2} \cdots \left( \frac{1}{\sqrt{p}} \right)^{m_l-2} \\
= O\left( \frac{1}{np} \right).
\]

The second kind of \textit{good} walk uses exactly \( l = m \) different edges and thus \( m + 1 \) different vertices. And the corresponding term for each walk has form

\[ \mathbb{E}\eta_{e_1}^{2} \cdots \eta_{e_l}^{2} = 1. \]

The number of this kind of \textit{good} walk is given by the following result in the paper ([4], Page 617–618), which completes the proof of (4.8).

**Lemma 56.** The number of the second kind of \textit{good} walk is

\[ \frac{n^{m+1} (1 + O(n^{-1}))}{m+1} \left( \frac{2m}{m} \right). \]

Then the second conclusion of (4.8) follows.

On the other hand,

\[
\text{Var} \left( \frac{1}{n} \mathbb{E}(\text{trace}(W_n^k)) \right) = \mathbb{E} \left( \frac{1}{n} \text{trace}(W_n^k) \right)^2 - (\mathbb{E} \left( \frac{1}{n} \text{trace}(W_n^k) \right))^2 \\
= \frac{1}{n^{2+k}} \left( \mathbb{E} \left( \sum_{1 \leq i_1, \ldots, i_k \leq n} \eta_{i_1} \eta_{i_2} \cdots \eta_{i_k} \right)^2 - (\mathbb{E} \sum_{1 \leq i_1, \ldots, i_k \leq n} \eta_{i_1} \eta_{i_2} \cdots \eta_{i_k})^2 \right) \\
= \frac{1}{n^{2+k}} \sum_{i, j} [\mathbb{E}(X(i)X(j)) - \mathbb{E}(X(i))\mathbb{E}(X(j))],
\]
where
\[ X(i) := \sum_{1 \leq i_1, \ldots, i_k \leq n} \eta_{i_1 i_2} \eta_{i_2 i_3} \cdots \eta_{i_k i_1} \]
and
\[ X(j) := \sum_{1 \leq j_1, \ldots, j_k \leq n} \eta_{j_1 j_2} \eta_{j_2 j_3} \cdots \eta_{j_k j_1} \].

We still consider the term \( X(i) \) as a correspondence to a closed walk \( K(i) \) of length \( k \) on the complete graph \( K_n \) on \( \{1, 2, \ldots, n\} \). If any edge in the closed walk in \( K(i) \) or \( K(j) \) appears just once, then \( \mathbf{E}(X(i)X(j) - \mathbf{E}(X(i))\mathbf{E}(X(j))) = 0 \) because the \( \mathbf{E}(\mathbf{1}_{i j}) = 0 \). If \( K(i) \) and \( K(j) \) do not share edges, then \( \mathbf{E}(X(i)X(j) = \mathbf{E}(X(i))\mathbf{E}(X(j)) \).

Thus the nonzero contribution in (4.11) comes from the closed walk on \( \{i_1, i_2, \ldots, i_k\} \cup\{j_1, j_2, \ldots, j_k\} \) that each edge appears at least twice and there is at one edge that appears four times. It is clear that such a closed walk can use at most \( k \) vertices. We can use a trial upper bound \( n^k \) for the number of such closed walks. Therefore,

\[ \text{Var}\left( \frac{1}{n} \mathbf{E}(\text{trace}(W_n^k)) \right) \ll n^{-2}. \]

For each \( k \), by Chebyshev’s inequality, for any constant \( \epsilon > 0 \),

\[ \mathbf{P}(\left| \frac{1}{n} \text{trace}(W_n^k) - \frac{1}{n} \mathbf{E}(\text{trace}(W_n^k)) \right| \geq \epsilon) = O_\epsilon(n^{-2}), \]

which implies that

\[ \sum_n \mathbf{P}(\left| \frac{1}{n} \text{trace}(W_n^k) - \frac{1}{n} \mathbf{E}(\text{trace}(W_n^k)) \right| \geq \epsilon) < \infty. \]

By Borel–Cantelli lemma, we have

\[ \frac{1}{n} \text{trace}(W_n^k) \to \frac{1}{n} \mathbf{E}(\text{trace}(W_n^k)) \]

in probability. And together with (4.8), we have shown for each \( k \), as \( n \) tends to infinity,

\[ \frac{1}{n} \text{trace}(W_n^k) \to \int_{-2}^{2} x^k \rho_{sc}(x) dx \]

in probability. This completes the proof of Theorem 42.
References


[23] F. J. Dyson. Statistical theory of the energy levels of complex systems. i. Journal of Mathematical Physics, 3(140), 1962.


[30] L. Erdős, B. Schlein, and H.T. Yau. Semicircle law on short scales and delocaliz-
ization of eigenvectors for Wigner random matrices. The Annals of Probability,

[31] L. Erdős, B. Schlein, and H.T. Yau. Wegner estimate and level repulsion for Wigner

[32] L. Erdős, 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.

[33] U. Feige and E. Ofek. Spectral techniques applied to sparse random graphs. Ran-

[34] J. Friedman. On the second eigenvalue and random walks in random d-regular


the thirty-fifth annual ACM symposium on Theory of computing, pages 720–724.

[37] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. Combi-

[38] F. Götze and A. Tikhomirov. On the rate of convergence to the semi-circular law.


[40] D.L. Hanson and F.T. Wright. A bound on tail probabilities for quadratic forms in
1083, 1971.

[41] K. Johansson. Universality of the local spacing distribution in certain ensembles of
hermitian wigner matrices. Communications in Mathematical Physics, 215(3):683–

[42] I. M. Johnstone. On the distribution of the largest eigenvalue in principal compo-


[45] V.A. Marčenko and L.A. Pastur. Distribution of eigenvalues for some sets of


Vita

Ke Wang

Education

9/2006-5/2013  Ph. D. in Mathematics, Rutgers University
9/2002-5/2006  B. Sc. in Mathematics from University of Science and Technology of China (USTC)

Employment

9/2006-5/2013  Teaching assistant, Department of Mathematics, Rutgers University
5/2010-8/2010  Intern, Bell Laboratory

Publications


4. Optimal local semi-circle law and delocalization (With V. Vu), preprint.

5. Computing singular vectors with noise (With S. O'Rourke and V. Vu), preprint.