Tracy-Widom law for the extreme eigenvalues of sample correlation matrices*

Zhigang Bao[†]   Guangming Pan[‡]   Wang Zhou[§]

Abstract

Let the sample correlation matrix be $W = YY^T$, where $Y = (y_{ij})_{p,n}$ with $y_{ij} = x_{ij}/\sqrt{\sum_{j=1}^n x_{ij}^2}$. We assume $\{x_{ij} : 1 \leq i \leq p, 1 \leq j \leq n\}$ to be a collection of independent symmetrically distributed random variables with sub-exponential tails. Moreover, for any i, we assume $x_{ij}, 1 \leq j \leq n$ to be identically distributed. We assume $0 < p < n$ and $p/n \to y$ with some $y \in (0,1)$ as $p, n \to \infty$. In this paper, we provide the Tracy-Widom law (TW_1) for both the largest and smallest eigenvalues of W. If x_{ij} are i.i.d. standard normal, we can derive the TW_1 for both the largest and smallest eigenvalues of the matrix $\mathcal{R} = RR^T$, where $R = (r_{ij})_{p,n}$ with $r_{ij} = (x_{ij} - \bar{x}_i)/\sqrt{\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2}$, $\bar{x}_i = n^{-1} \sum_{j=1}^n x_{ij}$.

Keywords: extreme eigenvalues; sample correlation matrices; sample covariance matrices; Stieltjes transform; Tracy-Widom law.
AMS MSC 2010: 15B52; 62H25; 62H10.
Submitted to EJP on April 20, 2012, final version accepted on October 3, 2012.

1 Introduction

Suppose we have a p-dimensional distribution with mean μ and covariance matrix Σ. In recent three or four decades, in many research areas, including signal processing, network security, image processing, genetics, stock marketing and other economic problems, people are interested in the case where p is quite large or proportional to the sample size. Naturally, one may ask how to test the independence among the p components of the population. From the principal component analysis point of view, the independence test statistic is usually the maximum eigenvalue of the sample covariance matrices. Under the additional normality assumption, Johnstone [12] derived the asymptotic distribution of the largest eigenvalue of the sample covariance matrices to study the test $H_0 : \Sigma = I$ assuming $\mu = 0$.

†Zhejiang University, P. R. China. E-mail: zhigangbao@zju.edu.cn
‡Nanyang Technological University, Singapore. E-mail: GMPAN@ntu.edu.sg
§National University of Singapore, Singapore. E-mail: stazw@nus.edu.sg

However, sample covariance matrices are not scale-invariant. So if $\mu = 0$, Johnstone [12] proposes to perform principal component analysis (PCA) by the maximum eigenvalue of the matrix $W = YY^T$, where

$$Y = (y_{ij})_{p,n} := \begin{pmatrix} \frac{x_{11}}{||\mathbf{x}_1||} & \frac{x_{12}}{||\mathbf{x}_1||} & \cdots & \frac{x_{1n}}{||\mathbf{x}_1||} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{x_{p1}}{||\mathbf{x}_p||} & \frac{x_{p2}}{||\mathbf{x}_p||} & \cdots & \frac{x_{pn}}{||\mathbf{x}_p||} \end{pmatrix}. \tag{1.1}$$

Here $\mathbf{x}_i = (x_{i1}, \cdots, x_{in})^T$ contains n observations for the i-th component of the population, $i = 1, \cdots, p$, and $|| \cdot ||$ represents the vector norm. Unfortunately, one can not derive the limiting distribution of the largest eigenvalue of W by Johnstone's machinery for one decade.

Performing PCA on W amounts to PCA on the sample correlations of the original data if $\mu = 0$. So for simplicity, we call W the *sample correlation matrix* in this paper. From now on, the eigenvalues of W will be denoted by

$$0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p.$$

Then the empirical distribution (ESD) of W is defined by

$$F_p(x) = \frac{1}{p} \sum_{i=1}^{p} \mathbf{1}_{\{\lambda_i \leq x\}}.$$

The asymptotic property of F_p was studied in [11] and [2]. For the almost sure convergence of λ_1 and λ_p, see [11].

In this paper, we will study the fluctuations of the extreme eigenvalues λ_1, λ_p of W for a general population, including multivariate normal one. The basic assumption on the distribution of our population throughout the paper is

Condition C$_1$. We assume $x_{ij}, 1 \leq i \leq p, 1 \leq j \leq n$ to be independent symmetrically distributed random variables with common variance 1. And for any i, we assume $x_{i1}, \cdots, x_{in}$ to be i.i.d. Furthermore, we request the distributions of the x'_{ij}s have sub-exponential tails, i.e., there exist positive constants C, C' such that for all $1 \leq i \leq p, 1 \leq j \leq n$ one has

$$\mathbb{P}(|x_{ij}| \geq t^C) \leq e^{-t}$$

for all $t \geq C'$. And we also assume $p/n \to y$ as $p, n = n(p) \to \infty$, where $0 < y < 1$.

Remark 1.1. *The sample correlation matrix W is invariant under the scaling on the elements x_{ij}, so the assumption $Var(x_{ij}) = 1$ is not necessary indeed. We specify it to be 1 here just for convenience.*

Remark 1.2. *Owing to the exponential tails, we can always truncate the variables so that $|x_{ij}| \leq K$ with some $K \geq \log^{O(1)} n$ for all $1 \leq i \leq p, 1 \leq j \leq n$. It is easy to see that the truncated matrix equals to the original one with probability $1 - o(1)$. Note that x_{ij} are symmetric and $x_{i1}, \cdots, x_{in}$ are i.i.d for any $1 \leq i \leq p$, thus the truncated variables $x_{i1}\mathbf{1}_{\{|x_{i1}| \leq K\}}, \cdots, x_{in}\mathbf{1}_{\{|x_{in}| \leq K\}}$ are still symmetric and i.i.d for any $1 \leq i \leq p$. Moreover, as mentioned above, the sample correlation matrix is invariant under scaling, so we can adjust the variances of the truncated variables to be 1 by scaling. Therefore, without loss of generality. We can always work with the condition $\mathbf{C_1}$ and the additional assumption that $|x_{ij}| \leq K$ with some $K \geq \log^{O(1)} n$ for all $1 \leq i \leq p, 1 \leq j \leq n$.*

A special sample correlation matrix model is the Bernoulli case, i.e. x_{ij} takes value of 1 or -1 with equal probability. Notice that if x_{ij} are Bernoulli, we always have for all $1 \leq i \leq p$

$$||\mathbf{x}_i||^2 = x_{i1}^2 + \cdots + x_{in}^2 = n.$$

As a consequence, the sample correlation matrix with Bernoulli elements coincides with its corresponding sample covariance matrix for which the limiting distribution of the extreme eigenvalues are well known under some moment assumptions. One can refer to [4],[10], [13], [15] and [20]. We only summarize their results for the special Bernoulli case as the following theorem.

Theorem 1.3 (Bernoulli case). *For the matrix W in (1.1), if x_{ij} are ± 1 Bernoulli variables, we have*

$$\frac{n\lambda_p - (p^{1/2} + n^{1/2})^2}{(n^{1/2} + p^{1/2})(p^{-1/2} + n^{-1/2})^{1/3}} \xrightarrow{d} TW_1,$$

and

$$\frac{n\lambda_1 - (p^{1/2} - n^{1/2})^2}{(n^{1/2} - p^{1/2})(p^{-1/2} - n^{-1/2})^{1/3}} \xrightarrow{d} TW_1.$$

as $p, n \to \infty$ with $p/n \to y \in (0, 1)$.

Here TW_1 is the famous Tracy-Widom distribution of type 1, which was firstly raised by Tracy and Widom in [19] for the Gaussian orthogonal ensemble. The distribution function $F_1(t)$ of TW_1 admits the representation

$$F_1(t) = \exp(-\frac{1}{2} \int_t^\infty [q(x) + (x - t)q(x)^2]dx),$$

where q statisfies the Painlev$\acute{e}$ *II* equation

$$q'' = tq + 2q^3, \quad q(t) \sim \mathrm{Ai(t)}, \text{as t} \to \infty.$$

Here $\mathrm{Ai(t)}$ is the Airy function.

The main purpose of this paper is to generalize Theorem 1.3 to the population satisfying the basic condition $\mathbf{C}_1$. Our main results are the following two theorems.

Theorem 1.4. *Let W be a sample correlation matrix satisfying the basic condition $\mathbf{C}_1$. We have*

$$\frac{n\lambda_p - (p^{1/2} + n^{1/2})^2}{(n^{1/2} + p^{1/2})(p^{-1/2} + n^{-1/2})^{1/3}} \xrightarrow{d} TW_1.$$

and

$$\frac{n\lambda_1 - (p^{1/2} - n^{1/2})^2}{(n^{1/2} - p^{1/2})(p^{-1/2} - n^{-1/2})^{1/3}} \xrightarrow{d} TW_1.$$

as $p \to \infty$.

Remark 1.5. *For technical reasons, it is convenient to work with the continuous random variables x_{ij}. As a result, the events such as eigenvalue collision will only occur with probability zero (see Lemma 3.10). Because none of our bounds depends on how continuous the x_{ij} are, one can recover the discrete case from the continuous one by a standard limiting argument by using Weyl's inequality (see Lemma 2.2), especially for the Bernoulli case.*

If the population is normal, then we can derive the Tracy-Widom law for both the largest and smallest eigenvalues of the matrix $\mathcal{R} = RR^T$, where

$$R = (r_{ij})_{p,n} := \begin{pmatrix} \frac{x_{11}-\bar{x}_1}{||\mathbf{x}_1-\bar{x}_1||} & \frac{x_{12}-\bar{x}_1}{||\mathbf{x}_1-\bar{x}_1||} & \cdots & \frac{x_{1n}-\bar{x}_1}{||\mathbf{x}_1-\bar{x}_1||} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{x_{p1}-\bar{x}_p}{||\mathbf{x}_p-\bar{x}_p||} & \frac{x_{p2}-\bar{x}_p}{||\mathbf{x}_p-\bar{x}_p||} & \cdots & \frac{x_{pn}-\bar{x}_p}{||\mathbf{x}_p-\bar{x}_p||} \end{pmatrix}. \tag{1.2}$$

Here $\bar{x}_i = n^{-1}\sum_{j=1}^n x_{ij}$ and $\mathbf{x}_i - \bar{x}_i$ means each element x_{ij} of $\mathbf{x}_i$ will be subtracted by $\bar{x}_i$, $i = 1, \cdots, p$. We denote the ordered eigenvalues of $\mathcal{R}$ by $0 \leq \lambda_1(\mathcal{R}) \leq \cdots \leq \lambda_p(\mathcal{R})$ below. Actually $\mathcal{R}$ is the sample correlation matrix when the population mean is unknown.

Theorem 1.6. *For the sample correlation matrix $\mathcal{R}$ with i.i.d $N(0,1)$ elements, if $p/n \to y \in (0,1)$, we have*

$$\frac{n\lambda_p(\mathcal{R}) - (p^{1/2} + n^{1/2})^2}{(n^{1/2} + p^{1/2})(p^{-1/2} + n^{-1/2})^{1/3}} \xrightarrow{d} TW_1.$$

and

$$\frac{n\lambda_1(\mathcal{R}) - (p^{1/2} - n^{1/2})^2}{(n^{1/2} - p^{1/2})(p^{-1/2} - n^{-1/2})^{1/3}} \xrightarrow{d} TW_1$$

as $p \to \infty$.

Throughout the paper, We will use C, C_0, C_1, C_2, C' to denote some positive constants independent of p, which may differ from line to line. And we will use C_α to denote some positive constants depending on the parameter α. The notation $|| \cdot ||_{op}, || \cdot ||_F$ represent the operator norm and Frobenius norm of a matrix respectively. And $|| \cdot ||$ represents a Euclidean norm of a vector.

And we will use the following ad hoc definitions on the frequent events provided in [16].

Definition 1.7 (Frequent events, [16]). *Let E be an event depending on n.*

- *E holds* asymptotically almost surely *if $\mathbb{P}(E) = 1 - o(1)$.*
- *E holds with* high probability *if $\mathbb{P}(E) \geq 1 - O(n^{-c})$ for some constant $c > 0$ (independent of n).*
- *E holds with* overwhelming probability *if $\mathbb{P}(E) \geq O_C(n^{-C})$ for every constant $C > 0$.*
- *E holds* almost surely *if $\mathbb{P}(E) = 1$.*

The main strategy is to prove a so-called "Green function comparison theorem", which was raised by Erdös, Yau and Yin in [9] for generalized Wigner matrices. Very recently, Pillai and Yin provided an analogous comparison theorem for sample covariance matrices in [14]. Moreover, a rigidity result was derived for a class of matrices including sample correlation matrices, which is crucial for our proof (see Theorem 1.5 of [14]). Different from the Wigner matrix case, the authors of [14] used a strategy to replace the matrix entries column by column instead of replacing entries one at a time in the swapping procedure. We will pursue this strategy to sample correlation matrices to provide a "Green function comparison theorem" to the sample correlation matrices obeying the assumption $\mathbf{C}_1$ in Section 4, see Theorem 4.3. Then by the comparison theorem, we can compare the general distributed case with the Bernoulli case to get Theorem 1.4. And as an application, we can also get Theorem 1.6.

Our article is organized as follows. In Section 2, we state some basic tools, which can be also found in the series work [16], [17], [18] and [20]. And we provide some main technical lemmas and theorems in Section 3. The most important one is the so-called delocalization property of singular vectors, which will be shown as an obstacle to establish the Green function comparison theorem in the sample correlation matrices case. And in Section 4, we provide a Green function comparison theorem (Theorem 4.3). The proof is a combination of Pillai and Yin's arguments in [14] and the delocalization property of singular vectors proved in Section 3. In Section 5, we state the proofs for our main results: Theorem 1.4 and Theorem 1.6.

2   Basic Tools

In this section, we state some basic tools from linear algebra and probability theory. Firstly, we denote the ordered singular values of Y by

$$0 \le \sigma_1 \le \sigma_2 \le \cdots \le \sigma_p,$$

then we have $\sigma_i = \lambda_i^{1/2}$. If we further denote the unit right singular vector of Y corresponding σ_i by u_i and the left one by v_i, we have

$$Yu_i = \sigma_i v_i \qquad (2.1)$$

and

$$Y^T v_i = \sigma_i u_i. \qquad (2.2)$$

Below we shall state some tools for eigenvalues, singular values and singular vectors without proof.

Lemma 2.1 (Cauchy's interlacing law). *. Let $1 \le p \le n$*

(i) *If A_n is an $n \times n$ Hermitian matrix, and A_{n-1} is an $n-1 \times n-1$ minor, then $\lambda_i(A_n) \le \lambda_i(A_{n-1}) \le \lambda_{i+1}(A_n)$ for all $1 \le i < n$.*
(ii) *If $A_{n,p}$ is a $p \times n$ matrix, and $A_{n,p-1}$ is a $p-1 \times n$ minor, then $\sigma_i(A_{n,p}) \le \sigma_i(A_{n,p-1}) \le \sigma_{i+1}(A_{n,p})$ for all $1 \le i < p$.*
(iii) *If $p < n$, $A_{n,p}$ is a $p \times n$ matrix, and $A_{n-1,p}$ is a $p \times n-1$ minor, then $\sigma_{i-1}(A_{n,p}) \le \sigma_i(A_{n-1,p}) \le \sigma_i(A_{n,p})$ for all $1 \le i \le p$, with the understanding that $\sigma_0(A_{n,p}) = 0$. (For $p = n$, one can consider its transpose and use* (ii) *instead.)*

Lemma 2.2 (Weyl's inequality). *Let $1 \le p \le n$*

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

The following lemma is on the components of a singular vector, which can be found in [18].

Lemma 2.3. *[18] Let $p, n \ge 1$, and let*

$$A_{p,n} = (A_{p,n-1} \quad h)$$

be a $p \times n$ matrix with $h \in \mathbb{C}^p$, and let $\binom{u}{x}$ be a right unit singular vector of $A_{p,n}$ with singular value $\sigma_i(A_{p,n})$, where $x \in \mathbb{C}$ and $u \in \mathbb{C}^{n-1}$. Suppose that none of the singular values of $A_{p,n-1}$ is equal to $\sigma_i(A_{p,n})$. Then

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

where $\{v_1(A_{p,n-1}), \cdots, v_{\min(p,n-1)}(A_{p,n-1}) \in \mathbb{C}^p\}$ is an orthonormal system of left singular vectors corresponding to the non-trivial singular values of $A_{p,n-1}$ and $v_j(A_{p,n-1}) \cdot h = v_j(A_{p,n-1})^* h$ with $v_j(A_{p,n-1})^*$ being the complex conjugate of $v_j(A_{p,n-1})$.

Similarly, if

$$A_{p,n} = \begin{pmatrix} A_{p-1,n} \\ l^* \end{pmatrix}$$

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

$$|\mathfrak{y}|^2 = \frac{1}{1 + \sum_{j=1}^{\min(p-1,n)} \frac{\sigma_j(A_{p-1,n})^2}{(\sigma_j(A_{p-1,n})^2 - \sigma_i(A_{p,n})^2)^2} |u_j(A_{p-1,n}) \cdot l|^2},$$

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

Further, we need a frequently used tool in the Random Matrix Theory: the Stieltjes transform of ESD $F_p(x)$, which is defined by

$$s_p(z) = \int \frac{1}{x - z} dF_p(x)$$

for any $z = E + i\eta$ with $E \in \mathbb{R}$ and $\eta > 0$. If we introduce the Green function $G(z) = (W - z)^{-1}$, we also have

$$s_p(z) = \frac{1}{p} Tr G(z) = \frac{1}{p} \sum_{k=1}^{p} G_{kk}. \tag{2.3}$$

Here we denote G_{jk} as the (j,k) entry of $G(z)$. As is well known, the convergence of a tight probability measure sequence is equivalent to the convergence of its Stieltjes transform sequence towards the corresponding transform of the limiting measure. So corresponding to the convergence of $F_p(x)$ towards $F_{MP,y}(x)$, the famous Marčenko-Pastur law $F_{MP,y}(x)$ whose density function is given by

$$\rho_{MP,y} = \frac{1}{2\pi xy} \sqrt{(b-x)(x-a)} \mathbf{1}_{[a,b]}(x), \tag{2.4}$$

where $a = (1 - \sqrt{y})^2, b = (1 + \sqrt{y})^2$, $s_p(z)$ almost surely converges to the Stieltjes transform $s(z)$ of $F_{MP,y}(x)$. Here

$$s(z) = \frac{1 - y - z + \sqrt{(z - 1 - y)^2 - 4y}}{2yz}, \tag{2.5}$$

where the square root is defined as the analytic extension of the positive square root of the positive numbers. Moreover, $s(z)$ satisfies the equation

$$s(z) + \frac{1}{y + z - 1 + yzs(z)} = 0. \tag{2.6}$$

If we denote the k-th row of Y by $\mathbf{y}_k^T$ and the remaining $(p-1) \times n$ matrix after deleting $\mathbf{y}_k^T$ by $Y^{(k)}$, one has

$$W = \begin{pmatrix} 1 & \mathbf{y}_1^T Y^{(1)T} \\ Y^{(1)} \mathbf{y}_1 & Y^{(1)} Y^{(1)T} \end{pmatrix}.$$

By Schur's complement,

$$
\begin{aligned}
G_{11} &= \frac{1}{1 - z - \mathbf{y}_1^T Y^{(1)T} (Y^{(1)} Y^{(1)T} - z)^{-1} Y^{(1)} \mathbf{y}_1} \\
&= \frac{1}{1 - z - \mathbf{y}_1^T Y^{(1)T} Y^{(1)} (Y^{(1)T} Y^{(1)} - z)^{-1} \mathbf{y}_1}.
\end{aligned}
\tag{2.7}
$$

The formula of G_{kk} is analogous. By (2.3), we have the following lemma on the decomposition of $s_p(z)$:

Lemma 2.4. *For the matrix W, we have*

$$s_p(z) = \frac{1}{p} \sum_{k=1}^{p} \frac{1}{1 - z - \mathbf{y}_k^T Y^{(k)T} Y^{(k)} (Y^{(k)T} Y^{(k)} - z)^{-1} \mathbf{y}_k}.$$

The last main tool we need comes from the probability theory, which is a concentration inequality for projections of random vectors. The details of the proof can also be found in [16].

Lemma 2.5. *Let $\mathcal{X} = (\xi_1, \cdots, \xi_n)^T \in \mathbb{C}^n$ be a random vector whose entries are independent with mean zero, variance 1, and are bounded in magnitude by K almost surely for some K, where $K \geq 10(\mathbb{E}|\xi|^4 + 1)$. Let H be a subspace of dimension d and π_H the orthogonal projection onto H. Then*

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

In particular, one has

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

with overwhelming probability.

3 Main Technical Results

In this section, we provide our main technical results: the local MP law for sample correlation matrices, and the delocalization property for the singular vectors. Both results will be proved under much weaker assumption than $\mathbf{C}_1$. We form them into the following two theorems.

Let us introduce more notation. For any interval $I \subset \mathbb{R}$, we use N_I to denote the number of the eigenvalues of W falling into I, and use $|I|$ to denote the length of I.

Theorem 3.1 (Local MP law). *. Assume that $p/n \to y$ with $0 < y < 1$. And $\{x_{ij} : 1 \leq i \leq p, 1 \leq j \leq n\}$ is a collection of independent (but not necessary identically distributed) random variables with mean zero and variance 1. If $|x_{ij}| \leq K$ almost surely for some $K = o(p^{1/C_0} \delta^2 \log^{-1} p)$ with some $0 < \delta < 1/2$ and some large constant C_0 for all i, j, one has with overwhelming probability that the number of eigenvalues N_I for any interval $I \subset [a/2, 2b]$ with $|I| \geq \frac{K^2 \log^7 p}{\delta^9 p}$ obeys*

$$\left| N_I - p \int_I \rho_{MP,y}(x) dx \right| \leq \delta p |I| \tag{3.1}$$

when p is sufficiently large.

Remark 3.2. *The topic of the limiting spectrum distribution on short scales was firstly raised by Erdős, Schlein and Yau in [6] for Wigner matrices. Such type of results are shown to be quite necessary for the proof of the famous universality conjectures in the Random Matrix Theory, for example, see [8] and [16].*

Remark 3.3. *The local MP law for sample covariance matrices was proved in [8], [18] and [20]. And a strong type of the local MP law has been established for more general matrix models in a very recent paper of Pillai and Yin, see Theorem 1.5, [14]. In fact, from Theorem 1.5 of [14], one can get a more precise bound than that in (3.1) if we replace $\rho_{MP,y}(x)$ by the nonasymptotic MP law $\rho_W(x)$ defined in Section 4. Moreover, Pillai and Yin's strong local MP law also provides some crucial estimates on individual elements of the Green function G, which will be used to establish our Green function comparison theorem in Section 4.*

Theorem 3.4 (Delocalization of singular vectors). *Under the assumptions of Theorem 3.1 and $\mathbb{E}x_{ij}^3 = 0$, if we assume $x_{ij}'s$ are continuous random variables, then with overwhelming probability all the left and right unit singular vectors of W have all components uniformly of size at most $p^{-1/2}K^{C_0/2}\log^{O(1)}p$.*

Remark 3.5. *Note that a little weaker delocalization property for the left singular vector v_i can also be found in Theorem 1.2 (iv) of Pillai and Yin [14].*

Now if we denote

$$X = \begin{pmatrix} \frac{x_{11}}{\sqrt{n}} & \frac{x_{12}}{\sqrt{n}} & \cdots & \frac{x_{1n}}{\sqrt{n}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{x_{p1}}{\sqrt{n}} & \frac{x_{p2}}{\sqrt{n}} & \cdots & \frac{x_{pn}}{\sqrt{n}} \end{pmatrix},$$

then $S := XX^T$ is the sample covariance matrix corresponding to W. We further denote the ordered eigenvalues of S by $0 \le \tilde{\lambda}_1 \le \cdots \le \tilde{\lambda}_p$ and introduce the matrix

$$D = \begin{pmatrix} \frac{\sqrt{n}}{||\mathbf{x}_1||} & & \\ & \ddots & \\ & & \frac{n}{||\mathbf{x}_p||} \end{pmatrix}.$$

By Theorem 5.9 of [1], we have $\tilde{\lambda}_p = b + o(1)$ holds with overwhelming probability. In fact, it is easy to see $\tilde{\lambda}_1 = a + o(1)$ holds with overwhelming probability as well by a similar discussion through moment method. Observe that $W = DSD$, and $||D - I||_{op} = o(1)$ holds with overwhelming probability. By Lemma 2.2, we also have

$$\lambda_1 = a + o(1), \quad \lambda_p = b + o(1) \tag{3.2}$$

holds with overwhelming probability. So below we always assume $\lambda_i \in (a/2, 2b), 1 \le i \le p$.

The proof of Theorem 3.1 is partially based on the lemmas of Section 2. It turns out to be quite similar to the case of sample covariance matrices and Wigner matrices, see [7], [8], [18] and [20]. However, the delocalization of the right singular vector u_i of Y is an obstacle, owing to the lack of independence between the columns of Y.

For the convenience of the reader, we provide a short proof of Theorem 3.1 at first. Our main task in this section is the proof of Theorem 3.4, more precisely, the right singular vector part of the theorem.

Proof of Theorem 3.1. To show the local MP law, we will pursue the approach provided in [8]. Though the argument in [8] was presented for the sample covariance matrices, with slight modification it can also be used to the sample correlation matrices. We need the following crude upper bound on N_I at first.

Lemma 3.6. *Under the assumptions of Theorem 3.1, we have for any interval $I \subset \mathbb{R}$ with $|I| \gg K^2 \log^2 p/p$, and large enough $C > 0$*

$$N_I \leq Cp|I|$$

with overwhelming probability.

Proof. Firstly we introduce the notation

$$W^{(k)} = Y^{(k)}Y^{(k)T}, \quad \mathcal{W}^{(k)} = Y^{(k)^T}Y^{(k)}, \quad G^{(k)} = (W^{(k)} - z)^{-1}, \quad \mathcal{G}^{(k)} = (\mathcal{W}^{(k)} - z)^{-1}.$$

Let $\lambda_\alpha^{(1)}, \alpha = 1, \cdots, p - 1$ denote the eigenvalues of the $(p - 1) \times (p - 1)$ matrix $W^{(1)}$. Thus $\lambda_\alpha^{(1)}, \alpha = 1, \cdots, p - 1$ are also the eigenvalues of the $n \times n$ matrix $\mathcal{W}^{(1)}$, whose other eigenvalues are all zeros. We further use ν_α to denote the eigenvector of $\mathcal{W}^{(1)}$ corresponding to the eigenvalue $\lambda_\alpha^{(1)}$, and introduce the quantity

$$\xi_\alpha = n|\mathbf{y}_1 \cdot \nu_\alpha|^2 = \frac{n}{||\mathbf{x}_1||^2}|\mathbf{x}_1 \cdot \nu_\alpha|^2 =: \frac{n}{||\mathbf{x}_1||^2}\tilde{\xi}_\alpha. \tag{3.3}$$

We can rewrite (2.7) as

$$G_{11} = \frac{1}{1 - z - \frac{1}{n}\sum_{\alpha=1}^{p-1}\frac{\lambda_\alpha^{(1)}\xi_\alpha}{\lambda_\alpha^{(1)} - z}}. \tag{3.4}$$

By Cauchy's interlacing law, we also have $\lambda_\alpha^{(1)} \in [a/2, 2b]$ with overwhelming probability. Then for any $z = E + i\eta$ such that $E \in [a/2, 2b]$, we have

$$|\Im G_{kk}| \leq \frac{1}{\eta + \frac{\eta}{n}\sum_{\alpha=1}^{p-1}\frac{\lambda_\alpha^{(1)}\xi_\alpha}{(\lambda_\alpha^{(1)} - E)^2 + \eta^2}} \leq \frac{C_1 p\eta}{\sum_{\alpha:|\lambda_\alpha^{(1)} - E| \leq \eta}\xi_\alpha} \tag{3.5}$$

for any $k \in \{1, \cdots, p\}$. Now we set $I = [E - \eta/2, E + \eta/2]$. Notice that there always exists some positive constant C_2 such that

$$N_I \leq C_2 p\eta \Im s_p(z) = C_2 \eta \sum_{k=1}^{p}\Im G_{kk}. \tag{3.6}$$

If we set $C_3 = C_1 C_2$, it follows from (3.5) and (3.6) that

$$\mathbb{P}(N_I \geq Cp\eta)$$

$$= \mathbb{P}\left(\sum_{k=1}^{p}\Im G_{kk} \geq C_2^{-1}Cp \text{ and } N_I \geq Cp\eta\right)$$

$$\leq p\mathbb{P}\left(\sum_{\alpha:|\lambda_\alpha^{(1)} - E| \leq \eta/2}\xi_\alpha \leq C_3 C^{-1}p\eta \text{ and } N_I \geq Cp\eta\right)$$

$$\leq p\mathbb{P}\left(\frac{n}{||\mathbf{x}_1||^2}\sum_{\alpha:|\lambda_\alpha^{(1)} - E| \leq \eta/2}\tilde{\xi}_\alpha \leq C_3 C^{-1}p\eta \text{ and } N_I \geq Cp\eta\right)$$

$$\leq p\mathbb{P}(||\mathbf{x}_1||^2 \geq 2n) + p\mathbb{P}\left(\sum_{\alpha:|\lambda_\alpha^{(1)} - E| \leq \eta/2}\tilde{\xi}_\alpha \leq 2C_3 C^{-1}p\eta \text{ and } N_I \geq Cp\eta\right). \tag{3.7}$$

The first term of (3.7) is obviously exponential small by the Hoeffding inequality. For the second term, we use Lemma 2.5. Now we specialize $\mathcal{X}$ in Lemma 2.5 to be $\mathbf{x}_1$ and the subspace H to be the one generated by eigenvectors $\{\nu_\alpha : \lambda_\alpha^{(1)} \in I\}$. Thus one has

$$d = N_I \geq Cp\eta \gg CK^2 \log^2 n.$$

Then by Lemma 2.5 we have

$$\sum_{\alpha:|\lambda_\alpha^{(1)}-E|\leq \eta/2} \tilde{\xi}_\alpha = ||\pi_H(\mathcal{X})||^2 > \frac{1}{2}Cp\eta$$

with overwhelming probability. This implies that the second term of (3.7) is exponential small when C is large enough. So we conclude the proof of Lemma 3.6. $\qquad\square$

Now we proceed to prove Theorem 3.1. We follow the strategy in [8] to compare $s_p(z)$ and $s(z)$ with small imaginary part η. In fact, we have the following proposition.

Proposition 3.7. *Let* $1/10 \geq \eta \geq \frac{1}{n}$, *and* $L_1, L_2, \epsilon, \delta > 0$. *Suppose that one has the bound*

$$|s_p(z) - s(z)| \leq \delta$$

with (uniformly) overwhelming probability for all $z = E + i\eta$ *such that* $E \in [L_1, L_2]$ *and* $\Im z \geq \eta$. *Then for any interval* $I \subset [L_1 + \epsilon, L_2 - \epsilon]$ *with* $|I| \geq \max(2\eta, \frac{\eta}{\delta}\log\frac{1}{\delta})$, *one has*

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

with overwhelming probability.

Remark 3.8. *Proposition 3.7 is an extension of Lemma 30 of [18] up to the edge, whose proof can be found in [20]. In fact, the proof can be taken in the same manner as that of Lemma 64 in [16] for the Wigner matrix.*

So in view of Proposition 3.7, to prove Theorem 3.1, we only need to prove that the bound

$$|s_p(z) - s(z)| \leq \delta \tag{3.8}$$

holds with (uniformly) overwhelming probability for all $z = E + i\eta$ such that $E \in [a/2 - \epsilon, 2b + \epsilon]$ and $1/10 \geq \eta \geq \frac{K^2 \log^6 n}{n\delta^8}$. To prove (3.8) we need to derive a consistent equation for $s_p(z)$, which is similar to the equation (2.6) for $s(z)$.

Firstly by Lemma 2.4 we can rewrite $s_p(z)$ as

$$s_p(z) = \frac{1}{p}\sum_{k=1}^p \frac{1}{1-z-d_k},$$

with

$$d_k = \mathbf{y}_k^T \mathcal{W}^{(k)} \mathcal{G}^{(k)} \mathbf{y}_k.$$

Then the proof of (3.8) can be taken in the same manner as the counterpart of the sample covariance matrix case (see the proof of formula (4.12) of [20]). We only state the different parts below and leave the details to the reader. We remark here that we consider the domain $[L_1, L_2] = [a/2 - \epsilon, 2b + \epsilon]$ rather than $[a, b]$ in [20]. However, if one goes through the proof in [20], it is not difficult to see that the proof towards any

domain $[L_1, L_2]$ containing $[a, b]$ is the same. The only minor difference between our case and the sample covariance matrix in [20] is the estimation of d_k. We will only deal with d_1 in the sequel. The others are analogous. By (3.3) and (3.4), we have

$$d_1 = \frac{1}{n} \sum_{\alpha=1}^{p-1} \frac{\lambda_\alpha^{(1)} \xi_\alpha}{\lambda_\alpha^{(1)} - z} = \frac{1}{n} \sum_{\alpha=1}^{p-1} \frac{\lambda_\alpha^{(1)}}{\lambda_\alpha^{(1)} - z} + \frac{1}{n} \sum_{\alpha=1}^{p-1} \frac{\lambda_\alpha^{(1)} (\xi_\alpha - 1)}{\lambda_\alpha^{(1)} - z}. \tag{3.9}$$

For the first term of (3.9) we have

$$\frac{1}{n} \sum_{\alpha=1}^{p-1} \frac{\lambda_\alpha^{(1)}}{\lambda_\alpha^{(1)} - z} = \frac{p-1}{n} + \frac{z}{n} \sum_{j=1}^{p-1} \frac{1}{\lambda_\alpha^{(1)} - z} := \frac{p-1}{n}(1 + z s_p^{(1)}(z)),$$

where

$$s_p^{(1)}(z) = \frac{1}{p-1} \sum_{j=1}^{p-1} \frac{1}{\lambda_\alpha^{(1)} - z}$$

is the Stieltjes transform of the ESD of $W^{(1)}$. Then by the Cauchy's interlacing property, we have

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

Consequently one has

$$\frac{1}{n} \sum_{\alpha=1}^{p-1} \frac{\lambda_\alpha^{(1)}}{\lambda_\alpha^{(1)} - z} = \frac{p-1}{n} + z \frac{p}{n} s_p(z) + o(\delta^2). \tag{3.10}$$

Now we provide the following lemma on the second term of (3.9).

Lemma 3.9. *For all $z = E + i\eta$ with $E \in [a/2 - \epsilon, 2b + \epsilon]$ and $\eta \geq \frac{K^2 \log^6 n}{n\delta^8}$,*

$$\frac{1}{n} \sum_{\alpha=1}^{p-1} \frac{\lambda_\alpha^{(1)} (\xi_\alpha - 1)}{\lambda_\alpha^{(1)} - z} = o(\delta^2)$$

uniformly in z with overwhelming probability.

Proof. We set $R_j = (\xi_j - 1)$. By (3.3) and the fact that

$$\frac{n}{||\mathbf{x}_1||^2} = 1 + O\left(\frac{K^2 \log^2 n}{\sqrt{n}}\right) \tag{3.11}$$

holds with overwhelming probability, we have for any $T \subset \{1, \cdots, p-1\}$

$$\sum_{j \in T} R_j = \frac{n}{||\mathbf{x}_1||^2} \sum_{j \in T} |\mathbf{x}_1 \cdot \nu_j|^2 - |T|. \tag{3.12}$$

By using Lemma 2.5, we have

$$\sum_{j \in T} |\mathbf{x}_1 \cdot \nu_j|^2 = T + O\left(\sqrt{T} K \log n \vee K^2 \log^2 n\right), \tag{3.13}$$

where $a \vee b = \max(a, b)$. By inserting (3.11) and (3.13) into (3.12), we have

$$\sum_{j \in T} R_j = \sum_{j \in T} |\mathbf{x}_1 \cdot \nu_j|^2 - |T| + O\left(\frac{T K^4 \log^4 n}{\sqrt{n}}\right).$$

If we choose $T = \log^{O(1)} n$, we always have

$$\sum_{j \in T} R_j = \sum_{j \in T} |\mathbf{x}_1 \cdot \nu_j|^2 - |T| + o(\delta^2).$$

Then the following part of the proof is the same as that in the sample covariance matrix case. One can refer to the proof of Proposition 4.6 of [20] for details. $\square$

Now we proceed to the proof of Theorem 3.1. By (3.9), (3.10) and Lemma 3.9 we can get the following equation

$$s_p(z) + \frac{1}{\frac{p}{n} + z - 1 + z\frac{p}{n}s_p(z) + o(\delta^2)} = 0. \tag{3.14}$$

Now we claim that there exists a sufficiently large constant $L > 0$ such that

$$|zs_p(z)| \le L \tag{3.15}$$

if $s_p(z)$ satisfies the relation (3.14). Note that $z = E + i\eta$, where $E \in [a/2 - \epsilon, 2b + \epsilon]$ and $1/10 \ge \eta \ge \frac{K^2 \log^6 n}{n\delta^8}$. Thus we have

$$\max\{|z|, |\frac{p}{n} + z - 1 + o(\delta^2)|\} \le C \tag{3.16}$$

for some constant $C > 0$. From (3.14), we have

$$|s_p(z)||\frac{p}{n}zs_p(z) + \frac{p}{n} + z - 1 + o(\delta^2)| = 1.$$

Thus by (3.16) one has

$$|zs_p(z)|(|\frac{p}{n}zs_p(z) + \frac{p}{n} + z - 1 + o(\delta^2)|) \le C,$$

which implies

$$|zs_p(z)||\frac{p}{n}|zs_p(z)| - |\frac{p}{n} + z - 1 + o(\delta^2)|| \le C.$$

By assumption, for p large enough we have $p/n > c$ for some positive constant c, thus it is elementary to see that there exists some positive constant L such that $|zs_p(z)| \le L$ by using (3.16) again. With the aid of (3.15) and the fact that $|p/n - y| = o(1)$, we have

$$s_p(z) + \frac{1}{y + z - 1 + yzs_p(z) + o(\delta^2)} = 0 \tag{3.17}$$

when p is sufficiently large such that $|p/n - y| = o(\delta^2)$. Then by a standard comparison of (3.17) and (2.6) (see [20] for example), we have (3.8). Thus by Proposition 3.7 we conclude the proof of Theorem 3.1. $\square$

Now we turn to the proof of Theorem 3.4. At first, we introduce the matrix $\widehat{W}_{(n)} := \widehat{Y}_{(n)}\widehat{Y}_{(n)}^T$ with

$$\widehat{Y}_{(n)} = \begin{pmatrix} \frac{x_{11}}{||\widehat{\mathbf{x}}_1||} & \frac{x_{12}}{||\widehat{\mathbf{x}}_1||} & \cdots & \frac{x_{1,n-1}}{||\widehat{\mathbf{x}}_1||} \\ \frac{x_{21}}{||\widehat{\mathbf{x}}_2||} & \frac{x_{22}}{||\widehat{\mathbf{x}}_2||} & \cdots & \frac{x_{2,n-1}}{||\widehat{\mathbf{x}}_2||} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{x_{p1}}{||\widehat{\mathbf{x}}_p||} & \frac{x_{p2}}{||\widehat{\mathbf{x}}_p||} & \cdots & \frac{x_{p,n-1}}{||\widehat{\mathbf{x}}_p||} \end{pmatrix},$$

where

$$\widehat{\mathbf{x}}_j = (x_{j1}, x_{j2}, \cdots, x_{j,n-1})^T.$$

We will need the following lemma on eigenvalue collision.

Lemma 3.10. *If we assume the random variables $x'_{ij}s$ are continuous, we have the following events hold with probability one.*
i): W has simple eigenvalues, i.e. $\lambda_1 < \lambda_2 < \cdots < \lambda_p$.
ii): W and $W^{(p)}$ have no eigenvalue in common.
iii): W and $\widehat{W}_{(n)}$ have no eigenvalue in common.

The proof of Lemma 3.10 will be postponed to Appendix A.

Proof of Theorem 3.4. The proof for the left singular vectors is nearly the same as the sample covariance matrix case shown in [20] by using Lemma 2.3, *ii*) of Lemma 3.10 and Theorem 3.1. Moreover, as we have mentioned in the Remark 3.5, a slightly weaker delocalization property for the left singular vectors has been provided in [14]. So we will only present the proof for the right singular vectors below.

Below we denote the k-th column of Y by h_k, and the remaining $p \times (n-1)$ matrix after deleting h_k by $Y_{(k)}$. Note that $Y_{(n)}$ is not independent of the last column h_n. However, for the sample covariance matrix case, the independence between the column and the corresponding submatrix is essential for one to use the concentration results such as Lemma 2.5. To overcome the inconvenience caused by the dependence, we will use the modified matrix $\widehat{Y}_{(n)}$ defined above. Notice that the matrix $\widehat{Y}_{(n)}$ is independent of the random vector $(x_{1n}, x_{2n} \cdots, x_{pn})^T$.

Now we define

$$\Delta_1 = Y_{(n)}^T Y_{(n)} - \widehat{Y}_{(n)}^T \widehat{Y}_{(n)},$$

and

$$\Delta_2 = Y_{(n)}^T - \widehat{Y}_{(n)}^T. \tag{3.18}$$

The following lemma handles the operator norms of Δ_1 and Δ_2.

Lemma 3.11. *Under the assumption of Theorem 3.4, we have*

$$||\Delta_1||_{op}, ||\Delta_2||_{op} = O(\frac{K^2}{n})$$

with overwhelming probability.

Proof. Observe that

$$\Delta_1 = (Y_{(n)}^T - \widehat{Y}_{(n)}^T)Y_{(n)} + \widehat{Y}_{(n)}^T(Y_{(n)} - \widehat{Y}_{(n)}) = \Delta_2 Y_{(n)} + \widehat{Y}_{(n)}^T \Delta_2^T.$$

We only discuss the second term since the first one is analogous. It is easy to see the entries of Δ_2^T satisfy

$$\frac{x_{ij}}{||\mathbf{x}_i||} - \frac{x_{ij}}{||\widehat{\mathbf{x}}_i||} = \frac{x_{ij}(||\widehat{\mathbf{x}}_i||^2 - ||\mathbf{x}_i||^2)}{||\mathbf{x}_i|| ||\widehat{\mathbf{x}}_i||(||\mathbf{x}_i|| + ||\widehat{\mathbf{x}}_i||)} = -\frac{x_{in}^2}{||\mathbf{x}_i||(||\mathbf{x}_i|| + ||\widehat{\mathbf{x}}_i||)} \cdot \frac{x_{ij}}{||\widehat{\mathbf{x}}_i||}.$$

It follows that

$$\widehat{Y}_{(n)}^T \Delta_2^T := -\widehat{Y}_{(n)}^T \Delta_3 \widehat{Y}_{(n)},$$

where Δ_3 is a $p \times p$ diagonal matrix with (i,i)-th entry to be

$$\frac{x_{in}^2}{||\mathbf{x}_i||(||\mathbf{x}_i|| + ||\widehat{\mathbf{x}}_i||)}.$$

Thus it is easy to see

$$||\Delta_3||_{op} = O(\frac{K^2}{n}),$$

with overwhelming probability. Together with the fact that $||\widehat{Y}_{(n)}||_{op} \leq C$ holds with overwhelming probability, we can conclude the proof of Lemma 3.11. $\square$

Now we proceed to the proof of Theorem 3.4. If we denote

$$u_i = \begin{pmatrix} \mathbf{w} \\ x \end{pmatrix},$$

where x is the last component of u_i. Without loss of generality, we can only prove the theorem for x. Notice that u_i is the eigenvector of $\mathcal{W} = Y^T Y$ corresponding to the eigenvalue λ_i. From

$$\begin{pmatrix} Y_{(n)}^T Y_{(n)} & Y_{(n)}^T h_n \\ h_n^T Y_{(n)} & h_n^T h_n \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ x \end{pmatrix} = \lambda_i \begin{pmatrix} \mathbf{w} \\ x \end{pmatrix},$$

we have

$$Y_{(n)}^T Y_{(n)} \mathbf{w} + x Y_{(n)}^T h_n = \lambda_i \mathbf{w}, \tag{3.19}$$

and

$$h_n^T Y_{(n)} \mathbf{w} + x h_n^T h_n = \lambda_i x. \tag{3.20}$$

(3.19) can be rewritten as

$$(\widehat{Y}_{(n)}^T \widehat{Y}_{(n)} + \Delta_1) \mathbf{w} + x(\widehat{Y}_{(n)}^T + \Delta_2) h_n = \lambda_i \mathbf{w}.$$

It follows that

$$(\widehat{Y}_{(n)}^T \widehat{Y}_{(n)} - \lambda_i) \mathbf{w} = -x \widehat{Y}_{(n)}^T h_n - x \Delta_2 h_n - \Delta_1 \mathbf{w}. \tag{3.21}$$

Note that $\widehat{Y}_{(n)}^T \widehat{Y}_{(n)}$ share the same nonzero eigenvalues with $\widehat{W}_{(n)}$, so by $iii)$ of Lemma 3.10, we can always view that the matrix $\widehat{Y}_{(n)}^T \widehat{Y}_{(n)} - \lambda_i$ is invertible. Consequently,

$$||\mathbf{w}||^2 = [x\widehat{Y}_{(n)}^T h_n + x\Delta_2 h_n + \Delta_1 \mathbf{w}]^T (\widehat{Y}_{(n)}^T \widehat{Y}_{(n)} - \lambda_i)^{-2} [x\widehat{Y}_{(n)}^T h_n + x\Delta_2 h_n + \Delta_1 \mathbf{w}]$$

If $x = 0$ then Theorem 3.4 is evidently true. Consider $x \neq 0$ below. Together with the fact that $x^2 = 1 - ||\mathbf{w}||^2$, we have

$$x^2 = \frac{1}{1 + [\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w}]^T (\widehat{Y}_{(n)}^T \widehat{Y}_{(n)} - \lambda_i)^{-2} [\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w}]}.$$

Now if we use $\hat{\lambda}_j$ to denote the ordered nonzero eigenvalue of $\widehat{Y}_{(n)}^T \widehat{Y}_{(n)}$ and $\hat{u}_j$ the corresponding unit eigenvector. And set the projection

$$\widehat{P} = I - \sum_{j=1}^{p} \hat{u}_j \hat{u}_j^T.$$

Then by the spectral decomposition one has

$$x^2 = \frac{1}{1 + \sum_{j=1}^{p} \frac{1}{(\hat{\lambda}_j - \lambda_i)^2} |\hat{u}_j \cdot (\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})|^2 + \Delta}, \tag{3.22}$$

where

$$\Delta = \frac{1}{\lambda_i^2} ||\widehat{P}(\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})||^2.$$

Therefore to show $|x| \leq n^{-1/2} K^{C_0/2} \log^{O(1)} n$, it suffices to prove

$$\sum_{j=1}^{p} \frac{1}{(\hat{\lambda}_j - \lambda_i)^2} |\hat{u}_j \cdot (\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1} \Delta_1 \mathbf{w})|^2 \geq nK^{-C_0} \log^{-O(1)} n. \quad (3.23)$$

To verify (3.23), we need to separate the issue into the bulk case and the edge case. Before that, we shall provide the following lemma which will be used in both cases.

Lemma 3.12. *If we denote the unit eigenvector of $\widehat{W}_{(n)}$ corresponding to $\hat{\lambda}_j$ by $\hat{v}_j$, under the assumption of Theorem 3.4 we have for any $J \subseteq \{1, \cdots, p\}$ with $|J| = d \leq nK^{-3}$,*

$$\sqrt{n}[\sum_{j \in J} (\hat{v}_j \cdot h_n)^2]^{1/2} = \sqrt{d} + O(K \log n)$$

with overwhelming probability.

We will postpone the proof of Lemma 3.12 to Appendix B. In fact, it can be viewed as a minor modification of Lemma 2.5.

Now we decompose the proof of Theorem 3.4 into two parts: bulk case and edge case.

• Bulk case: $\lambda_i \in [a + \epsilon, b - \epsilon]$ *for some* $\epsilon > 0$

Note that the local MP law (Theorem 3.1) can also be applied to the matrix $\widehat{Y}_{(n)}^T \widehat{Y}_{(n)}$. Thus we can find a set $J \subseteq \{1, \cdots, p\}$ with $|J| \geq K^2 \log^{20} n$ such that $\hat{\lambda}_j = \lambda_i + O(K^2 \log^{20} n/n)$ for any $j \in J$ when λ_i is in the bulk region of the MP law. It follows that

$$\sum_{j \in J} \frac{1}{(\hat{\lambda}_j - \lambda_i)^2} |\hat{u}_j \cdot (\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1} \Delta_1 \mathbf{w})|^2$$
$$\geq C \frac{n^2}{K^4 \log^{40} n} \sum_{j \in J} |\hat{u}_j \cdot (\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1} \Delta_1 \mathbf{w})|^2. \quad (3.24)$$

By the singular value decomposition, we have

$$\hat{u}_j \cdot \widehat{Y}_{(n)}^T h_n = \hat{\lambda}_j^{1/2} \hat{v}_j \cdot h_n. \quad (3.25)$$

Now we compare

$$\sum_{j \in J} |\hat{u}_j \cdot \widehat{Y}_{(n)}^T h_n|^2 = \sum_{j \in J} \hat{\lambda}_j |\hat{v}_j \cdot h_n|^2 \quad (3.26)$$

with

$$\sum_{j \in J} |\hat{u}_j \cdot (\Delta_2 h_n + x^{-1} \Delta_1 \mathbf{w})|^2 \quad (3.27)$$

for any $J \subset \{1, \cdots, p\}$ such that $K^2 \log^{20} n \leq |J| \leq nK^{-3}$.

If $|x| \leq n^{-1/2} K^{C_0/2} \log^{O(1)} n$, then we get the conclusion for the bulk case. So we assume $|x| \geq n^{-1/2} K^{C_0/2} \log^{O(1)} n$ below to get (3.23). By Lemma 3.11, if we choose $C_0 \geq 20$ (say), we have

$$
\begin{aligned}
(3.27) &\leq 2|J|(||\Delta_2||_{op}||h_n||)^2 + 2x^{-2}|J|(||\Delta_1||_{op}||\mathbf{w}||)^2 \\
&\leq |J|n^{-1} K^{-C_0/2} \log^{-O(1)} n
\end{aligned}
\quad (3.28)
$$

with overwhelming probability. On the other side, Lemma 3.12 implies

$$(3.26) = Cn^{-1}(|J| + O(K^2 \log^2 n)) \qquad (3.29)$$

with overwhelming probability. So one has

$$\sum_{j \in J} |\hat{u}_j \cdot \widehat{Y}_{(n)}^T h_n|^2 \gg \sum_{j \in J} |\hat{u}_j \cdot (\Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})|^2, \qquad (3.30)$$

where $\gg$ means "much larger than", i.e.

$$\Big(\sum_{j \in J} |\hat{u}_j \cdot (\Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})|^2\Big) / \Big(\sum_{j \in J} |\hat{u}_j \cdot \widehat{Y}_{(n)}^T h_n|^2\Big) = o(1).$$

Notice that for any real number sequence $\{S_1, \cdots, S_m\}$ and $\{T_1, \cdots, T_m\}$ with $\sum_{i=1}^m S_i^2 \gg \sum_{i=1}^m T_i^2$, there exists some c near 1 such that $\sum_{i=1}^m (S_i + T_i)^2 \geq c \sum_{i=1}^m S_i^2$. Therefore by (3.29),(3.30) and (3.24) we can obtain

$$\sum_{j \in J} \frac{1}{(\hat{\lambda}_j - \lambda_i)^2} |\hat{u}_j \cdot (\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})|^2 \geq CnK^{-2} \log^{-20} n,$$

which implies (3.23) directly. So we conclude the proof for the bulk case.

Next, we turn to the edge case.

• Edge case: $a - o(1) \leq \lambda_i \leq a + \epsilon$ or $b - \epsilon \leq \lambda_i \leq b + o(1)$ with some $\epsilon > 0$.

For the edge case we also begin with the representation (3.22). By (3.21), we have

$$\mathbf{w} = -x(\widehat{Y}_{(n)}^T \widehat{Y}_{(n)} - \lambda_i)^{-1}(\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w}). \qquad (3.31)$$

Inserting (3.31) and (3.18) into (3.20) we find

$$(\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n)^T (\widehat{Y}_{(n)}^T \widehat{Y}_{(n)} - \lambda_i)^{-1}(\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w}) = h_n^T h_n - \lambda_i.$$

Furthermore,

$$|x^{-1}\mathbf{w}^T \Delta_1^T (\widehat{Y}_{(n)}^T \widehat{Y}_{(n)} - \lambda_i)^{-1}(\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})|$$

$$= |x^{-2}\mathbf{w}^T \Delta_1^T \mathbf{w}| \leq |x|^{-2}||\Delta_1||_{op}||\mathbf{w}||^2 = ||\Delta_1||_{op} \frac{1 - x^2}{x^2}.$$

Thus one has

$$(\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})^T (\widehat{Y}_{(n)}^T \widehat{Y}_{(n)} - \lambda_i)^{-1}(\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})$$

$$= h_n^T h_n - \lambda_i + O\left(||\Delta_1||_{op} \frac{1 - x^2}{x^2}\right). \qquad (3.32)$$

Similarly to the bulk case, we only need to get (3.23). Below we also assume $|x| \geq Cn^{-1/2}K^{C_0/2}\log^{O(1)} n$ to get (3.23). Similar to (3.28), by using Lemma 3.11 we have

$$|\hat{u}_j \cdot (\Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})|^2 \leq n^{-1}K^{-C_0/2}\log^{-O(1)} n. \qquad (3.33)$$

Moreover, by Lemma 3.12 and (3.25), we also have

$$|\hat{u}_j \cdot \widehat{Y}_{(n)}^T h_n|^2 = \hat{\lambda}_i |\hat{v}_j \cdot h_n| \leq Cn^{-1}K^2 \log^2 n \qquad (3.34)$$

with overwhelming probability. Thus to provide (3.23), it suffices to show

$$\sum_{j=1}^{p} \frac{1}{(\hat{\lambda}_j - \lambda_i)^2} |\hat{u}_j \cdot (\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})|^4 \geq K^{-C_0+2} \log^{-O(1)} n$$

instead. By the Cauchy-Schwarz inequality, we only need to prove

$$\sum_{i-T_- \leq j \leq i+T_+} \frac{1}{|\hat{\lambda}_j - \lambda_i|} |\hat{u}_j \cdot (\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})|^2 \geq \log^{-O(1)} n \qquad (3.35)$$

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

Notice that under the assumption $|x| \geq C n^{-1/2} K^{C_0/2} \log^{O(1)} n$, by Lemma 3.11 we have

$$\|\Delta_1\|_{op} \frac{1-x^2}{x^2} = o(1).$$

Moreover, it is not difficult to see $h_n^T h_n = y + o(1)$ with overwhelming probability. Thus by (3.32), we have with overwhelming probability

$$\sum_{j=1}^{p} \frac{1}{(\hat{\lambda}_j - \lambda_i)} |\hat{u}_j \cdot (\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})|^2 - \frac{1}{\lambda_i} \|\widehat{P}(\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})\|^2$$

$$= h_n^T h_n - \lambda_i + O(\|\Delta_1\|_{op} \frac{1-x^2}{x^2}) = y - \lambda_i + o(1).$$

Observing that

$$\widehat{P}\widehat{Y}_{(n)}^T h_n = 0$$

and

$$\|\widehat{P}(\Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})\|^2 \ll n^{-1},$$

we also have

$$\sum_{j=1}^{p} \frac{1}{(\hat{\lambda}_j - \lambda_i)} |\hat{u}_j \cdot (\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})|^2 = y - \lambda_i + o(1). \qquad (3.36)$$

So to prove (3.35) we only need to evaluate

$$\sum_{j<i-T_- \ or \ j>i+T_+} \frac{1}{(\hat{\lambda}_j - \lambda_i)} |\hat{u}_j \cdot (\widehat{Y}_{(n)}^T h_n + \Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w})|^2. \qquad (3.37)$$

To do this, we let $A > 100$ be a constant large enough. For any interval I of length $|I| = K^2 \log^A n/n$, we set $d_I := \frac{\text{dist}(\lambda_i, I)}{|I|}$, where

$$\text{dist}(\lambda_i, I) = \min_{x \in I} |\lambda_i - x| \text{sgn}(\lambda_i, I).$$

Here $\text{sgn}(\lambda_i, I) = 1 (resp. -1)$ when λ_i is on the left (resp. right) hand side of I.

By Theorem 3.1, the interval I with $|d_I| < \log n$ contains at most $K^2 \log^{O(1)} n$ eigenvalues. So we can set T_-, T_+ accordingly so that such intervals don't contain any $\hat{\lambda}_j$ if $j < i - T_-$ or $j > i + T_+$. In the following we only consider I such that $|d_I| \geq \log n$ in the estimation of (3.37). Note that for $\hat{\lambda}_j \in I$,

$$\frac{1}{\hat{\lambda}_j - \lambda_i} = \frac{1}{d_I |I|} + O(\frac{1}{d_I^2 |I|}).$$

Using (3.33) and (3.34) again one has

$$2|(\hat{u}_j \cdot \widehat{Y}_{(n)}^T h_n)(\hat{u}_j \cdot (\Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w}))| + |\hat{u}_j \cdot \Delta_2 h_n + x^{-1}\hat{u}_j \cdot \Delta_1 \mathbf{w}|^2$$
$$\leq \quad Cn^{-1}K^{-O(1)}\log^{-O(1)} n$$

when $|x| \geq n^{-1/2}K^{C_0/2}\log^{O(1)} n$. Thus we can find

$$\sum_{j \in I} \frac{1}{|\hat{\lambda}_j - \lambda_i|}(2|(\hat{u}_j \cdot \widehat{Y}_{(n)}^T h_n)(\hat{u}_j \cdot (\Delta_2 h_n + x^{-1}\Delta_1 \mathbf{w}))| + |\hat{u}_j \cdot \Delta_2 h_n + x^{-1}\hat{u}_j \cdot \Delta_1 \mathbf{w}|^2)$$

$$\leq C\frac{N_I}{|d_I||I|n}K^{-O(1)}\log^{-O(1)} n \leq C\frac{1}{|d_I|}K^{-O(1)}\log^{-O(1)} n. \tag{3.38}$$

Here we used Lemma 3.6 in the last inequality. Now we partition the real line into intervals I of length $K^2 \log^A n/n$, and sum (3.38) over all intervals I with $|d_I| \geq \log n$. Then

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

So we can evaluate

$$\sum_{j<i-T_- \ or \ j>i+T_+} \frac{1}{(\hat{\lambda}_j - \lambda_i)}|\hat{u}_j \cdot \widehat{Y}_{(n)}^T h_n|^2 = \sum_{j<i-T_- \ or \ j>i+T_+} \frac{\hat{\lambda}_j}{(\hat{\lambda}_j - \lambda_i)}|\hat{v}_j \cdot h_n|^2. \tag{3.39}$$

instead of (3.37). The evaluation of (3.39) is really the same as the counterpart in the sample covariance matrix case (see (4.5) in [20]) by inserting Lemma 3.12, so we omit the details here. In fact, we can finally get

$$\sum_{j<i-T_- \ or \ j>i+T_+} \frac{\hat{\lambda}_j}{(\hat{\lambda}_j - \lambda_i)}|\hat{v}_j \cdot h_n|^2 = p.v. \int_a^b y\frac{x}{x - \lambda_i}\rho_{MP,y}(x)dx + o(1)$$

$$= y + \lambda_i \ p.v. \int_a^b \frac{\rho_{MP,y}(x)}{x - \lambda_i}dx + o(1)$$

where $p.v.$ means the principal value.

Using the formula for the Stieltjes transform $s(z)$, one can get from residue calculus that for $\lambda_i \in [a, b]$,

$$p.v. \int_a^b \frac{\rho_{MP,y}(x)}{x - \lambda_i}dx = \frac{1 - y - \lambda_i}{2y\lambda_i},$$

and for $\lambda_i \notin [a, b]$

$$p.v. \int_a^b \frac{\rho_{MP,y}(x)}{x - \lambda_i}dx = \frac{1 - y - \lambda_i + \sqrt{(\lambda_i - 1 - y)^2 - 4y}}{2y\lambda_i}.$$

Consequently by the definition of a and b, if $|\lambda_i - a| \leq o(1)$, we have

$$(3.36) = -1 + 2\sqrt{y} + o(1), \quad (3.39) = \sqrt{y} + o(1).$$

And if $|\lambda_i - b| \leq o(1)$, we have

$$(3.36) = -1 - 2\sqrt{y} + o(1), \quad (3.39) = -\sqrt{y} + o(1).$$

Then it is easy to see when $0 < y < 1$, (3.35) holds with overwhelming probability for the case where $|\lambda_i - a| = o(1)$ or $|\lambda_i - b| = o(1)$. Moreover by continuity we can adjust the value of ϵ to get the conclusion for the general case $a - o(1) \leq \lambda_i \leq a + \epsilon$ or $b - \epsilon \leq \lambda_i \leq b + o(1)$. Thus we complete the proof of the delocalization for u_i. $\quad\square$

4 Green function comparison theorem

In this section, we provide a Green function comparison theorem for the sample correlation matrices satisfying $\mathbf{C_1}$. The proof heavily relies on the recent results of Pillai and Yin [14] on sample covariance matrices and the delocalization property for the right singular vectors proved in the last section. At first, we will borrow some results from [14] directly with only minor notation change. In fact, by Theorem 1.5 in [14], it is not difficult to see Theorem 1.2 and Theorem 1.3 of [14] also hold for sample correlation matrices under our basic condition $\mathbf{C_1}$.

To state the results in [14], we need to introduce some notation. Define the parameter

$$\varphi := (\log p)^{\log \log p},$$

and

$$\lambda_\pm := (1 \pm (\frac{p}{n})^{1/2})^2.$$

Moreover we introduce the "nonasymptotic Marchenko-Pastur law "

$$\rho_W(x) = \frac{n}{2\pi x p} \sqrt{(\lambda_+ - x)(x - \lambda_-)} \mathbf{1}_{[\lambda_-, \lambda_+]}(x)$$

and the corresponding distribution function $F_W(x)$ and Stieltjes transform

$$s_W(z) = \int_{\mathbb{R}} \frac{\rho_W(x)}{x - z} dx.$$

For $\zeta \geq 0$, define the set

$$\underline{S}(\zeta) := \{z \in \mathbb{C} : 0 \leq E \leq 5\lambda_+, \varphi^\zeta p^{-1} \leq \eta \leq 10(1 + \frac{p}{n})\}. \tag{4.1}$$

And we say that an event Ω holds with ζ-high probability if there exists a constant $C > 0$ such that

$$\mathbb{P}(\Omega^c) \leq p^C \exp(-\varphi^\zeta) \tag{4.2}$$

for large enough p. Note that (4.2) implies that the event Ω holds with overwhelming probability if $\zeta > 0$. We further denote

$$\Lambda_d := \max_k |G_{kk} - s_W(z)|, \quad \Lambda_o := \max_{k \neq l} |G_{kl}|, \quad \Lambda := |s_p(z) - s_W(z)|.$$

Lemma 4.1 (Theorem 1.5, [14]). *Under the condition $\mathbf{C_1}$, for any $\zeta > 0$ there exists a constant C_ζ such that the following events hold with ζ-high probability.*

(i) The Stieltjes transform of the ESD of W satisfies

$$\bigcap_{z \in \underline{S}(C_\zeta)} \left\{ \Lambda(z) \leq \varphi^{C_\zeta} \frac{1}{p\eta} \right\}.$$

(ii) The individual matrix elements of the Green function satisfy

$$\bigcap_{z \in \underline{S}(C_\zeta)} \left\{ \Lambda_o(z) + \Lambda_d(z) \leq \varphi^{C_\zeta} \left(\sqrt{\frac{\Im s_W(z)}{p\eta}} + \frac{1}{p\eta} \right) \right\}.$$

(iii) Uniformly in $E \in \mathbb{R}$,

$$|F_p(E) - F_W(E)| \leq \varphi^{C_\zeta} p^{-1}.$$

We also need the following lemma on $s_W(z)$.

Lemma 4.2 (Lemma 26, [14]). *Set* $\kappa := \min(|\lambda_+ - E|, |E - \lambda_-|)$. *For* $z = E + i\eta \in \underline{S}(0)$, *(see (4.1)) we have the following relations:*

$$|s_W(z)| \sim 1, \qquad |1 - s_W^2(z)| \sim \sqrt{\kappa + \eta}, \tag{4.3}$$

$$\Im s_W(z) \sim \begin{cases} \dfrac{\eta}{\sqrt{\kappa + \eta}} & if \quad \kappa \geq \eta \quad and \quad |E| \notin [\lambda_-, \lambda_+] \\[2mm] \sqrt{\kappa + \eta} & if \quad \kappa \leq \eta \quad and \quad |E| \in [\lambda_-, \lambda_+] \end{cases} \tag{4.4}$$

where $A \sim B$ *means* $C^{-1}B \leq A \leq CB$ *for some constant* C. *Furthermore*

$$\frac{\Im s_W(z)}{p\eta} \geq O(\frac{1}{p}) \quad and \quad \partial_\eta \frac{\Im s_W(z)}{\eta} \leq 0.$$

Now we set $Y^{\mathbf{v}} = (\mathbf{y}_{ij}^{\mathbf{v}}) := (x_{ij}^{\mathbf{v}}/\|\mathbf{x}_i^{\mathbf{v}}\|)_{p,n}$, with elements $x_{ij}^{\mathbf{v}}$ satisfying our basic condition $\mathbf{C_1}$. Correspondingly we let $W^{\mathbf{v}} = Y^{\mathbf{v}}Y^{\mathbf{v}T}$, $G^{\mathbf{v}}(z) = (W^{\mathbf{v}} - z)^{-1}$ and $s_p^{\mathbf{v}}(z) = \frac{1}{p}TrG^{\mathbf{v}}(z)$. Define the matrix $W^{\mathbf{w}}$, the Green function $G^{\mathbf{w}}(z)$ and the Stieltjes transform $s_p^{\mathbf{w}}(z)$ analogously for another random sequence $\{x_{ij}^{\mathbf{w}}\}$ satisfying $\mathbf{C_1}$ which is independent of $\{x_{ij}^{\mathbf{v}}\}$. The aim in this section is to prove the following Green function comparison theorem.

Below we only state the results and proofs for the largest eigenvalue. The smallest one is just analogous.

Theorem 4.3 (Green function comparison theorem on the edge). . *Let* $F : \mathbb{R} \to \mathbb{R}$ *be a function whose derivatives* $F^{(\alpha)}$ *satisfy*

$$\max_x |F^{(\alpha)}(x)|(|x| + 1)^{-C_1} \leq C_1, \quad \alpha = 1, 2, 3, 4$$

with some constant $C_1 > 0$. *Then there exists* $\epsilon_0 > 0$ *depending only on* C_1 *such that for any positive* $\epsilon < \epsilon_0$ *and for any real numbers* E, E_1 *and* E_2 *satisfying*

$$|E - \lambda_+| \leq p^{-2/3+\epsilon}, \quad |E_1 - \lambda_+| \leq p^{-2/3+\epsilon}, \quad |E_2 - \lambda_+| \leq p^{-2/3+\epsilon},$$

and $\eta = p^{-2/3-\epsilon}$, *we have*

$$\left| \mathbb{E}^{\mathbf{v}} F(p\eta \Im s_p^{\mathbf{v}}(z)) - \mathbb{E}^{\mathbf{w}} F(p\eta \Im s_p^{\mathbf{w}}(z)) \right| \leq Cp^{-1/6+C\epsilon}, \quad z = E + i\eta, \tag{4.5}$$

and

$$\left| \mathbb{E}^{\mathbf{v}} F\left(p \int_{E_1}^{E_2} dx \Im s_p^{\mathbf{v}}(x + i\eta)\right) - \mathbb{E}^{\mathbf{w}} F\left(p \int_{E_1}^{E_2} dx \Im s_p^{\mathbf{w}}(x + i\eta)\right) \right| \leq Cp^{-1/6+C\epsilon} \tag{4.6}$$

for some positive constant C *and large enough* p.

Proof of Theorem 4.3. The proof is similar to that of Theorem 6.3 of [14]. Moreover, the proof of (4.6) can be taken in a same manner as that of (4.5), so we will just present the proof for (4.5) below. The basic strategy is to estimate the successive difference of matrices which differ by a row. For $1 \leq \gamma \leq p$, we denote by Y_γ the random matrix whose j-th row is the same as that of $Y^{\mathbf{v}}$ if $j \leq \gamma$ and that of $Y^{\mathbf{w}}$ otherwise; in particular $Y_0 = Y^{\mathbf{v}}$ and $Y_p = Y^{\mathbf{w}}$. And we set

$$W_\gamma = Y_\gamma Y_\gamma^T.$$

We shall compare $W_{\gamma-1}$ with W_γ by using the following lemma. For simplicity, we denote

$$s_p^{(i)}(z) = \frac{1}{p}TrG^{(i)}(z), \quad \tilde{s}_p^{(i)}(z) = s_p^{(i)}(z) - \frac{1}{pz}.$$

Lemma 4.4. *For any sample correlation matrix W with elements satisfying the basic assumption* $\mathbf{C_1}$, *if $|E - \lambda_+| \leq p^{-2/3+\epsilon}$ and $p^{-2/3} \gg \eta \gg p^{-2/3-\epsilon}$ for some small enough $\epsilon > 0$, then we have*

$$\mathbb{E}F(p\eta\Im s_p(z)) - \mathbb{E}F(p\eta\Im \tilde{s}_p^{(i)}(z)) = A(Y^{(i)}, m_1, m_2) + O(p^{-7/6+C\epsilon})$$

where the functional $A(Y^{(i)}, m_1, m_2)$ only depends on the distribution of $Y^{(i)}$ and the first two moments m_1, m_2 of x_{ij}.

Remark 4.5. *We always assume $m_1 = 0$, $m_2 = 1$ in our case.*

Note that

$$W_{\gamma-1}^{(\gamma)} = W_\gamma^{(\gamma)},$$

thus Lemma 4.4 implies that

$$\mathbb{E}F\left(\eta\Im Tr(W_{\gamma-1} - z)^{-1}\right) - \mathbb{E}F\left(\eta\Im Tr(W_\gamma - z)^{-1}\right) = O(p^{-7/6+C\epsilon}).$$

Then the proof of Theorem 4.3 can be completed by the telescoping argument.

Therefore it suffices to prove Lemma 4.4 in the sequel. To do this, we need to provide some bounds about $\mathcal{G}^{(i)}$. We only state the result for $i = 1$ as the following lemma since the others are analogous.

Lemma 4.6. *Under the assumptions in Lemma 4.4, we have for $\epsilon > 0$ small enough,*

$$|\mathbf{y}_1^T(\mathcal{G}^{(1)})^2\mathbf{y}_1| \leq p^{1/3+C\epsilon} \tag{4.7}$$

and

$$|(\mathcal{G}^{(1)})_{ij}| \leq p^{C\epsilon}, \quad |((\mathcal{G}^{(1)})^2)_{ij}| \leq p^{1/3+C\epsilon} \tag{4.8}$$

hold with overwhelming probability.

The proof of Lemma 4.6 will be postponed to the end of this section. Now we begin to prove Lemma 4.4 assuming Lemma 4.6.

Proof of Lemma 4.4. The proof is in a similar manner to that of Lemma 6.5 in [14]. At first we rewrite (2.7) as

$$G_{11} = \frac{1}{-z - z\mathbf{y}_1^T\mathcal{G}^{(1)}(z)\mathbf{y}_1} \tag{4.9}$$

by using the facts that

$$\mathcal{W}^{(1)}\mathcal{G}^{(1)}(z) = I + z\mathcal{G}^{(1)}(z), \quad \mathbf{y}_1^T\mathbf{y}_1 = 1.$$

Moreover, by Schur's complement, we also have

$$TrG - TrG^{(1)} = G_{11} + \frac{\mathbf{y}_1^T Y^{(1)T}(G^{(1)})^2 Y^{(1)}\mathbf{y}_1}{-z - z\mathbf{y}_1^T\mathcal{G}^{(1)}(z)\mathbf{y}_1}, \tag{4.10}$$

Inserting (4.9) and the identity

$$Y^{(1)T}(G^{(1)})^2 Y^{(1)} = \mathcal{W}^{(1)}(\mathcal{G}^{(1)})^2 = \mathcal{G}^{(1)} + z(\mathcal{G}^{(1)})^2$$

into (4.10) we can get

$$TrG - TrG^{(1)} + z^{-1} = zG_{11}(\mathbf{y}_1^T(\mathcal{G}^{(1)})^2(z)\mathbf{y}_1). \tag{4.11}$$

Now we define the quantity B as

$$B = -zs_W(z)\left[\mathbf{y}_1^T \mathcal{G}^{(1)}(z)\mathbf{y}_1 - \left(\frac{-1}{zs_W(z)} - 1\right)\right].$$

Thus by (4.9) we have

$$B = -zs_W(z)\left[\left(\frac{-1}{zG_{11}(z)} - 1\right) - \left(\frac{-1}{zs_W(z)} - 1\right)\right] = \frac{s_W(z) - G_{11}}{G_{11}}.$$

By (ii) of Lemma 4.1 and (4.4) we can get

$$|B| \le p^{-1/3+2\epsilon} \ll 1$$

with overwhelming probability. Thus we have the expansion

$$G_{11} = \frac{s_W(z)}{B+1} = s_W(z)\sum_{k\ge 0}(-B)^k. \tag{4.12}$$

Now we set

$$\phi := \eta(TrG - TrG^{(1)} + z^{-1}).$$

It follows from (4.11) and (4.12) that

$$\phi = \eta zG_{11}\mathbf{y}_1^T(\mathcal{G}^{(1)})^2\mathbf{y}_1 = \sum_{k=1}^{\infty}\phi_k,$$

where

$$\phi_k := \eta zs_W(z)(-B)^{k-1}\mathbf{y}_1^T(\mathcal{G}^{(1)})^2\mathbf{y}_1.$$

Since z and $s_W(z)$ are $O(1)$ by (4.3), by definitions and Lemma 4.6, we have

$$|\phi_k| \le O(p^{-k/3+C\epsilon}) \quad\text{and}\quad |\phi| \le O(p^{-1/3+C\epsilon}) \tag{4.13}$$

with overwhelming probability. Thus we have

$$F(p\eta\Im s_p(z)) - F(p\eta\Im\tilde{s}_p^{(1)}(z))$$
$$= \sum_{k=1}^{3}\frac{1}{k!}F^{(k)}(p\eta\Im\tilde{s}_p^{(1)}(z))(\Im\phi)^k + O(p^{-4/3+C\epsilon})$$

with overwhelming probability.

Similarly to the counterpart proof of Lemma 6.5 in [14], we only need to show

$$\mathbb{E}F^{(k)}(p\eta\Im\tilde{s}_p^{(1)}(z))(\Im\phi)^k = A_k(Y^{(1)}, m_1, m_2) + O(p^{-4/3+C\epsilon}), \quad k = 1, 2, 3 \tag{4.14}$$

with some functional A_k only depending on the distribution of $Y^{(1)}$, m_1 and m_2.

Since the proof of (4.14) is similar to the counterpart in [14], we will only state the proof for $k = 3$ below. We use $\mathbb{E}_1$ to denote the expectation with respect to $\mathbf{y}_1$ in the sequel. By using (4.13) we obtain

$$F^{(3)}(p\eta\Im\tilde{s}_p^{(1)}(z))(\Im\phi)^3 = F^{(3)}(p\eta\Im\tilde{s}_p^{(1)}(z))(\Im\phi_1)^3 + O(p^{-4/3+C\epsilon}) \tag{4.15}$$

with overwhelming probability. If we write $r_1 = \Re(\eta z s_W(z)), r_2 = \Im(\eta z s_W(z))$, then we have

$$
\begin{aligned}
\mathbb{E}_1(\Im\phi_1)^3 &= \mathbb{E}_1 r_1^3(\Im(\mathbf{y}_1^T(\mathcal{G}^{(1)})^2\mathbf{y}_1))^3 + \mathbb{E}_1 r_2^3(\Re(\mathbf{y}_1^T(\mathcal{G}^{(1)})^2\mathbf{y}_1))^3 \\
&\quad +3\mathbb{E}_1 r_1^2 r_2(\Im(\mathbf{y}_1^T(\mathcal{G}^{(1)})^2\mathbf{y}_1))^2(\Re(\mathbf{y}_1^T(\mathcal{G}^{(1)})^2\mathbf{y}_1)) \\
&\quad +3\mathbb{E}_1 r_1 r_2^2(\Im(\mathbf{y}_1^T(\mathcal{G}^{(1)})^2\mathbf{y}_1))(\Re(\mathbf{y}_1^T(\mathcal{G}^{(1)})^2\mathbf{y}_1))^2 \\
&= r_1^3 \sum_{k_1,\cdots,k_6} \mathbb{E}_1(\prod_{i=1}^{6}\frac{x_{1k_i}}{||\mathbf{x}_1||})\prod_{i=1}^{3}\Im((\mathcal{G}^{(1)})^2)_{k_{2i-1},k_{2i}} \\
&\quad +r_2^3 \sum_{k_1,\cdots,k_6} \mathbb{E}_1(\prod_{i=1}^{6}\frac{x_{1k_i}}{||\mathbf{x}_1||})\prod_{i=1}^{3}\Re((\mathcal{G}^{(1)})^2)_{k_{2i-1},k_{2i}} \\
&\quad +3r_1 r_2^2 \sum_{k_1,\cdots,k_6} \mathbb{E}_1(\prod_{i=1}^{6}\frac{x_{1k_i}}{||\mathbf{x}_1||})\prod_{i=1}^{2}\Re((\mathcal{G}^{(1)})^2)_{k_{2i-1},k_{2i}}\Im((\mathcal{G}^{(1)})^2)_{k_5,k_6} \\
&\quad +3r_1^2 r_2 \sum_{k_1,\cdots,k_6} \mathbb{E}_1(\prod_{i=1}^{6}\frac{x_{1k_i}}{||\mathbf{x}_1||})\prod_{i=1}^{2}\Im((\mathcal{G}^{(1)})^2)_{k_{2i-1},k_{2i}}\Re((\mathcal{G}^{(1)})^2)_{k_5,k_6}(4.16)
\end{aligned}
$$

Notice that if there exists a k_i which appears only once in the above product, then by the assumption that x_{ij} is symmetric, we have

$$
\mathbb{E}_1(\prod_{i=1}^{6}\frac{x_{1k_i}}{||\mathbf{x}_1||}) = 0 = m_1. \tag{4.17}
$$

So we first consider the case where k_i appears exactly twice. Note that

$$
||\mathbf{x}_1||^6 = \sum_{k_1,k_2,k_3} x_{1k_1}^2 x_{1k_2}^2 x_{1k_3}^2 := \sum_{(1)} x_{1k_1}^2 x_{1k_2}^2 x_{1k_3}^2 + \sum_{(2)} x_{1k_1}^2 x_{1k_2}^2 x_{1k_3}^2,
$$

where the first summation goes through the indices k_1, k_2, k_3 such that they are not equal to each other, and the second summation goes through the left part of the indices. Then it is not difficult to see the number of the terms in the second summation is of the order $O(n^2)$. By the exponential tail assumption and the Hoeffding inequality, we can see

$$
\mathbb{E}_1 \sum_{(2)} \frac{x_{1k_1}^2 x_{1k_2}^2 x_{1k_3}^2}{||\mathbf{x}_1||^6} = O(\frac{\log^{O(1)} n}{n}).
$$

Furthermore, since $x_{11},\cdots,x_{1n}$ are i.i.d., we have for k_1, k_2, k_3 not equal to each other

$$
\mathbb{E}_1 \frac{x_{1k_1}^2 x_{1k_2}^2 x_{1k_3}^2}{||\mathbf{x}_1||^6} = \frac{1}{n(n-1)(n-2)}(1 - O(\frac{\log^{O(1)} n}{n})) = \frac{m_2^3}{n^3} + O(\frac{\log^{O(1)} n}{n^4}). \tag{4.18}
$$

Therefore by (4.16), (4.17), (4.18) and the fact that $\mathcal{G}^{(1)}$ only depends on $Y^{(1)}$, we have

$$
\begin{aligned}
&|\mathbb{E}_1(\Im\phi_1)^3 - \tilde{A}_3(Y^{(1)}, m_1, m_2)| \\
\leq\quad &\frac{\log^{O(1)} n}{n^4}|\eta z s_W(z)|^3 \sum_{(3)} |[(\mathcal{G}^{(1)})^2]_{k_1,k_2}[(\mathcal{G}^{(1)})^2]_{k_3,k_4}[(\mathcal{G}^{(1)})^2]_{k_5,k_6}| \\
&+C|\eta z s_W(z)|^3 \sum_{(4),(5)} \mathbb{E}_1\left|\prod_{i=1}^{6}\frac{x_{1k_i}}{||\mathbf{x}_1||}\right| \cdot |[(\mathcal{G}^{(1)})^2]_{k_1,k_2}[(\mathcal{G}^{(1)})^2]_{k_3,k_4}[(\mathcal{G}^{(1)})^2]_{k_5,k_6}|(4.19)
\end{aligned}
$$

with some functional $\tilde{A}_3$ only depending on the distribution of $Y^{(1)}$, m_1 and m_2. Here the first summation $\sum_{(3)}$ in (4.19) goes through the terms such that each $k_i, i = 1, \cdots, 6$ appears exactly twice. It is easy to see that there are $O(n^3)$ such terms totally. And the second summation goes through the terms such that (4) no k_i appears only once and (5) at least one k_i appears three times. Thus we have the total number of the terms in the second summation is of the order $O(n^2)$. Then by using Lemma 4.6 and the fact

$$\mathbb{E}_1 \left| \prod_{i=1}^{6} \frac{x_{1k_i}}{||\mathbf{x}_1||} \right| = O\left(\frac{\log^{O(1)} n}{n^3} \right),$$

we have

$$\mathbb{E}_1(\Im\phi_1)^3 = \tilde{A}_3(Y^{(1)}, m_1, m_2) + O(p^{-2+C\epsilon}) \tag{4.20}$$

By inserting (4.20) into (4.15), we can get (4.14) for $k = 3$. The cases of $k = 1$ and $k = 2$ can be proved similarly by inserting Lemma 4.6. So we conclude the proof. $\qquad \square$

Now we begin to prove Lemma 4.6.

Proof of Lemma 4.6. The proof of (4.7) is the same as the counterpart in [14], (see (6.36) of [14]). So we only state the proof of (4.8) below. For the ease of the presentation, we prove (4.8) for $\mathcal{G} = (\mathcal{W} - z)^{-1} := (Y^TY - z)^{-1}$ instead of $\mathcal{G}^{(1)}$. By the spectral decomposition, we have

$$\mathcal{G}^\alpha = \sum_{k=1}^{p} \frac{1}{(\lambda_k - z)^\alpha} u_k u_k^T + \frac{1}{(-z)^\alpha} P, \qquad \alpha = 1, 2,$$

where the projection $P = I - \sum_{k=1}^{p} u_k u_k^T$. Consequently, we have

$$(\mathcal{G}^\alpha)_{ij} = \sum_{k=1}^{p} \frac{1}{(\lambda_k - z)^\alpha} u_{ki} u_{kj} + \frac{1}{(-z)^\alpha} P_{ij}.$$

Note that $|P_{ij}| \leq 1, |z| \geq \lambda_+/2$. By the delocalization property of u_k in Theorem 3.4 one has

$$|(\mathcal{G}^\alpha)_{ij}| \leq \frac{\log^{O(1)} p}{p} \sum_{k=1}^{p} \frac{1}{|\lambda_k - z|^\alpha} + C$$

with overwhelming probability. For $\alpha = 2$, by using $i)$ of Lemma 4.1 and (4.4) we have

$$\sum_{k=1}^{p} \frac{1}{|\lambda_k - z|^2} = p\eta^{-1}\Im s_p(z) \leq p^\epsilon p\eta^{-1} \frac{1}{p\eta} \leq p^{4/3+C\epsilon},$$

which implies

$$|(\mathcal{G}^2)_{ij}| \leq p^{1/3+C\epsilon}.$$

For $\alpha = 1$, we have

$$\sum_{k=1}^{p} \frac{1}{|\lambda_k - z|} = p \int \frac{1}{|x - z|} dF_p(x).$$

Observe that

$$\left| p \int \frac{1}{|x - z|} dF_p(x) - p \int \frac{1}{|x - z|} dF_W(x) \right| \leq Cp \int \frac{|F_p(x) - F_W(x)|}{|x - z|^2} dx \leq \eta^{-1} p^{C\epsilon}$$

with overwhelming probability. Here we used (iii) of Lemma 4.1 in the last inequality. Consequently, we have

$$|\mathcal{G}_{ij}| \leq (\log^{O(1)} p) \int \frac{1}{|x-z|} dF_W(x) + C.$$

It remains to estimate $\int \frac{1}{|x-z|} dF_W(x)$. For $E < \lambda_+$ such that $\lambda_+ - E \leq p^{-2/3+\epsilon}$

$$\int \frac{1}{|x-z|} dF_W(x) = \left(\int_{\lambda_-}^{2E-\lambda_+} + \int_{2E-\lambda_+}^{\lambda_+} \right) \frac{1}{\sqrt{(x-E)^2 + \eta^2}} dF_W(x).$$

By the formula for the MP law, one has

$$\begin{aligned}
\int_{\lambda_-}^{2E-\lambda_+} \frac{1}{\sqrt{(x-E)^2 + \eta^2}} dF_W(x) &\leq C \int_{\lambda_-}^{2E-\lambda_+} \frac{\sqrt{\lambda_+ - x}}{E - x} dx \\
&\leq C \int_{\lambda_-}^{2E-\lambda_+} \frac{1}{\sqrt{E-x}} dx = O(1), \quad (4.21)
\end{aligned}$$

and

$$\int_{2E-\lambda_+}^{\lambda_+} \frac{1}{\sqrt{(x-E)^2 + \eta^2}} dF_W(x) \leq \eta^{-1} \int_{2E-\lambda_+}^{\lambda_+} dF_W(x) = o(1).$$

When $E \geq \lambda_+$, we still have (4.21). Therefore, we have

$$|\mathcal{G}_{ij}| \leq p^{C\epsilon}$$

with overwhelming probability. Thus we complete the proof. $\qquad\square$

Theorem 4.3 is proved. $\qquad\square$

5 Proofs of main theorems

In this section, we provide the proofs of Theorem 1.4 and Theorem 1.6.

Proof of Theorem 1.4. The proof of Theorem 1.4 is totally based on Theorem 1.5 of [14] and our Theorem 4.3. Let $W^{\mathbf{v}}$ and $W^{\mathbf{w}}$ be two independent sample correlation matrix satisfying $\mathbf{C}_1$. We claim that there is an $\varepsilon > 0$ and $\delta > 0$ such that for any real number s (which may depend on p) one has

$$\begin{aligned}
\mathbb{P}^{\mathbf{v}}(p^{2/3}(\lambda_p - \lambda_+) \leq s - p^{-\varepsilon}) - p^{-\delta} &\leq \mathbb{P}^{\mathbf{w}}(p^{2/3}(\lambda_p - \lambda_+) \leq s) \\
&\leq \mathbb{P}^{\mathbf{v}}(p^{2/3}(\lambda_p - \lambda_+) \leq s + p^{-\varepsilon}) + p^{-\delta} \quad (5.1)
\end{aligned}$$

for $p \geq p_0$ sufficiently large, where p_0 is independent of s. The proof of (5.1) is independent of the matrix model and totally based on Theorem 1.5 of [14] and our Theorem 4.3, we refer to the proof of Theorem 1.7 of [14] for details.

Now if we choose $W^{\mathbf{v}}$ to be the Bernoulli case, it is not difficult to get Theorem 1.4 by combining (5.1) and Theorem 1.3. $\qquad\square$

Proof of Theorem 1.6. Set the matrix

$$A = \begin{pmatrix}
\frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\
\frac{1}{\sqrt{3\cdot2}} & \frac{1}{\sqrt{3\cdot2}} & -\frac{2}{\sqrt{3\cdot2}} & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & -\frac{n-1}{\sqrt{n(n-1)}}
\end{pmatrix}.$$

It is easy to see A is an orthogonal matrix. Moreover, it is elementary that

$$A(x_{i1} - \bar{x}_i, \cdots, x_{in} - \bar{x}_i)^T = (0, z_{i1}, \cdots, z_{i,n-1})^T,$$

where $z_{i1}, \cdots, z_{in-1}$ is a sequence of i.i.d $N(0,1)$ variables. Further, if we denote the vector $\mathbf{z}_i = (z_{i1}, \cdots, z_{in-1})^T$, we also have

$$||\mathbf{x}_i - \bar{x}_i||^2 = \sum_{k=1}^{n-1} z_{ik}^2 = ||\mathbf{z}_i||^2.$$

Thus one has

$$\mathcal{R} = RR^T = RA^T AR^T =: \mathcal{Z}.$$

Here

$$\mathcal{Z} = ZZ^T$$

with

$$Z = \begin{pmatrix} \frac{z_{11}}{||\mathbf{z}_1||} & \frac{z_{12}}{||\mathbf{z}_1||} & \cdots & \frac{z_{1,n-1}}{||\mathbf{z}_1||} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{z_{p1}}{||\mathbf{z}_p||} & \frac{z_{p2}}{||\mathbf{z}_p||} & \cdots & \frac{z_{p,n-1}}{||\mathbf{z}_p||} \end{pmatrix}.$$

Consequently, in the Gaussian case, $\mathcal{R}$ is also a W-type sample correlation matrix defined in (1.1) with parameters $p, n-1$. Thus by Theorem 1.4, we have

$$\frac{(n-1)\lambda_p(\mathcal{R}) - (p^{1/2} + (n-1)^{1/2})^2}{((n-1)^{1/2} + p^{1/2})(p^{-1/2} + (n-1)^{-1/2})^{1/3}} \xrightarrow{d} TW_1 \tag{5.2}$$

and

$$\frac{(n-1)\lambda_1(\mathcal{R}) - (p^{1/2} - (n-1)^{1/2})^2}{((n-1)^{1/2} - p^{1/2})(p^{-1/2} - (n-1)^{-1/2})^{1/3}} \xrightarrow{d} TW_1. \tag{5.3}$$

as $p \to \infty$. Replacing $n-1$ by n in (5.2) and (5.3), we can complete the proof of Theorem 1.6. $\qquad\square$

A   Appendix A

In this appendix we prove Lemma 3.10

Proof of Lemma 3.10. At first we prove i). Note that $W = DSD$. For W and SD^2 share the same eigenvalues, it is equivalent to prove that the eigenvalues of SD^2 are simple. We further introduce the polynomial $P_1(X)$ of $\{x_{ij}, 1 \leq i \leq p, 1 \leq j \leq n\}$ as

$$P_1(X) = \prod_{k=1}^{p} ||\mathbf{x}_k||^2.$$

It is easy to see $P_1(X)$ vanishes with zero Lebesgue measure, so we can always assume $P_1(X) \neq 0$. As a consequence, we can reduce our problem to prove the matrix

$$Q := SD^2 P_1(X)$$

has no multiple eigenvalue. Now we denote the discriminant of the characteristic polynomial of Q by $P_Q(X)$. Observe that all the entries of Q are polynomials of $\{x_{ij}, 1 \leq i \leq$

$p, 1 \leq j \leq n\}$, so $P_Q(X)$ is also a polynomial of $\{x_{ij}, 1 \leq i \leq p, 1 \leq j \leq n\}$. For the set of zeros of any non null polynomial in real variables only has zero Lebesgue measure, it suffices to prove that $P_Q(X)$ is not a null polynomial. In other words, it suffices to find a family $\{x_{ij}, 1 \leq i \leq p, 1 \leq j \leq n\}$ such that $P_Q(X) \neq 0$. It is equivalent to show that W has no multiple eigenvalue for one sample of the collection $\{x_{ij}, 1 \leq i \leq p, 1 \leq j \leq n\}$ such that $P_1(X) \neq 0$.

Now we choose the sample as

$$x_{ij} = \begin{cases} 1, & j = i \text{ or } i+1 \\ 0, & \text{others} \end{cases}$$

with $1 \leq i \leq p, 1 \leq j \leq n$. Then it is not difficult to see

$$W = \begin{pmatrix} 1 & \frac{1}{2} & & & \\ \frac{1}{2} & 1 & \frac{1}{2} & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{1}{2} & 1 & \frac{1}{2} \\ & & & \frac{1}{2} & 1 \end{pmatrix},$$

which is a Jacobi matrix with positive subdiagonal entries. Such a Jacobi matrix has simple eigenvalues, for example, see Proposition 2.40 of [5].

Next we turn to the proof of $ii)$. We use $X^{(p)}$ to denote the submatrix of X with p-th row deleted, and use $D^{(p)}$ to denote the $p-1 \times p-1$ upper left corner of D. And we set $S^{(p)} = X^{(p)} X^{(p)T}$, thus one has $W^{(p)} = D^{(p)} S^{(p)} D^{(p)}$. Similar to the proof of $i)$, we can prove that $SD^2 P_1(X)$ and $S^{(p)}(D^{(p)})^2 P_1(X)$ have no eigenvalue in common instead. It is easy to see the resultant of the characteristic polynomials of $SD^2 P_1(X)$ and $S^{(p)}(D^{(p)})^2 P_1(X)$ is a polynomial of $\{x_{ij}, 1 \leq i \leq p, 1 \leq j \leq n\}$. Therefore, it suffices to show the resultant is a non null polynomial. Equivalently, we shall provide a sample of $\{x_{ij}, 1 \leq i \leq p, 1 \leq j \leq n\}$ such that W and $W^{(p)}$ have no eigenvalue in common.

Using $i)$ to $W^{(p)}$ we can denote the ordered eigenvalues of $W^{(p)}$ by $\lambda_1^{(p)} < \lambda_2^{(p)} < \cdots < \lambda_{p-1}^{(p)}$. By Cauchy's interlacing property, one has

$$0 \leq \lambda_1 \leq \lambda_1^{(p)} \leq \lambda_2 \leq \cdots \leq \lambda_{p-1}^{(n)} \leq \lambda_p. \tag{A.1}$$

Moreover, we know that $\mathcal{W}^{(p)}$ shares the same nonzero eigenvalues with $W^{(p)}$. So we can provide an example such that $\mathcal{W}$ and $\mathcal{W}^{(p)}$ have no nonzero eigenvalue in common instead. Note

$$\mathcal{W} = \mathcal{W}^{(p)} + \mathbf{y}_p \mathbf{y}_p^T. \tag{A.2}$$

Taking trace on both side of (A.2), we obtain

$$\lambda_1 + \cdots + \lambda_p = \lambda_1^{(n)} + \cdots + \lambda_{p-1}^{(n)} + 1. \tag{A.3}$$

Now if we fix $\{x_{ij}, 1 \leq i \leq p-1, 1 \leq j \leq n\}$ such that $\lambda_1^{(p)} < \lambda_2^{(p)} < \cdots < \lambda_{p-1}^{(p)}$ and let $\{x_{pj}, 1 \leq j \leq n\}$ vary. When $\{x_{pj}, 1 \leq j \leq n\}$ runs through the set $\mathbb{R}^n$, the ordered nonzero eigenvalues of $\mathcal{W}$ describe the set of families $\lambda_1, \cdots, \lambda_p$ of real numbers obeying (A.1) and (A.3), see the proof of Lemma 11.2 of [3] for example. Thus it is easy to find a family $\lambda_1, \cdots, \lambda_p$ such that

$$\{\lambda_1, \cdots, \lambda_p\} \cap \{\lambda_1^{(p)}, \lambda_2^{(p)}, \lambda_{p-1}^{(p)}\} = \emptyset$$

Now we prove $iii)$. We set $X_{(n)}$ to be the submatrix of X with the n-th column deleted and set

$$\widehat{D}_{(n)} = \begin{pmatrix} \frac{\sqrt{n}}{||\widehat{\mathbf{x}}_1||} & & \\ & \ddots & \\ & & \frac{\sqrt{n}}{||\widehat{\mathbf{x}}_p||} \end{pmatrix}.$$

Let $S_{(n)} = X_{(n)}X_{(n)}^T$. It is obvious that $S_{(n)}\widehat{D}_{(n)}^2$ shares the same eigenvalues with $\widehat{W}_{(n)}$ Now we introduce the polynomials

$$P_2(X) = \prod_{k=1}^{p} ||\mathbf{x}_k||^2 \cdot ||\widehat{\mathbf{x}}_k||^2.$$

To prove that W and $\widehat{W}_{(n)}$ have no eigenvalue in common, we only need to show SD^2 and $S_{(n)}\widehat{D}_{(n)}^2$ have no eigenvalue in common. Moreover, if $P_2(X)$ does not vanish, it is equivalent to prove that the matrices $T := SD^2 P_2(X)$ and $\widehat{T}_{(n)} := S_{(n)}\widehat{D}_{(n)}^2 P_2(X)$ have no eigenvalue in common. Note that the event $P_2(X) = 0$ has zero Lebesgue measure. What's more, it is not difficult to see the entries of T and $\widehat{T}_{(n)}$ are all polynomials of the elements of X, thus the resultant $R(X)$ of the characteristic polynomials of T and $T_{(n)}$ is also a polynomial of the elements of X. Therefore, we only need to show $R(X)$ is a non null polynomial, it suffices to give only one example of X such that W and $\widehat{W}_{(n)}$ do not have eigenvalue in common. For example, we can choose

$$x_{ij} = \begin{cases} 1, & j = i \text{ or } j = n, \\ 0, & \text{others} \end{cases}$$

Then we have $\widehat{W}_{(n)} = I_p$ and

$$W = \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{2} \\ \frac{1}{2} & 1 & \cdots & \frac{1}{2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2} & \frac{1}{2} & \cdots & 1 \end{pmatrix}.$$

Thus it is easy to see $\widehat{W}_{(n)}$ and W have no eigenvalue in common for $\det(W - I) \neq 0$, which implies that $R(X)$ is not a null polynomial, so we conclude the proof. □

B Appendix B

In this appendix, we prove Lemma 3.12. If we denote

$$\hat{h}_n = \left(\frac{x_{1n}}{||\widehat{\mathbf{x}}_1||}, \frac{x_{2n}}{||\widehat{\mathbf{x}}_2||}, \cdots, \frac{x_{pn}}{||\widehat{\mathbf{x}}_p||} \right)^T.$$

Set

$$c_i = \frac{1}{||\mathbf{x}_i|| \cdot ||\widehat{\mathbf{x}}_i|| \cdot (||\mathbf{x}_i|| + ||\widehat{\mathbf{x}}_i||)}.$$

By the Hoeffding inequality, we have

$$c_i = \frac{1}{n^{3/2}} + \frac{K^{O(1)} \log^{O(1)} n}{n^2}. \tag{B.1}$$

holds with overwhelming probability. It is not difficult to see

$$\hat{v}_j \cdot h_n = \hat{v}_j \cdot \hat{h}_n - \hat{v}_j \cdot (c_1 x_{1n}^3, \cdots, c_p x_{pn}^3)^T := \hat{v}_j \cdot \hat{h}_n + d_j. \tag{B.2}$$

By (B.1), we can write d_j

$$d_j := \frac{1}{n^{3/2}} \hat{v}_j \cdot (x_{1n}^3, \cdots, x_{pn}^3)^T + f_j. \tag{B.3}$$

Observe that

$$\sum_{j \in J} (\hat{v}_j \cdot h_n)^2 = \sum_{j \in J} (\hat{v}_j \cdot \hat{h}_n)^2 + 2 \sum_{j \in J} d_j (\hat{v}_j \cdot \hat{h}_n) + \sum_{j \in J} d_j^2.$$

Since $(x_{1n}^3, \cdots, x_{pn}^3)^T$ is also a random vector with mean zero and finite variance entries, Lemma 2.5 can be used to the first part of the right hand side of (B.3). Thus if we set the projection

$$P_J = \sum_{j \in J} \hat{v}_j \hat{v}_j^T, \tag{B.4}$$

then we have

$$\sum_{j \in J} d_j^2 \leq C \frac{1}{n^3} |P_J \cdot (x_{1n}^3, \cdots, x_{pn}^3)^T|^2 + C \sum_{j \in J} f_j^2 = O(\frac{|J|}{n^3}) + O(\frac{K^{O(1)} \log^{O(1)} n}{n^3}).$$

with overwhelming probability. Here we have used the fact that for any J

$$\sum_{j \in J} f_j^2 \leq C \frac{K^{O(1)} \log^{O(1)} n}{n^4} \sum_{i=1}^{p} x_{in}^6 = O(\frac{K^{O(1)} \log^{O(1)} n}{n^3})$$

with overwhelming probability. Since

$$\sum_{j \in J} d_j (\hat{v}_j \cdot \hat{h}_n) \leq \left(\sum_{j \in J} d_j^2 \right)^{1/2} \left(\sum_{j \in J} (\hat{v}_j \cdot \hat{h}_n)^2 \right)^{1/2},$$

it suffices to prove the following lemma instead.

Lemma B.1. *Using the notation in Lemma 3.12, we have for any $J \in \{1, \cdots, p\}$ with $|J| = d \leq nK^{-3}$,*

$$\sqrt{n} [\sum_{j \in J} (\hat{v}_j \cdot \hat{h}_n)^2]^{1/2} = \sqrt{d} + O(K \log n)$$

with overwhelming probability.

Proof. Observe that

$$\hat{v}_j \cdot \hat{h}_n = (\frac{\hat{v}_{j1}}{||\hat{\mathbf{x}}_1||}, \frac{\hat{v}_{j2}}{||\hat{\mathbf{x}}_2||}, \cdots, \frac{\hat{v}_{jp}}{||\hat{\mathbf{x}}_p||})^T \cdot (x_{1n}, \cdots, x_{pn})^T.$$

Now we set

$$\tilde{v}_j = \sqrt{n} (\frac{\hat{v}_{j1}}{||\hat{\mathbf{x}}_1||}, \frac{\hat{v}_{j2}}{||\hat{\mathbf{x}}_2||}, \cdots, \frac{\hat{v}_{jp}}{||\hat{\mathbf{x}}_p||})^T$$

and

$$\tilde{h}_n = (x_{1n}, \cdots, x_{pn})^T.$$

It follows that

$$\hat{v}_j \cdot \hat{h}_n = \frac{1}{\sqrt{n}} \tilde{v}_j \cdot \tilde{h}_n.$$

We use the following concentration theorem, which is a consequence of Talagrand's inequality, (see Theorem 69 of [16]).

Theorem B.2 (Talagrand's inequality). *. Let $\mathbf{D}$ be the disk $\{z \in \mathbb{C}, |z| \leq K\}$. For every product probability μ on $\mathbf{D}^p$, every convex 1-Lipschitz function $F : \mathbb{C}^p \to \mathbb{R}$, and every $r \geq 0$,*

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

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

Remark B.3. *In fact, here we only need the real case of the theorem.*

It is easy to see

$$\sqrt{n}[\sum_{j \in J}(\hat{v}_j \cdot \hat{h}_n)^2]^{1/2} = [\sum_{j \in J}(\tilde{v}_j \cdot \tilde{h}_n)^2]^{1/2} =: F(\tilde{h}_n)$$

is a convex function of the vector $\tilde{h}_n$. Note

$$\frac{|F(\tilde{h}'_n) - F(\tilde{h}_n)|}{||\tilde{h}'_n - \tilde{h}_n||} = \frac{|F(\tilde{h}'_n) - F(\tilde{h}_n)|}{||\sqrt{n}\hat{h}'_n - \sqrt{n}\hat{h}_n||} \cdot \frac{||\sqrt{n}\hat{h}'_n - \sqrt{n}\hat{h}_n||}{||\tilde{h}'_n - \tilde{h}_n||},$$

where

$$\hat{h}'_n = (\frac{x'_{1n}}{||\widehat{\mathbf{x}}_1||}, \cdots, \frac{x'_{pn}}{||\widehat{\mathbf{x}}_p||})^T, \quad \tilde{h}'_n = (x'_{1n}, \cdots, x'_{pn})^T.$$

Since $F(\tilde{h}_n)$ is the norm of a projection of the vector $\sqrt{n}\hat{h}_n$, it is always 1-Lipschitz with respect to $\sqrt{n}\hat{h}_n$. And by the Hoeffding inequality, we also have

$$\frac{||\sqrt{n}\hat{h}'_n - \sqrt{n}\hat{h}_n||}{||\tilde{h}'_n - \tilde{h}_n||} \leq 2$$

with overwhelming probability. So $F(\tilde{h}_n)$ is a 2-Lipschitz function with overwhelming probability. Thus we can always consider $F(\tilde{h}_n)$ as a 2-Lipschitz function below. By Theorem B.2, we have

$$\mathbb{P}(|F(\tilde{h}_n) - M(F(\tilde{h}_n))| \geq r) \leq 4\exp(-r^2/64K^2).$$

So to conclude the proof of Lemma B.1, we only need to show that

$$|M(F(\tilde{h}_n)) - \sqrt{d}| \leq 2K.$$

Note that

$$\sqrt{n}\hat{h}_n = \begin{pmatrix} \frac{\sqrt{n}}{||\widehat{\mathbf{x}}_1||} & & \\ & \ddots & \\ & & \frac{\sqrt{n}}{||\widehat{\mathbf{x}}_p||} \end{pmatrix} \tilde{h}_n := \hat{D}\tilde{h}_n$$

So we have $F^2(\tilde{h}_n) = n\sum_{j \in J}|\hat{v}_j \cdot \hat{h}_n|^2 = \sum_{j \in J}\tilde{h}_n^T \hat{D}\hat{v}_j\hat{v}_j^T\hat{D}\tilde{h}_n := \tilde{h}_n^T \hat{D}P_J\hat{D}\tilde{h}_n$, where P_J is the projection defined in (B.4). Let $\hat{D}P_J\hat{D} =: (m_{kl})_{1 \leq k,l \leq p}$, then we have

$$F^2(\tilde{h}_n) = \sum_{1 \leq k,l \leq p} m_{kl}x_{kn}x_{ln} = \sum_{k=1}^p m_{kk}x_{kn}^2 + \sum_{1 \leq k \neq l \leq p} m_{kl}x_{kn}x_{ln}.$$

We fix all the variables except $x_{1n}, \cdots, x_{pn}$, so the probabilities and expectations are all taken with respect to $\tilde{h}_n$ below. Consider the event $\mathcal{E}_+$ that $F(\tilde{h}_n) \geq \sqrt{d} + 2K$, which implies $F^2(\tilde{h}_n) \geq d + 4\sqrt{d}K + K^2$. It follows that

$$\mathbb{P}(\mathcal{E}_+) \leq \mathbb{P}(\sum_{k=1}^p m_{kk}x_{kn}^2 \geq d + 2\sqrt{d}K) + \mathbb{P}(|\sum_{1 \leq k \neq l \leq p} m_{kl}x_{kn}x_{ln}| \geq 2\sqrt{d}K).$$

Observe that

$$\mathbb{E}\{\sum_{k=1}^{p} m_{kk}x_{kn}^2\} = \sum_{k=1}^{p} m_{kk} = d(1 + O(\frac{K^{2+\epsilon}}{\sqrt{n}}))$$

holds with overwhelming probability for any small $\epsilon > 0$. Here we have used the fact that

$$\lambda_{\min}^2(\hat{D})TrP_J \leq Tr\hat{D}P_J\hat{D} \leq \lambda_{\max}^2(\hat{D})TrP_J$$

and $TrP_J = d$. By the condition that $d \leq nK^{-3}$, we have

$$\mathbb{E}\{\sum_{k=1}^{p} m_{kk}x_{kn}^2\} = d + o(\sqrt{d}K).$$

Let $S_1 := \sum_{k=1}^{p} m_{kk}(x_{kn}^2 - 1)$. We have

$$\mathbb{P}(\sum_{k=1}^{p} m_{kk}x_{kn}^2 \geq d + 2\sqrt{d}K) \leq \mathbb{P}(|S_1| \geq \sqrt{d}K) \leq \frac{\mathbb{E}(|S_1|^2)}{dK^2}.$$

And by the assumption on K we also have

$$\mathbb{E}|S_1|^2 = \sum_{k=1}^{p} m_{kk}^2\mathbb{E}(x_{kn}^2 - 1)^2 = \sum_{k=1}^{p} m_{kk}^2(\mathbb{E}x_{kn}^4 - 1) \leq dK.$$

Thus,

$$\mathbb{P}(|S_1| \geq \sqrt{d}K) \leq \frac{\mathbb{E}(|S_1|^2)}{dK^2} \leq \frac{1}{K} \leq 1/10.$$

Set $S_2 := |\sum_{k \neq l} m_{kl}x_{kn}x_{ln}|$. Then we have

$$\mathbb{E}S_2^2 = 2\sum_{k \neq l} m_{kl}^2 \leq 2Tr\hat{D}P_J\hat{D}^2P_J\hat{D} \leq ||\hat{D}||_{op}^4 TrP_J = 2d(1 + O(\frac{K^{2+\epsilon}}{\sqrt{n}})).$$

By Chebyshev's inequality one has

$$\mathbb{P}(S_2 \geq 2\sqrt{d}K) \leq 1/10.$$

Similarly, we can define $\mathcal{E}_-$ as the event $F(\tilde{h}_n) \leq \sqrt{d} - 2K$ and use

$$\mathbb{P}(\mathcal{E}_-) \leq \mathbb{P}(S_1 \leq d - \sqrt{d}K) + \mathbb{P}(S_2 \geq \sqrt{d}K).$$

Both terms on the right hand side can be bounded by $1/5$ by the same argument as above. So we conclude the proof. $\qquad\square$

References

[1] Bai, Z. D. and Silverstein, J. W.: Spectral analysis of large dimensional random matrices. Mathematics Monograph Series 2, Science Press, Beijing, 2000.

[2] Bai, Z. D. and Zhou, W.: Large sample covariance matrices without independence structure in columns. *Statistica Sinica*. **18**, (2008), 425-442. MR-2411613

[3] Benaych-Georges, F. Guionnet, A. and Maida, M.: Large deviations of the extreme eigenvalues of random deformations of matrices. Probab. Theory and Related Fields, (2011), DOI:10.1007/s00440-011-0382-3.

[4] Borodin, A. and Forrester, P. J.: Increasing subsequences and the hard-to-soft edge transition in matrix ensembles. *Journal of Physics A: Math. and Gene.*, **36**, (2003), 2963-2981. MR-1986402

[5] Deift, P.: Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach, volume 3 of Courant Lecture Notes in Mathematics. New York Universality Courant Instituite of Mathematical Sciences, New York, NY, 1999. MR-1677884

[6] Erdös, L., Schlein, B. and Yau, H-T.: Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Annals of Probab.*, **37**, (2009), 815-852. MR-2537522

[7] Erdös, L., Schlein, B. and Yau, H-T.: Wegner estimate and level repulsion for Wigner random matrices. *Int. Math. Res. Not.* , **3**, (2010), 436-479. MR-2587574

[8] Erdös, L., Schlein, B., Yau, H-T. and Yin, J.: The local relaxation flow approach to universality of the local statistics for random matrices. Ann. Inst. H. Poincaré Probab. Statist., **48**, (2012), 1-46. MR-2919197

[9] Erdös, L., Yau, H-T. and Yin, J.: Bulk universality for generalized Wigner matrices. arXiv:1001.3453.

[10] Feldheim, O.N. and Sodin, S.: A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geometric And Functional Analysis*, **20**, (2010) , 88-123. MR-2647136

[11] Jiang, T. F.: The Limiting Distributions of Eigenvalues of Sample Correlation Matrices. *Sankhyā*, **66**, (2004), 35-48. MR-2082906

[12] Johnstone, I.: On the distribution of the largest eigenvalue in principal component analysis. *Ann. Stat.* **29**, (2001), 295-327. MR-1863961

[13] Péché, S.: Universality results for largest eigenvalues of some sample covariance matrix ensemble. *Probab. Theorey Related Fields*, **143**, (2009), 481-516. MR-2475670

[14] Pillai, N. S. and Yin, J.: Universality of covariance matrices . arXiv:1110.2501.

[15] Soshnikov, A.: A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist, Phys.*, **108**, (2002), 1033-1056. MR-1933444

[16] Tao, T. and Vu, V.: Random matrices: Universality of local eigenvalue statistics. *Acta Mathematica*, **206**, (2010), 127-204. MR-2784665

[17] Tao, T. and Vu, V.: Random matrices: Universality of local eigenvalue statistics up to the edge. *Communications in Mathematical Physics*, **298**, (2010), 549-572. MR-2669449

[18] Tao, T. and Vu, V.: Random covariance matrices: Universality of local statistics of eigenvalues. Ann. Probab. **40**, (2012), 1285-1315. MR-2962092

[19] Tracy, C. A. and Widom, H.: On orthogonal and symplectic matrix ensembles. *Commun. Math. Phys.*, **177**, (1996), 727-754. MR-1385083

[20] Wang, K.: Random covariance matrices: Universality of local statistics of eigenvalues up to edge. Random Matrices: Theory and Application, **01**, (2012), 1150005.