Tracy-Widom limit for Spearman's rho

Zhigang Bao*
Hong Kong University of Science and Technology
mazgbao@ust.hk

In this paper, we study the Spearman rank correlation matrix, which is a random matrix model from the non-parametric statistics. We focus on the high dimensional scenario when n is proportional to p. In the null case, we show that the Tracy-Widom law holds for the largest eigenvalues of the Spearman rank correlation matrix. The proof is based on a general strategy for the universality of the covariance type matrices from Pillai and Yin [17].

Date: December 18, 2017
Keywords: largest eigenvalue, non-parametric statistics, Spearman's ρ, Tracy-Widom law, random matrices

1. Introduction.

1.1. *Matrix model and main results.* Let $\boldsymbol{w} = (w_1, \ldots, w_p)$ be a p-dimensional random vector with independent but may not be identically distributed components. We further assume that w_i's are all continuous random variables. Let $\boldsymbol{w}_j = (w_{1j}, \ldots, w_{pj})', j \in [\![1, n]\!]$ be n i.i.d. samples of $\boldsymbol{w}$. Hereafter we use the notation $[\![a, b]\!] := [a, b] \cap \mathbb{Z}$. We then call $W = (w_{ij})_{p,n}$ the data matrix. In this paper, we focus on the setting when n are p are comparably large, i.e.,

$$p = p(n), \qquad c_n := \frac{p}{n} \to c \in (0, \infty), \qquad \text{if} \quad n \to \infty, \tag{1.1}$$

for some positive constant c.

We then construct the corresponding Spearman rank correlation matrix from the data matrix W as follows. For each fixed $i \in [\![1, p]\!]$, we can rank n samples $w_{i1}, \ldots, w_{in}$ according to their size. Let q_{ij} be the rank of w_{ij} among $w_{i1}, \ldots, w_{in}$. Observe that for each $i \in [\![1, p]\!]$, the random vector $(q_{i1}, \ldots, q_{in})$ is a random

*The author is partially supported by Hong Kong RGC grant ECS 26301517

permutation uniformly distributed on $\mathcal{S}_n$. Here $\mathcal{S}_n$ is the symmetric group of the set $\{1,2,\ldots,n\}$. Next, we normalize q_{ij}'s as

$$y_{ij} := \sqrt{\frac{12}{n^2-1}}\big(q_{ij} - \frac{n+1}{2}\big), \qquad (i,j) \in [\![1,p]\!] \times [\![1,n]\!].$$

We further set $Y = (y_{ij})_{p,n}$ as the matrix of rank. Observe that by the assumption on the independence of the components of $\boldsymbol{w}$, we see that the p rows of Y are i.i.d random vectors. It is also easy to check for any $i \in [\![1,p]\!]$, there are

$$\mathbb{E}y_{ij} = 0, \qquad \mathbb{E}y_{ij}^2 = 1, \qquad \mathbb{E}y_{ij}y_{ik} = -\frac{1}{n-1}, \quad \forall j \neq k. \tag{1.2}$$

We then do the scaling

$$X = \frac{1}{\sqrt{n}}Y, \tag{1.3}$$

and denote by $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ the i-th rows of X and Y, respectively. The Spearman rank correlation matrix is defined as

$$S \equiv S_n := XX' = \frac{1}{n}YY'. \tag{1.4}$$

Observe that the matrix entry S_{ab} is the Spearman rank correlation coefficient of the ranks of the samples of w_a and those of w_b. Hence, the matrix S is a natural multivariate extension of the Spearman rank correlation coefficient.

Since Marchenko and Pastur discovered the global spectral distribution (MP law) in their seminal work[15], there has been a vast of literature devoted to the spectral property of the large dimensional sample covariance matrix and its varieties. Especially, on the largest eigenvalue, Johnstone [12] proved the Tracy-Widom law (TW law) in the null case, i.e., the population covariance matrix is I_p. The TW law was then shown to be universal for sample covariance matrices in the null case, even under more general distribution assumptions, see [18, 17]. Later on, the Tracy-Widom law was further extended to more general population assumption, see [5, 14, 13]. In [4, 16], it was also shown that the TW law holds for the sample correlation matrix under in the null case.

Although many spectral statistics of the sample covariance matrices and correlation matrices turn out to be extremely useful for various statistical inference problems, these two matrix models are both parametric. Consequently, certain parametric assumption such as the moment condition is needed for the limiting theorem on the spectral statistics. For instance, on the TW law for the covariance matrices, we refer to [8] for a necessary moment assumption. Moreover, limiting results such as TW law are very often used to test the independence of the components of the population random vector $\boldsymbol{w}$. Mathematically, the idea is valid only for the Gaussian vectors. For general distribution, covariance matrix only contains the information on correlation rather than dependence. Due to the above reasons, it is very natural to consider the limiting spectral properties of the nonparametric random matrix models. Among the others, the Spearman rank correlation and Kendall rank correlation matrices are probably the most important and natural ones. However, the study of these multivariate non-parametric models under the high-dimension setting is much less, in contrast to the parametric ones. So far, there are a only a handful of results on this direction. The global spectral distributions for the Spearman rank correlation matrix and Kendall rank correlation matrix have been derived in [1] and [2], respectively. A CLT for the linear eigenvalue statistics of the Spearman rank correlation matrix has been established in [6]. On the local scale, we proved the TW law for the Kendall rank correlation matrices in [3] recently. It is the first TW law for a non-parametric matrix model, and is also the first TW law for a high-dimensional U-statistics. In this paper, our aim is to derive the companion result (TW law) for the Spearman rank correlation matrix.

Before stating our main results, we first recall the global spectral property of S from [1]. Let $\lambda_1(S) \geq \ldots \geq \lambda_p(S)$ be p ordered eigenvalue of S. We denote the empirical spectral distributions (ESD) of S by

$$F_n := \frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i(S)}.$$

In [1], it is proved that F_n is asymptotically given by the standard MP law. More specifically, we have the following theorem.

THEOREM 1.1 (Theorem 2.2 of [1]). *Under the assumption (1.1), we have that almost surely F_n converges weakly to F_c whose density is given by*

$$\rho_c(x) = \frac{1}{2\pi c}\frac{\sqrt{(d_{+,c}-x)(x-d_{-,c})}}{x}\mathbf{1}(d_{-,c} \leq x \leq d_{+,c})$$

where

$$d_{\pm,c} = (1 \pm \sqrt{c})^2.$$

In case $c > 1$, in addition, F_c has a singular part: a point mass $(1-c^{-1})\delta_0$.

Further, replacing c be c_n, we denote by ρ_{c_n}, F_{c_n}, $d_{\pm,c_n}$ the analogues of ρ_c, F_c, $d_{\pm,c}$, respectively.

To state our main results, we further denote by $Q := \frac{1}{n}\mathcal{X}\mathcal{X}'$ a Wishart matrix, where $\mathcal{X}$ is $p \times n$ data matrix with i.i.d. $N(0,1)$ variables. Let $\lambda_i(Q)$ be the i-th largest eigenvalue of Q.

Our main result is the following theorem.

Theorem 1.2 (Edge universality of Spearman rank correlation matrix). *Suppose that the assumption (1.1) holds. There exist positive constants ε and δ such that for any $s \in \mathbb{R}$,*

$$\mathbb{P}\Big(n^{\frac{2}{3}}(\lambda_1(Q) - d_{+,c_n}) \le s - n^{-\varepsilon}\Big) - n^{-\delta} \le \mathbb{P}\Big(n^{\frac{2}{3}}(\lambda_1(S) - d_{+,c_n}) \le s\Big) \\ \le \mathbb{P}\Big(n^{\frac{2}{3}}(\lambda_1(Q) - d_{+,c_n}) \le s + n^{-\varepsilon}\Big) + n^{-\delta}$$

holds when n is sufficiently large.

Remark 1.3. The above result can be generalized to the joint distribution of the first few eigenvalues. More specifically, there exist positive constants ε and δ such that for any fixed positive integer k and any $s_1, \dots, s_k \in \mathbb{R}$,

$$\mathbb{P}\Big(n^{\frac{2}{3}}(\lambda_1(Q) - d_{+,c_n}) \le s_1 - n^{-\varepsilon}, \dots, n^{\frac{2}{3}}(\lambda_k(Q) - d_{+,c_n}) \le s_k - n^{-\varepsilon}\Big) - n^{-\delta} \\ \le \mathbb{P}\Big(n^{\frac{2}{3}}(\lambda_1(S) - d_{+,c_n}) \le s_1, \dots, n^{\frac{2}{3}}(\lambda_k(S) - d_{+,c_n}) \le s_k\Big) \\ \le \mathbb{P}\Big(n^{\frac{2}{3}}(\lambda_1(Q) - d_{+,c_n}) \le s_1 + n^{-\varepsilon}, \dots, n^{\frac{2}{3}}(\lambda_k(Q) - d_{+,c_n}) \le s_k + n^{-\varepsilon}\Big) + n^{-\delta}$$

holds when n is sufficiently large. We refer to Remark 1.4 of [17] for a similar extension for the sample covariance matrix. The extension here can be proved in the same way.

From Theorem 1.2, we have the following corrolary on the largest eigenvalue.

Corollary 1.4 (Tracy-Widom law for $\lambda_1(S)$). *Under the assumption of Theorem 1.2, we have*

$$n^{\frac{2}{3}} c_n^{\frac{1}{6}} d_{+,c_n}^{-\frac{2}{3}} \big(\lambda_1(S) - d_{+,c_n}\big) \Longrightarrow \mathrm{TW}_1$$

1.2. *Proof strategy.* The proof of Theorem 1.2 will be done with the aid of a general strategy in Pillai and Yin [17] for the covariance type matrices, which itself is an adaptation of the method originally raised in [11] by Erdös, Yau and Yin for Wigner matrices. Roughly speaking, to prove the TW law for the largest eigenvalue, first one needs to prove a local law for the spectral distribution, which controls the location of the eigenvalues on an optimal local scale. Second, with the aid of the local law, one needs to perform a Green function comparison between the matrix of interest and certain reference matrix ensemble, whose edge spectral behavior is already known. In [17], an extended criterion of the local law for covariance type of matrices with independent columns (or rows) was given, see Theorem 3.6 of [17]. It allows one to relax the independence assumption on the entries within one columns (or rows) to certain extent, as long as some large deviation estimates hold for certain linear and quadratic forms of each column (or row) of the data matrix, see Lemma 3.4 of [17]. This general criterion was then used in [16] and [4] to establish the edge universality of the sample correlation matrices.

For the Spearman rank correlation matrix $S = XX'$ defined in (1.4), our main task is to show a large deviation estimate (c.f. Proposition 2.1) for each row of the matrix X. Once Proposition 2.1 is established, we can use the criterion in Theorem 3.6 of [17] to conclude the local law of S. It turns out that the Green function comparison part can be done similarly to that in [16] for the Pearson's correlation matrix, by choosing an appropriate reference matrix ensemble. The reference matrix to be chosen is this work turns out to be the traditional sample covariance matrix, which is subtracted by the sample mean and divided by $n-1$.

1.3. *Notation and organization.*

1.3.1. *Notation.* We need the following definition on high-probability estimates from [9].

DEFINITION 1.5. *Let $\mathcal{X} \equiv \mathcal{X}^{(N)}$ and $\mathcal{Y} \equiv \mathcal{Y}^{(N)}$ be two sequences of nonnegative random variables. We say that $\mathcal{Y}$ stochastically dominates $\mathcal{X}$ if, for all (small) $\epsilon > 0$ and (large) $D > 0$,*

$$\mathbb{P}\big(\mathcal{X}^{(N)} > N^{\epsilon}\mathcal{Y}^{(N)}\big) \le N^{-D}, \tag{1.5}$$

for sufficiently large $N \ge N_0(\epsilon, D)$, and we write $\mathcal{X} \prec \mathcal{Y}$ or $\mathcal{X} = O_{\prec}(\mathcal{Y})$. When $\mathcal{X}^{(N)}$ and $\mathcal{Y}^{(N)}$ depend on a parameter $v \in \mathcal{V}$ (typically an index label or a spectral parameter), then $\mathcal{X}(v) \prec \mathcal{Y}(v)$, uniformly in $v \in \mathcal{V}$, means that the threshold $N_0(\epsilon, D)$ can be chosen independently of v.

We use the symbols $O(\,\cdot\,)$ and $o(\,\cdot\,)$ for the standard big-O and little-o notation. We use c and C to denote strictly positive constants that do not depend on N. Their values may change from line to line. For any matrix A, we denote by $\|A\|$ its operator norm, while for any vector $\boldsymbol{a}$, we use $\|\boldsymbol{a}\|$ to denote its 2-norm. The matrix entries of A are denoted by A_{ij}. In addition, we use double brackets to denote index sets, *i.e.*, for $n_1, n_2 \in \mathbb{R}$, $[\![n_1, n_2]\!] := [n_1, n_2] \cap \mathbb{Z}$. We also use $\mathbf{1}$ to represent the all-one vector, whose dimension may be changed from one to another.

1.3.2. *Organization.* The paper is organized as follows: In Section 2, we will prove some large deviation estimates for certain linear and quadratic forms of $\boldsymbol{x}_i$'s, and then briefly state the proof of the local law of S based on Theorem 3.6 of [17]. In Section 3, we perform the Green function comparison and then prove our main results.

2. Local law of S. In this section, our final goal is to prove a strong local law for the matrix S: Proposition 2.3. To this end, we shall first establish some large deviation estimates for certain linear and quadratic forms of $\boldsymbol{x}_i$'s, which are the rows of the matrix X defined in (1.3). Then Proposition 2.3 will follows from these large deviation estimates and Theorem 3.6 of [17].

2.1. *Large deviation estimates for $\boldsymbol{x}_i$.* Let

$$s_i = \sqrt{\frac{12}{n(n^2-1)}}\left(i - \frac{n+1}{2}\right), \qquad i \in [\![1, n]\!].$$

We set the vector

$$\boldsymbol{s} := (s_1, \ldots, s_n). \tag{2.1}$$

Let $\boldsymbol{x}$ be a random permutation which is uniformly distributed on the symmetric group of $\boldsymbol{s}$. We can then regard $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ as i.i.d. copies of $\boldsymbol{x}$. Further, we let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ be a random vector with i.i.d. components which are uniformly distributed on the set $\boldsymbol{s}$, i.e., $\mathbb{P}(\xi_j = s_i) = \frac{1}{n}, i \in [\![1, n]\!], j \in [\![1, n]\!]$. Let $\boldsymbol{\xi}_i = (\xi_{ij})_{j=1}^n, i \in [\![1, p]\!]$ be i.i.d. copies of $\boldsymbol{\xi}$. Note that x_{ij}'s are also identically distributed as ξ_j, but x_{ij}'s are correlated. We further set

$$\tilde{\boldsymbol{x}} := \boldsymbol{\xi}\Sigma^{\frac{1}{2}}, \qquad \tilde{\boldsymbol{x}}_i := \boldsymbol{\xi}_i\Sigma^{\frac{1}{2}}, \tag{2.2}$$

where

$$\Sigma = \frac{n}{n-1}I_n - \frac{1}{n-1}\mathbf{1}\mathbf{1}'. \tag{2.3}$$

Here $\mathbf{1}$ represents the n-dimensional all-one vector. Let $\widetilde{X}$ and Ξ be the $p \times n$ matrix with $\tilde{\boldsymbol{x}}_i$ and $\boldsymbol{\xi}_i$ as the i-th row, respectively. We further denote by

$$\widetilde{S} := \widetilde{X}\widetilde{X}' = \Xi\Sigma\Xi'. \tag{2.4}$$

Observe that $\widetilde{S}$ is the classical sample covariance matrix in statistics theory, which is subtracted by sample mean and normalized by $n-1$, although in random matrix theory the simplified model $\Xi\Xi'$ is considered more often. Since $\Xi\Sigma\Xi'$ is just a rank one perturbation of $\frac{n}{n-1}\Xi\Xi'$, it is known that almost surely the empirical spectral distribution of $\widetilde{S}$ also converges weakly to F_c.

Below is a collection of large deviation estimates on $\boldsymbol{x}_i$'s, and also $\tilde{\boldsymbol{x}}_i$'s.

PROPOSITION 2.1. *Let $\boldsymbol{x}_i$'s be defined as (1.3). For any deterministic vector $\boldsymbol{a} = (a_j) \in \mathbb{C}^n$ and matrix $B = (b_{ij}) \in \mathbb{C}^{n\times n}$, we have*

$$\left|\boldsymbol{x}_i\boldsymbol{a}'\right| \prec \sqrt{\frac{\|\boldsymbol{a}\|^2}{n}}, \tag{2.5}$$

$$\left|\boldsymbol{x}_iB\boldsymbol{x}_i' - \frac{1}{n}\mathrm{Tr}B\Sigma\right| \prec \frac{1}{n}\sqrt{\mathrm{Tr}|B|^2}. \tag{2.6}$$

The same inequalities hold if we replace $\boldsymbol{x}_i$ by $\tilde{\boldsymbol{x}}_i$ (c.f. (2.2)).

Proof of Proposition 2.1. The proof relies on a martingale concentration argument. We start with (2.5). Recall the random vector $\boldsymbol{x} = (x_1, \dots, x_n)$ which is a random permutation uniformly distributed on the symmetric group of $\boldsymbol{s}$ (c.f. (2.1)). Since $\boldsymbol{x}_i$'s are i.i.d. copies of $\boldsymbol{x}$, it suffices to prove all the results in Proposition 2.1 with $\boldsymbol{x}_i$ replaced by $\boldsymbol{x}$ for brevity.

We first construct a martingale difference sequence. Define the filtration

$$\mathcal{F}_0 = \emptyset, \qquad \mathcal{F}_\ell := \sigma\{x_1, \dots, x_\ell\}, \qquad \ell \in [\![1, n]\!], \tag{2.7}$$

and set

$$M_\ell = \sum_{j=1}^{n} a_j \Big(\mathbb{E}\big(x_j | \mathcal{F}_\ell\big) - \mathbb{E}\big(x_j | \mathcal{F}_{\ell-1}\big) \Big) \tag{2.8}$$

It is clear that, conditioning on $\mathcal{F}_\ell$, x_j is uniformly distributed on the set $\boldsymbol{s} \setminus \{x_1, \dots, x_\ell\}$ for all $j \in [\![\ell+1, n]\!]$. Hence, we have

$$\mathbb{E}(x_j | \mathcal{F}_\ell) = -\frac{1}{n-\ell} \sum_{k=1}^{\ell} x_k, \qquad j \in [\![\ell+1, n]\!] \tag{2.9}$$

where we used the fact that $\sum_{i=1}^n x_i = \sum_{i=1}^n s_i = 0$. From the definition in (2.8), it is easy to check that $M_n = 0$ since $\mathcal{F}_n = \mathcal{F}_{n-1}$. In addition, we have

$$M_\ell = x_\ell a_\ell + \sum_{j=\ell+1}^{n} a_j \mathbb{E}\big(x_j | \mathcal{F}_\ell\big) - \sum_{j=\ell}^{n} a_j \mathbb{E}\big(x_j | \mathcal{F}_{\ell-1}\big).$$

$$= \Big(a_\ell - \frac{1}{n-\ell} \sum_{j=\ell+1}^{n} a_j \Big) \Big(x_\ell + \frac{1}{n-\ell+1} \sum_{k=1}^{\ell-1} x_k \Big), \qquad \ell \in [\![1, n-1]\!]. \tag{2.10}$$

Using the bound $|x_j| = O(\frac{1}{\sqrt{n}})$ and the fact

$$\Big| \sum_{k=1}^{\ell-1} x_k \Big| = \Big| \sum_{k=\ell}^{n} x_k \Big| \le \frac{C}{\sqrt{n}} \min\{\ell - 1, n - \ell + 1\}, \tag{2.11}$$

we can simply get from (2.10) that

$$|M_\ell| \le \frac{C}{\sqrt{n}} \Big(\frac{1}{n-\ell} \sum_{j=\ell+1}^{n} |a_j| + |a_\ell| \Big). \tag{2.12}$$

Applying Burkholder inequality, we have for any fixed positive integer $q \ge 2$

$$\mathbb{E} \Big| \sum_{\ell=1}^{n} M_\ell \Big|^q \le (Cq)^{\frac{3q}{2}} \mathbb{E} \Big(\sum_{\ell=1}^{n} M_\ell^2 \Big)^{\frac{q}{2}}. \tag{2.13}$$

From (2.12), we have

$$\sum_{\ell=1}^{n} M_\ell^2 = \sum_{\ell=1}^{n-1} M_\ell^2 \le \frac{C}{n}\sum_{\ell=1}^{n-1}\Big(\frac{1}{(n-\ell)^2}\Big(\sum_{j=\ell+1}^{n}|a_j|\Big)^2 + a_\ell^2\Big)$$

$$\le \frac{C}{n}\sum_{\ell=1}^{n-1}\Big(\frac{1}{n-\ell}\sum_{j=\ell+1}^{n} a_j^2 + a_\ell^2\Big) \le C\frac{\log n}{n}\sum_{\ell=1}^{n} a_j^2. \tag{2.14}$$

Plugging (2.14) into (2.13) and using Markov's inequality, we can conclude (2.5).

Next, we prove (2.6). It suffices to show the following two

$$\Big|\sum_{j=1}^{n} b_{jj}(x_{ij}^2 - \frac{1}{n})\Big| \prec \frac{1}{n}\sqrt{\sum_j (b_{jj})^2}, \tag{2.15}$$

$$\Big|\sum_{j\ne k} b_{jk}x_{ij}x_{ik} + \frac{1}{n(n-1)}\sum_{j\ne k} b_{jk}\Big| \prec \frac{1}{n}\sqrt{\sum_{j\ne k}(b_{jk})^2}. \tag{2.16}$$

For (2.15), again, we construct a sequence of martingale differences as

$$N_\ell = \sum_{j=1}^{n} b_{jj}\Big(\mathbb{E}\big(x_j^2|\mathcal{F}_\ell\big) - \mathbb{E}\big(x_j^2|\mathcal{F}_{\ell-1}\big)\Big). \tag{2.17}$$

We have $N_n = 0$ since $\mathcal{F}_n = \mathcal{F}_{n-1}$. Again, given $\{x_1, \dots, x_{\ell-1}\}$, we recall the fact that x_j is uniformly distributed on $\boldsymbol{s} \setminus \{x_1, \dots, x_{\ell-1}\}$ for all $j \ge \ell$. Moreover, since $\sum_{j=1}^{n} s_j^2 = 1$, we have

$$\mathbb{E}(x_j^2|\mathcal{F}_\ell) = \frac{1}{n-\ell}\Big(1 - \sum_{k=1}^{\ell} x_k^2\Big), \qquad \forall j \ge \ell+1. \tag{2.18}$$

Applying (2.18), we obtain

$$N_\ell = b_{\ell\ell}x_\ell^2 + \sum_{j=\ell+1}^{n} b_{jj}\mathbb{E}\big(x_j^2|\mathcal{F}_\ell\big) - \sum_{j=\ell}^{n} b_{jj}\mathbb{E}\big(x_j^2|\mathcal{F}_{\ell-1}\big).$$

$$= \Big(b_{\ell\ell} - \frac{1}{n-\ell}\sum_{j=\ell+1}^{n} b_{jj}\Big)\Big(x_\ell^2 - \frac{1}{n-\ell+1}\Big(1 - \sum_{k=1}^{\ell-1} x_k^2\Big)\Big)$$

Using the fact $x_k = O(\frac{1}{\sqrt{n}})$, we have

$$|N_\ell| \le \frac{C}{n}\Big(\frac{1}{n-\ell}\sum_{j=\ell+1}^{n}|b_{jj}| + |b_{\ell\ell}|\Big)$$

The remaining proof of (2.15) is nearly the same as that for (2.5). We thus omit the details.

Next, we prove (2.16). It suffices to estimate a half of the quadratic form. Recall the filtration defined in (2.7). We further set

$$L_\ell := \sum_{i<j} b_{ij}\Big(\mathbb{E}\big(x_ix_j|\mathcal{F}_\ell\big) - \mathbb{E}\big(x_ix_j|\mathcal{F}_{\ell-1}\big)\Big) = L_{\ell 1} + L_{\ell 2} + L_{\ell 3} + L_{\ell 4}, \tag{2.19}$$

where

$$\begin{aligned} L_{\ell 1} &:= \sum_{i=1}^{\ell-1} b_{i\ell} x_i \Big(x_\ell - \mathbb{E}(x_\ell|\mathcal{F}_{\ell-1})\Big) \\ L_{\ell 2} &:= \sum_{j=\ell+1}^{n} b_{\ell j}\Big(\mathbb{E}(x_j|\mathcal{F}_\ell)x_\ell - \mathbb{E}(x_jx_\ell|\mathcal{F}_{\ell-1})\Big) \\ L_{\ell 3} &:= \sum_{i=1}^{\ell-1}\sum_{j=\ell+1}^{n} b_{ij}x_i\Big(\mathbb{E}(x_j|\mathcal{F}_\ell) - \mathbb{E}(x_j|\mathcal{F}_{\ell-1})\Big) \\ L_{\ell 4} &:= \sum_{i=\ell+1}^{n}\sum_{j=i+1}^{n} b_{ij}\Big(\mathbb{E}(x_ix_j|\mathcal{F}_\ell) - \mathbb{E}(x_ix_j|\mathcal{F}_{\ell-1})\Big). \end{aligned} \tag{2.20}$$

First, using (2.5) we can improve (2.11) to

$$\Big|\sum_{i=1}^{\ell-1} x_i\Big| = \Big|\sum_{i=\ell}^{n} x_i\Big| \prec \min\Big\{\sqrt{\frac{\ell-1}{n}}, \sqrt{\frac{n-\ell+1}{n}}\Big\}. \tag{2.21}$$

Hence, in light of (2.9), we have

$$|\mathbb{E}(x_j|\mathcal{F}_{\ell-1})| \prec \frac{1}{\sqrt{n(n-\ell+1)}}. \tag{2.22}$$

Moreover, we also have for $i, j \geq \ell$ and $i \neq j$,

$$\mathbb{E}\big(x_ix_j|\mathcal{F}_{\ell-1}\big) = -\frac{1}{(n-\ell+1)(n-\ell)}\Big(\big(\sum_{i=1}^{\ell-1} x_i\big)^2 + \big(1 - \sum_{i=1}^{\ell-1} x_i^2\big)\Big). \tag{2.23}$$

Observe that

$$|1 - \sum_{i=1}^{\ell-1} x_i^2| \leq C\frac{n-\ell+1}{n}. \tag{2.24}$$

Combining (2.21), (2.23) with (2.24), we obtain for $\ell \in [\![1, n-1]\!]$

$$\Big|\mathbb{E}\big(x_ix_j|\mathcal{F}_{\ell-1}\big)\Big| \leq C\frac{1}{n(n-\ell)}, \tag{2.25}$$

and for $\ell = n$ we simply use the bound $|x_i x_j| = O(\frac{1}{n})$.

Let $q \geq 2$ be any given integer. Using Burkholder inequality again, we have

$$\mathbb{E}\Big|\sum_\ell L_\ell\Big|^q \leq (Cq)^{\frac{3q}{2}}\mathbb{E}\Big(\sum_\ell L_\ell^2\Big)^{\frac{q}{2}}. \tag{2.26}$$

Then, applying generalized Minkowski inequality, we obtain

$$\Big(\mathbb{E}\Big(\sum_\ell L_\ell^2\Big)^{\frac{q}{2}}\Big)^{\frac{2}{q}} \leq \sum_\ell \Big(\mathbb{E}|L_\ell|^q\Big)^{\frac{2}{q}}. \tag{2.27}$$

Hence, it suffices to estimate $\mathbb{E}|L_{\ell a}|^q$ for $a = 1, 2, 3, 4$. For $\mathbb{E}|L_{\ell 1}|^q$, using the bound $|x_i| = O(\frac{1}{\sqrt{n}})$, we have

(2.28)
$$\mathbb{E}|L_{\ell 1}|^q = \mathbb{E}\Big(\big|x_\ell - \mathbb{E}(x_\ell|\mathcal{F}_{\ell-1})\big|^q\Big|\sum_{i=1}^{\ell-1} b_{i\ell}x_i\Big|^q\Big) \leq \frac{C}{n^{\frac{q}{2}}}\mathbb{E}\Big|\sum_{i=1}^{\ell-1} b_{i\ell}x_i\Big|^q \prec \frac{(\sum_{i=1}^{\ell-1} b_{i\ell}^2)^{\frac{q}{2}}}{n^q},$$

where the last step follows from (2.5).

Next, we estimate $\mathbb{E}|L_{\ell 2}|^q$. Plugging the bounds (2.22), (2.25) and $|x_i| = O(\frac{1}{\sqrt{n}})$ into the definition in (2.20), we have

$$|L_{\ell 2}| \prec \frac{\sum_{j=\ell+1}^n |b_{\ell j}|}{n\sqrt{n-\ell+1}} \leq \frac{\sqrt{\sum_{j=\ell+1}^n b_{\ell j}^2}}{n}.$$

Consequently, we have

$$\mathbb{E}|L_{\ell 2}|^q \prec \frac{(\sum_{j=\ell+1}^n b_{\ell j}^2)^{\frac{q}{2}}}{n^q}. \tag{2.29}$$

Next, we estimate $\mathbb{E}|L_{\ell 3}|^q$. From (2.9) and (2.11), we have

$$\big|\mathbb{E}(x_j|\mathcal{F}_\ell) - \mathbb{E}(x_j|\mathcal{F}_{\ell-1})\big| = \frac{1}{n-\ell}\Big|\frac{\sum_{i=1}^{\ell-1} x_i}{n-\ell+1} + x_\ell\Big| \prec \frac{1}{(n-\ell)\sqrt{n}}, \qquad \ell \in [\![1, n-1]\!],$$

and we also have the trivial fact $\mathbb{E}(x_j|\mathcal{F}_n) - \mathbb{E}(x_j|\mathcal{F}_{n-1}) = 0$. Therefore, by (2.5), we have

$$\begin{aligned} |L_{\ell 3}| &\prec \frac{1}{(n-\ell)\sqrt{n}} \sum_{j=\ell+1}^n \Big|\sum_{i=1}^{\ell-1} b_{ij}x_i\Big| \\ &\prec \frac{1}{(n-\ell)n} \sum_{j=\ell+1}^n \sqrt{\sum_{i=1}^{\ell-1} b_{ij}^2} \prec \frac{1}{n\sqrt{n-\ell}}\sqrt{\sum_{i=1}^{\ell-1}\sum_{j=\ell+1}^n b_{ij}^2}. \end{aligned}$$

Hence, we have

$$\mathbb{E}|L_{\ell 3}|^q \prec \frac{\left(\sum_{i=1}^{\ell-1}\sum_{j=\ell+1}^{n} b_{ij}^2\right)^{\frac{q}{2}}}{n^q(n-\ell)^{\frac{q}{2}}}. \tag{2.30}$$

Next, we estimate $\mathbb{E}|L_{\ell 4}|^q$. From (2.23), we have for $i,j>\ell$

$$\begin{aligned}
&\big|\mathbb{E}\big(x_ix_j|\mathcal{F}_\ell\big)-\mathbb{E}\big(x_ix_j|\mathcal{F}_{\ell-1}\big)\big|\\
&=\Big|\frac{1}{(n-\ell-1)(n-\ell)(n-\ell+1)}\Big(\Big(\sum_{i=1}^{\ell-1}x_i\Big)^2+\Big(1-\sum_{i=1}^{\ell-1}x_i^2\Big)\Big)\\
&\qquad+\frac{2}{(n-\ell)(n-\ell-1)}x_\ell\sum_{i=1}^{\ell-1}x_i\Big| \prec \frac{1}{n(n-\ell)^{\frac{3}{2}}}.
\end{aligned}$$

Therefore, we have

$$L_{\ell 4} \prec \frac{1}{n(n-\ell)^{\frac{3}{2}}}\sum_{i=\ell+1}^{n}\sum_{j=i+1}^{n}|b_{ij}| \le C\frac{1}{n\sqrt{n-\ell}}\sqrt{\sum_{i=\ell+1}^{n}\sum_{j=i+1}^{n}b_{ij}^2}.$$

Hence, we also have

$$\mathbb{E}|L_{\ell 4}|^q \prec \frac{\left(\sum_{i=\ell+1}^{n}\sum_{j=i+1}^{n} b_{ij}^2\right)^{\frac{q}{2}}}{n^q(n-\ell)^{\frac{q}{2}}}. \tag{2.31}$$

Using generalized Minkowski inequality again, we can conclude from (2.28), (2.29), (2.30) and (2.31) that

$$\begin{aligned}
\left(\mathbb{E}|L_\ell|^q\right)^{\frac{2}{q}} &= \Big(\mathbb{E}|\sum_{a=1}^{4}L_{\ell a}|^q\Big)^{\frac{2}{q}} \le 4\Big(\mathbb{E}(\sum_{a=1}^{4}L_{\ell a}^2)^{\frac{q}{2}}\Big)^{\frac{2}{q}} \le 4\sum_{a=1}^{4}\left(\mathbb{E}|L_{\ell a}|^q\right)^{\frac{2}{q}}\\
&\prec \frac{1}{n^2}\Big(\sum_{i=1}^{\ell-1}b_{i\ell}^2+\sum_{j=\ell+1}^{n}b_{\ell j}^2+\frac{1}{n-\ell}\sum_{j=\ell+1}^{n}\sum_{i=1}^{j-1}b_{ij}^2\Big).
\end{aligned} \tag{2.32}$$

Recall (2.27). We see from (2.32) that

$$\Big(\mathbb{E}\Big(\sum_{\ell}L_\ell^2\Big)^{\frac{q}{2}}\Big)^{\frac{2}{q}} \le \sum_{\ell}\left(\mathbb{E}|L_\ell|^q\right)^{\frac{2}{q}} \prec \frac{1}{n^2}\sum_{i<j}b_{ij}^2, \tag{2.33}$$

where we used the fact

$$\begin{aligned}
\sum_{\ell}\sum_{j=\ell+1}^{n}\sum_{i=1}^{j-1}\frac{1}{n-\ell}b_{ij}^2 &= \sum_{j}\sum_{i=1}^{j-1}\Big(\sum_{\ell=1}^{j-1}\frac{1}{n-\ell}\Big)b_{ij}^2\\
&\le C\log n\sum_{j}\sum_{i=1}^{j-1}b_{ij}^2 \prec \sum_{i<j}b_{ij}^2.
\end{aligned}$$

Plugging (2.33) into (2.26), we have

$$\mathbb{E}\Big|\sum_{\ell} L_\ell\Big|^q \leq (Cq)^{\frac{3q}{2}} \frac{\big(\sum_{i<j} b_{ij}^2\big)^{\frac{q}{2}}}{n^q}.$$

Hence, we have

$$\Big|\sum_{i\neq j} b_{ij}x_ix_j - \mathbb{E}\Big(\sum_{i\neq j} b_{ij}x_ix_j\Big)\Big| \prec \frac{1}{n}\sqrt{\sum_{i\neq j} b_{ij}^2} \tag{2.34}$$

This together with the fact $\mathbb{E}x_ix_j = -\frac{1}{n(n-1)}$ (c.f. (1.2)) concludes the proof of (2.16).

Recall the definition of $\tilde{\boldsymbol{x}}$ from (2.2) and also observe that the entries of $\boldsymbol{\xi}$'s are i.i.d.. Then using the large deviation estimates for linear and quadratic forms of the i.i.d. random variables (c.f. Corollary B.3 of [10] for instance), we see that both (2.5) and (2.6) still hold if we replace $\boldsymbol{x}$ by $\tilde{\boldsymbol{x}}$.

Hence, we completed the proof of Proposition 2.1. □

2.2. *Strong local law for S.* Recall the notation F_{c_n} as the distribution defined in Theorem 1.1 with c replaced by c_n. In addition, we dente by $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_{p\wedge n}$ the ordered p-quantiles of F_{c_n}, i.e., γ_j is the smallest real number such that

$$\int_{-\infty}^{\gamma_j} \mathrm{d}F_{c_n}(x) = \frac{p-j+1}{p}, \qquad j \in [\![1, n\wedge p]\!], \tag{2.35}$$

We denote by $\underline{m}$ the Stieltjes transform of F_{c_n} in the sequel. It is known that $\underline{m} : \mathbb{C}^+ \to \mathbb{C}^+$ satisfies the following equation

$$\underline{m}(z) = \frac{1}{1 - c_n - z - c_n z\underline{m}(z)}. \tag{2.36}$$

The following lemma on $\underline{m}(z)$ is elementary.

Lemma 2.2. *For any $z \in E + \mathrm{i}\eta \in \mathcal{D}(\varepsilon)$, we have*

$$|\underline{m}(z)| \sim 1, \tag{2.37}$$

$$\operatorname{Im}\underline{m}(z) \sim \begin{cases} \sqrt{\kappa+\eta}, & \text{if } E \leq d_{+,c_n} \\ \frac{\eta}{\sqrt{\kappa+\eta}}, & \text{if } E \geq d_{+,c_n} \end{cases} \tag{2.38}$$

where $\kappa \equiv \kappa(E) := |E - d_{+,c_n}|$.

Recall the matrix $\widetilde{S}$ defined in (2.4). We introduce some intermediate matrices between S and $\widetilde{S}$. Starting from X, we replace $\boldsymbol{x}_i$'s by $\tilde{\boldsymbol{x}}_i$'s one by one and get the sequence of intermediate matrices

$$X = X_0, \quad X_1, \quad \dots, \quad X_\ell, \quad X_{\ell+1}, \quad \dots, X_{p-1}, \quad X_p = \widetilde{X}. \tag{2.39}$$

Correspondingly, we set

$$S_\ell = X_\ell X_\ell', \qquad G_\ell \equiv G_\ell(z) := (S_\ell - z)^{-1}, \qquad m_\ell(z) := \frac{1}{p}\mathrm{Tr}G_\ell(z) \tag{2.40}$$

For $\ell = 0$, we simply write S_0, G_0, m_0 as S, G, m. We further introduce the notations

$$\Lambda_{\mathrm{d}} := \max_k |G_{kk} - \underline{m}|, \qquad \Lambda_{\mathrm{o}} := \max_{k \neq \ell} |G_{k\ell}|, \qquad \Lambda := |m - \underline{m}|. \tag{2.41}$$

We then set the domain

$$\mathcal{D}(\epsilon) := \Big\{ z = E + \mathrm{i}\eta : \frac{1}{2} d_{+,c} \le E \le 2d_{+,c}, n^{-1+\epsilon} \le \eta \le 1 \Big\} \tag{2.42}$$

We remind here that the following proof works well for a larger domain with $E \in [\frac{1}{2}d_{-,c}, 2d_{+,c}]$ (say), which covers the whole spectrum, in case $c \neq 1$. Here in $\mathcal{D}(\epsilon)$ we focus on a neighborhood of the right edge $d_{+,c}$ only to avoid the discussion on the regime of c. The following discussion restricted to the domain $\mathcal{D}(\epsilon)$ is sufficient for the universality of the largest eigenvalues. We then further define the control parameter

$$\Psi \equiv \Psi(z) := \sqrt{\frac{\mathrm{Im}\, \underline{m}}{N\eta}} + \frac{1}{N\eta}.$$

We claim that the following local law holds.

Proposition 2.3. *Under the assumption (1.1), the following bounds hold.*
(i): (Entrywise local law)

$$\Lambda_{\mathrm{d}}(z) \prec \Psi(z), \qquad \Lambda_{\mathrm{o}}(z) \prec \Psi(z)$$

holds uniformly on $\mathcal{D}(\epsilon)$.
(ii): (Strong local law)

$$\Lambda(z) \prec \frac{1}{N\eta}$$

holds uniformly on $\mathcal{D}(\epsilon)$.
(iii): (Rigidity on the right edge). For $i \in [1, \delta p]$ *with any sufficiently small constant* $\delta \in (0,1)$, *we have*

$$|\lambda_i(S) - \gamma_i| \prec n^{-\frac{2}{3}} i^{-\frac{1}{3}}.$$

All the above hold if we replace S *by* S_ℓ *for all* $\ell \in [\![1, p]\!]$.

In the sequel, we will prove Proposition 2.3, based on the large deviation estimate in Proposition 2.1 and the general framework developed in [17].

Proof of Proposition 2.3. With the large deviation estimate in Proposition 2.1, the proof of Proposition 2.3 is nearly the same as that for Theorem 3.1 in [17]. The main difference is that here we state all the estimates with the notation $\prec$ (c.f. Definition 1.5) instead of the more quantitive statements in [17]. More directly, we can regard Proposition 2.3 as a consequence of Proposition 2.1 and Theorem 3.6 in [17]. Nevertheless, here in (2.6), we do have $\frac{1}{n}\mathrm{Tr}B\Sigma$ which is not exactly the same as $\frac{1}{n}\mathrm{Tr}B$ in Lemma 3.4 of [17] (set $\sigma^2 = \frac{1}{n}$ therein). In the sequel, we justify this minor issue.

We first define a random control parameter

$$\Pi \equiv \Pi(z) := \sqrt{\frac{\mathrm{Im}\,\underline{m}(z) + |\Lambda(z)|}{n\eta}} + \frac{1}{n\eta}.$$

Observe that since $\mathrm{Im}\,\underline{m}(z) \gtrsim \eta$, we always have $\Psi(z), \Pi(z) \gtrsim n^{-\frac{1}{2}}$. We denote by $X^{(i)}$ the submatrix of X with $\boldsymbol{x}_i$ deleted. Further, we denote by $S^{(i)} = X^{(i)}(X^{(i)})'$ and $G^{(i)} := (S^{(i)} - z)^{-1}$. Denote by

(2.43)
$$\mathcal{S} := X'X, \quad \mathcal{G}(z) := (\mathcal{S} - z)^{-1}, \quad \mathcal{S}^{(i)} := (X^{(i)})'X^{(i)}, \quad \mathcal{G}^{(i)}(z) := (\mathcal{S}^{(i)} - z)^{-1}.$$

Observe that

$$\mathrm{Tr}\mathcal{G}(z) = \mathrm{Tr}G(z) - \frac{n-p}{z}, \quad \mathrm{Tr}\mathcal{G}^{(i)}(z) = \mathrm{Tr}G^{(i)}(z) - \frac{n-p+1}{z}. \tag{2.44}$$

Let G_{ij} be the (i,j)th entry of G. Using Schur complement, we see that

$$G_{ii} = \frac{1}{\boldsymbol{x}_i\boldsymbol{x}_i' - z - \boldsymbol{x}_i(X^{(i)})'G^{(i)}X^{(i)}\boldsymbol{x}_i'}.$$

The place we need to use (2.6) is the following

(2.45)
$$\boldsymbol{x}_i(X^{(i)})'G^{(i)}X^{(i)}\boldsymbol{x}_i' - \frac{1}{n}\mathrm{Tr}(X^{(i)})'G^{(i)}X^{(i)}\Sigma = O_{\prec}\Big(\frac{1}{n}\sqrt{\mathrm{Tr}((X^{(i)})'G^{(i)}X^{(i)})^2}\Big).$$

The key observation is that

$$\mathrm{Tr}(X^{(i)})'G^{(i)}X^{(i)}\mathbf{1}'\mathbf{1} = \mathbf{1}(X^{(i)})'G^{(i)}X^{(i)}\mathbf{1}' = 0, \tag{2.46}$$

since $\boldsymbol{x}_k\mathbf{1}' = \sum_j x_{kj} = 0$. Moreover, we can write

$$(X^{(i)})'G^{(i)}X^{(i)} = \mathcal{S}^{(i)}\mathcal{G}^{(i)} = (I_n + z\mathcal{G}^{(i)}), \tag{2.47}$$

where $\mathcal{S}^{(i)}$ and $\mathcal{G}^{(i)}$ are defined in (2.43). Hence, from (2.3), (2.46), (2.47), we see that

$$\frac{1}{n}\mathrm{Tr}(X^{(i)})'G^{(i)}X^{(i)}\Sigma = \frac{1}{n-1}\mathrm{Tr}(X^{(i)})'G^{(i)}X^{(i)}$$
$$= \frac{1}{n-1}\mathrm{Tr}(I_n + z\mathcal{G}^{(i)}) = \frac{n}{n-1} + z\frac{1}{n-1}\mathrm{Tr}\mathcal{G}^{(i)}.$$

Hence, we can write (2.45) as

$$\begin{aligned}\boldsymbol{x}_i(X^{(i)})'G^{(i)}X^{(i)}\boldsymbol{x}_i' =& \frac{n}{n-1} + z\frac{1}{n-1}\mathrm{Tr}\mathcal{G}^{(i)} + O_\prec\Big(\frac{1}{n}\sqrt{n + z\mathrm{Tr}\mathcal{G}^{(i)} + z^2\mathrm{Tr}(\mathcal{G}^{(i)})^2}\Big)\\ =& \frac{n}{n-1} + z\frac{1}{n-1}\mathrm{Tr}\mathcal{G}^{(i)} + O_\prec(\Pi) = 1 + z\mathrm{Tr}\mathcal{G}^{(i)} + O_\prec(\Pi),\end{aligned} \tag{2.48}$$

where the second step follows from (2.44), the fact $|\mathrm{Tr}G - \mathrm{Tr}G^{(i)}| \prec \frac{1}{\eta}$ and the fact $|m| \leq |\underline{m}| + |\Lambda| \leq C + |\Lambda|$, and the fact $\Pi(z) \gtrsim n^{-\frac{1}{2}}$. Moreover, (2.46) also holds even we replace X by any intermediate matrix X_ℓ defined in (2.39), since $\tilde{\boldsymbol{x}}_i\mathbf{1}' = \boldsymbol{\xi}_i\Sigma^{\frac{1}{2}}\mathbf{1}' = 0$. Hence, the estimate (2.48) also holds if we replace X by any X_ℓ. All the remaining proof is the same as the counterpart in [17]. We thus omit the details.

Hence, we conclude the proof of Proposition 2.3. □

3. Edge universality for $\boldsymbol{S}$. In this section, we prove the edge universality of S by Green function comparison.

3.1. *Green function comparison.* Recall the intermediate matrices defined in (2.39) and the notations introduced in (2.40). Our aim is to show the following lemma.

LEMMA 3.1. *Fix any $\gamma \in [\![0, p]\!]$. Let $\varepsilon > 0$ be any sufficiently small constant. Let $E, E_1, E_2 \in \mathbb{R}$ satisfy $E_1 < E_2$ and*

$$|E|, |E_1|, |E_2| \leq n^{-\frac{2}{3}+\varepsilon}, \tag{3.1}$$

and set $\eta_0 = n^{-\frac{2}{3}-\varepsilon}$. Let $F : \mathbb{R} \to \mathbb{R}$ be a smooth function satisfying

$$\max_{x\in\mathbb{R}} |F^{(\ell)}(x)|(|x|+1)^{-C} \leq C, \qquad \ell = 1, 2, 3, 4$$

for some positive constant C. Then, there exists a constant $\delta > 0$ such that, for sufficiently large n we have

$$\begin{aligned}\Big|&\mathbb{E}F(n\eta_0\mathrm{Im}\, m_\gamma(d_{+,c_n} + E + \mathrm{i}\eta_0))\\ &- \mathbb{E}F(n\eta_0\mathrm{Im}\, m_{\gamma+1}(d_{+,c_n} + E + \mathrm{i}\eta_0))\Big| \prec n^{-1-\delta},\end{aligned} \tag{3.2}$$

and also

$$
\begin{aligned}
&\Big|\mathbb{E}F\Big(n\int_{E_1}^{E_2}\operatorname{Im} m_\gamma(d_{+,c_n}+x+\mathrm{i}\eta_0)\mathrm{d}x\Big)\\
&\qquad -\mathbb{E}F\Big(n\int_{E_1}^{E_2}\operatorname{Im} m_{\gamma+1}(d_{+,c_n}+x+\mathrm{i}\eta_0)\mathrm{d}x\Big)\Big| \prec n^{-1-\delta}. \qquad (3.3)
\end{aligned}
$$

PROOF OF LEMMA 3.1. In the sequel, we only show the proof for (3.2). The proof of (3.3) can be done similarly. For brevity, throughout the proof, we will simply write $C\varepsilon$ with any positive constant C (independent of ε) by ε in the sequel. In other words, we allow ε to vary from line to line, up to C. Suppose that X_γ and $X_{\gamma+1}$ differ by the i-th row for some $i\in [\![1,p]\!]$.

We first define the matrix $X_\ell^{(i)}$ to be the submatrix of X_ℓ with i-th row removed. Hence, $X_\gamma^{(i)}=X_{\gamma+1}^{(i)}$. Further, we set

$$
\begin{aligned}
&S_\gamma^{(i)}:=X_\gamma^{(i)}(X_\gamma^{(i)})', \qquad G_\gamma^{(i)}:=(S_\gamma^{(i)}-z)^{-1}, \qquad m_\gamma^{(i)}:=\frac{1}{p}\mathrm{Tr}G_\gamma^{(i)},\\
&\mathcal{S}_\gamma^{(i)}:=(X_\gamma^{(i)})'X_\gamma^{(i)}, \qquad \mathcal{G}_\gamma^{(i)}:=(\mathcal{S}_\gamma^{(i)}-z)^{-1}.
\end{aligned}
$$

We now expand $m_\gamma(z)$ around $m_\gamma^{(i)}(z)$ as follows

$$
\begin{aligned}
m_\gamma&=\frac{1}{p}\mathrm{Tr}G_\gamma(z)=\frac{1}{p}\mathrm{Tr}\mathcal{G}_\gamma(z)+\frac{n-p}{pz}=\frac{1}{p}\mathrm{Tr}\Big(\mathcal{G}_\gamma^{(i)}-\frac{\mathcal{G}_\gamma^{(i)}\boldsymbol{x}_i'\boldsymbol{x}_i\mathcal{G}_\gamma^{(i)}}{1+\boldsymbol{x}_i\mathcal{G}_\gamma^{(i)}\boldsymbol{x}_i'}\Big)+\frac{n-p}{pz}\\
&=m_\gamma^{(i)}-\frac{1}{pz}-\frac{1}{p}\frac{\boldsymbol{x}_i(\mathcal{G}_\gamma^{(i)})^2\boldsymbol{x}_i}{1+\boldsymbol{x}_i\mathcal{G}_\gamma^{(i)}\boldsymbol{x}_i}:=\mu_\gamma^{(i)}-\frac{1}{p}\frac{\boldsymbol{x}_i(\mathcal{G}_\gamma^{(i)})^2\boldsymbol{x}_i}{1+\boldsymbol{x}_i\mathcal{G}_\gamma^{(i)}\boldsymbol{x}_i}.
\end{aligned}
$$

We further denote

$$
\mathcal{E}_i:=\boldsymbol{x}_i\mathcal{G}_\gamma^{(i)}\boldsymbol{x}_i'-\frac{p}{n}\underline{m}(z).
$$

We then do the following expansion

$$
\frac{n\eta_0}{p}\frac{\boldsymbol{x}_i(\mathcal{G}_\gamma^{(i)})^2\boldsymbol{x}_i'}{1+\boldsymbol{x}_i\mathcal{G}_\gamma^{(i)}\boldsymbol{x}_i'}=\delta_{i1}+\delta_{i2}+\delta_{i3}+O_\prec(n^{-\frac{4}{3}}), \qquad (3.4)
$$

where

$$
\delta_{ik}:=\frac{n\eta_0}{p}\frac{1}{(1+\frac{p}{n}\underline{m}(z))^k}(-\mathcal{E}_i)^{k-1}\boldsymbol{x}_i(\mathcal{G}_\gamma^{(i)})^2\boldsymbol{x}_i', \qquad k=1,2,3.
$$

In (3.4), we also used the estimates

$$
|\mathcal{E}_i|\prec n^{-\frac{1}{3}+\varepsilon}, \qquad |\boldsymbol{x}_i(\mathcal{G}_\gamma^{(i)})^2\boldsymbol{x}_i'|\prec n^{\frac{1}{3}+\varepsilon} \qquad (3.5)
$$

which follow from Proposition 2.1, the fact $\mathcal{G}^{(i)}\mathbf{1}'=0$, Lemma 2.2 and Proposition 2.3. From (3.5), we also have

$$|\delta_{ik}| \prec n^{-\frac{k}{3}+\varepsilon}, \qquad k=1,2,3. \tag{3.6}$$

Consequently, we have the expansion

$$\begin{aligned}
&F(n\eta_0 \mathrm{Im}\, m_\gamma(z)) - F(n\eta_0 \mathrm{Im}\, \mu_\gamma^{(i)}(z)) \\
&= -F^{(1)}(n\eta_0 \mathrm{Im}\, \mu_\gamma^{(i)}(z))\big(\mathrm{Im}\,\delta_{i1} + \mathrm{Im}\,\delta_{i2} + \mathrm{Im}\,\delta_{i3}\big) \\
&\quad + F^{(2)}(n\eta_0 \mathrm{Im}\, \mu_\gamma^{(i)}(z))\Big(\frac{1}{2}(\mathrm{Im}\,\delta_{i1})^2 + \mathrm{Im}\,\delta_{i1}\mathrm{Im}\,\delta_{i2}\Big) \\
&\quad - F^{(3)}(n\eta_0 \mathrm{Im}\, \mu_\gamma^{(i)}(z))\Big(\frac{1}{6}(\mathrm{Im}\,\delta_{i1})^3\Big) + O_\prec(n^{-\frac{4}{3}+\varepsilon}),
\end{aligned}$$

where we used (3.6). Hence, to prove (3.2), it suffices to show that for all nonnegative integers a,b satisfying $a\ge 1$ and $a+b\le 3$, the following holds

$$\eta_0^a\Big|\mathbb{E}(\boldsymbol{x}_i(\mathcal{G}_\gamma^{(i)})^2\boldsymbol{x}_i')^a(\boldsymbol{x}_i\mathcal{G}_\gamma^{(i)}\boldsymbol{x}_i')^b - \mathbb{E}(\tilde{\boldsymbol{x}}_i(\mathcal{G}_\gamma^{(i)})^2\tilde{\boldsymbol{x}}_i')^a(\tilde{\boldsymbol{x}}_i\mathcal{G}_\gamma^{(i)}\tilde{\boldsymbol{x}}_i')^b\Big| \prec n^{-1-\delta}. \tag{3.7}$$

Observe that the LHS is 0 if $a+b=1$ (i.e., $a=1,b=0$), due to the fact that the covariance structure of $\boldsymbol{x}_i$ is the same as that of $\tilde{\boldsymbol{x}}_i$. For the case of $a+b=2,3$, we need the following technical lemma which can be obtained via elementary calculation.

Lemma 3.2. *Let $\boldsymbol{x}=(x_1,\dots,x_n)$ and $\tilde{\boldsymbol{x}}=(\tilde{x}_1,\dots,\tilde{x}_n)$ be defined in Section 2.1. Then for any vector of index $\boldsymbol{k}=(k_1,k_2,\dots,k_{2d})$ for $d=2,3$, we have*

$$\Big|\mathbb{E}\Big(\prod_{i=1}^{2d} x_{k_i}\Big) - \mathbb{E}\Big(\prod_{i=1}^{2d}\tilde{x}_{k_i}\Big)\Big| \prec n^{-d-\lceil\frac{d_1(\boldsymbol{k})}{2}\rceil-1}, \tag{3.8}$$

where $d_1(\boldsymbol{k})$ represents the number of the lone index in $\boldsymbol{k}$ and $\lceil\frac{d_1(\boldsymbol{k})}{2}\rceil$ represents the smallest integer greater than or equal to $\frac{d_1(\boldsymbol{k})}{2}$.

With the aid of Lemma 3.2, we proceed to the proof of (3.7). In case of the Pearson's sample correlation matrix in [16], the counterpart of (3.8) has a sharper bound $n^{-d-\max\{d_1(\boldsymbol{k}),1\}}$, see Lemma 5.5 therein. Nevertheless, the bound in (3.8) is as good as that in [16] when $d_1(\boldsymbol{k})\le 3$. Consequently, in the sequel, we only need to check those terms with $d_1(\boldsymbol{k})\ge 4$. The case of $d_1(\boldsymbol{k})\le 3$ can be done in the same way as [16].

We start with the case $a=1,b=1$. In this case, we have

$$\begin{aligned}
\eta_0\mathbb{E}_i(\boldsymbol{x}_i(\mathcal{G}_\gamma^{(i)})^2\boldsymbol{x}_i')(\boldsymbol{x}_i\mathcal{G}_\gamma^{(i)}\boldsymbol{x}_i') &= \eta_0\sum_{\boldsymbol{k}:d_1(\boldsymbol{k})\le 3}((\mathcal{G}_\gamma^{(i)})^2)_{k_1k_2}(\mathcal{G}_\gamma^{(i)})_{k_3k_4}\mathbb{E}x_{k_1}x_{k_2}x_{k_3}x_{k_4} \\
&\quad + \eta_0\sum_{\boldsymbol{k}:d_1(\boldsymbol{k})>3}((\mathcal{G}_\gamma^{(i)})^2)_{k_1k_2}(\mathcal{G}_\gamma^{(i)})_{k_3k_4}\mathbb{E}x_{k_1}x_{k_2}x_{k_3}x_{k_4}.
\end{aligned} \tag{3.9}$$

Apparently, (3.9) and (3.10) still hold if we replace $\boldsymbol{x}$ by $\tilde{\boldsymbol{x}}$. As mentioned above, we can use the argument in [16] to conclude

$$\Big|\eta_0 \sum_{\boldsymbol{k}:d_1(\boldsymbol{k})\le 3} ((\mathcal{G}_\gamma^{(i)})^2)_{k_1k_2}(\mathcal{G}_\gamma^{(i)})_{k_3k_4}\big(\mathbb{E}x_{k_1}x_{k_2}x_{k_3}x_{k_4} - \mathbb{E}\tilde{x}_{k_1}\tilde{x}_{k_2}\tilde{x}_{k_3}\tilde{x}_{k_4}\big)\Big| \prec n^{-1-\delta}.$$

For the second part in (3.9), we observe that

$$\mathbf{1}(d_1(\boldsymbol{k}) > 3)\mathbb{E}x_{k_1}x_{k_2}x_{k_3}x_{k_4} = \mathbb{E}x_1x_2x_3x_{k_4}. \tag{3.10}$$

Using (3.10) and (3.8), we get

$$\begin{aligned}&\Big|\eta_0 \sum_{\boldsymbol{k}:d_1(\boldsymbol{k})> 3} ((\mathcal{G}_\gamma^{(i)})^2)_{k_1k_2}(\mathcal{G}_\gamma^{(i)})_{k_3k_4}\big(\mathbb{E}x_{k_1}x_{k_2}x_{k_3}x_{k_4} - \mathbb{E}\tilde{x}_{k_1}\tilde{x}_{k_2}\tilde{x}_{k_3}\tilde{x}_{k_4}\big)\Big| \\ &\prec n^{-5}\eta_0\sum_{\boldsymbol{k}}|((\mathcal{G}_\gamma^{(i)})^2)_{k_1k_2}||(\mathcal{G}_\gamma^{(i)})_{k_3k_4}| \le n^{-3}\eta_0\sqrt{\mathrm{Tr}|\mathcal{G}_\gamma^{(i)}|^4}\sqrt{\mathrm{Tr}|\mathcal{G}_\gamma^{(i)}|^2} \prec n^{-\frac{5}{3}+\varepsilon},\end{aligned} \tag{3.11}$$

where in the second step we used Cauchy-Schwarz inequality. The case of $a = 2, b = 0$ can be proved similarly. More specifically, we can again decompose the sum into two parts according to $d_1(\boldsymbol{k}) \le 3$ or $d_1(\boldsymbol{k}) > 3$. In the first case, we can simply use the argument in [16] to conclude the estimate. For the part with $d_1(\boldsymbol{k}) > 3$, instead of the bound $n^{-\frac{5}{3}+\varepsilon}$ in (3.11), we will have $n^{-\frac{4}{3}+\varepsilon}$ for the case of $a = 2, b = 0$.

For the case of $a + b = 3$, we have

$$\eta_0^a\mathbb{E}_i(\boldsymbol{x}_i(\mathcal{G}_\gamma^{(i)})^2\boldsymbol{x}_i')^a(\boldsymbol{x}_i\mathcal{G}_\gamma^{(i)}\boldsymbol{x}_i')^b = \eta_0^a\sum_{\boldsymbol{k}}\prod_{j=1}^{a}((\mathcal{G}_\gamma^{(i)})^2)_{k_{2j-1}k_{2j}}\prod_{j=a+1}^{3}(\mathcal{G}_\gamma^{(i)})_{k_{2j-1}k_{2j}}\mathbb{E}\Big(\prod_{j=1}^{6}x_{k_j}\Big)$$

The above also holds if we replace $\boldsymbol{x}$ by $\tilde{\boldsymbol{x}}$. Again, we can decompose the sum over $\boldsymbol{k}$ into two parts according to whether $d_1(\boldsymbol{k}) \le 3$ or $d_1(\boldsymbol{k}) > 3$. The estimate of the first part follows from the discussion in [16] again. Hence, it suffices to consider the cases $d_1(\boldsymbol{k}) = 6$ or $d_1(\boldsymbol{k}) = 4$. For the first case, by (3.8) we have

$$\begin{aligned}&\Big|\eta_0^a\sum_{\boldsymbol{k}:d_1(\boldsymbol{k})=6}\prod_{j=1}^{a}((\mathcal{G}_\gamma^{(i)})^2)_{k_{2j-1}k_{2j}}\prod_{j=a+1}^{3}(\mathcal{G}_\gamma^{(i)})_{k_{2j-1}k_{2j}}\Big(\mathbb{E}\big(\prod_{j=1}^{6}x_{k_j}\big) - \mathbb{E}\big(\prod_{j=1}^{6}\tilde{x}_{k_j}\big)\Big)\Big| \\ &\prec n^{-7}\eta_0^a\Big|\sum_{\boldsymbol{k}:d_1(\boldsymbol{k})=6}\prod_{j=1}^{a}((\mathcal{G}_\gamma^{(i)})^2)_{k_{2j-1}k_{2j}}\prod_{j=a+1}^{3}(\mathcal{G}_\gamma^{(i)})_{k_{2j-1}k_{2j}}\Big| \\ &\prec n^{-4}\eta_0^a\big(\mathrm{Tr}|\mathcal{G}_\gamma^{(i)}|^4\big)^{\frac{a}{2}}\big(\mathrm{Tr}|\mathcal{G}_\gamma^{(i)}|^2\big)^{\frac{b}{2}} \prec n^{-2+\varepsilon}.\end{aligned}$$

When $a + b = 3$ and $d_1(\boldsymbol{k}) = 4$. We shall further decompose the discussion into three cases $(a, b) = (1, 2)$, $(2, 1)$ or $(3, 0)$. The discussion for all three cases are similar, we thus only present the details for the first case in the sequel. In this case, using (3.8) and the fact $d_1(\boldsymbol{k})=4$, we have

$$\Big|\eta_0 \sum_{\boldsymbol{k}:d_1(\boldsymbol{k})=4} ((\mathcal{G}_\gamma^{(i)})^2)_{k_1k_2}(\mathcal{G}_\gamma^{(i)})_{k_3k_4}(\mathcal{G}_\gamma^{(i)})_{k_5k_6}\Big(\mathbb{E}\big(\prod_{j=1}^6 x_{k_j}\big)-\mathbb{E}\big(\prod_{j=1}^6 \tilde{x}_{k_j}\big)\Big)\Big|$$
$$\prec n^{-6}\eta_0\Big|\sum_{\boldsymbol{k}:d_1(\boldsymbol{k})=4} ((\mathcal{G}_\gamma^{(i)})^2)_{k_1k_2}(\mathcal{G}_\gamma^{(i)})_{k_3k_4}(\mathcal{G}_\gamma^{(i)})_{k_5k_6}\Big| =: n^{-6}\eta_0|\mathcal{I}_1+\mathcal{I}_2+\mathcal{I}_3+\mathcal{I}_4|,$$

where $\mathcal{I}_1$ represents the sum of the terms with $k_1=k_2$; $\mathcal{I}_2$ represents the sum of the terms with $k_3=k_4$ or $k_5=k_6$; $\mathcal{I}_3$ represents the sum of the terms with $\sharp(\{k_1,k_2\}\cap\{k_3,k_4,k_5,k_6\})=1$; and $\mathcal{I}_4$ represents the sum of the terms with $\sharp(\{k_3,k_4\}\cap\{k_5,k_6\})=1$. Observe that

$$|\mathcal{I}_1|\le \mathrm{Tr}|\mathcal{G}_\gamma^{(i)}|^2\big(\sum_{k,\ell}|(\mathcal{G}_\gamma^{(i)})_{k\ell}|\big)^2 \le n^2(\mathrm{Tr}|\mathcal{G}_\gamma^{(i)}|^2)^2 \prec n^4\eta_0^{-1},$$
$$|\mathcal{I}_2|\le \mathrm{Tr}|\mathcal{G}_\gamma^{(i)}|\big(\sum_{k,\ell}|(\mathcal{G}_\gamma^{(i)})_{k\ell}|\big)\big(\sum_{k,\ell}|((\mathcal{G}_\gamma^{(i)})^2)_{k\ell}|\big) \le n^2\mathrm{Tr}|\mathcal{G}_\gamma^{(i)}|\sqrt{\mathrm{Tr}|\mathcal{G}_\gamma^{(i)}|^2}\sqrt{\mathrm{Tr}|\mathcal{G}_\gamma^{(i)}|^4} \prec n^4\eta_0^{-\frac{3}{2}},$$
$$|\mathcal{I}_3|\le \big(\sum_{k,\ell}|(\mathcal{G}_\gamma^{(i)})_{k\ell}|\big)\big(\sum_{k,\ell}|((\mathcal{G}_\gamma^{(i)})^3)_{k\ell}|\big) \le n^2\sqrt{\mathrm{Tr}|\mathcal{G}_\gamma^{(i)}|^2}\sqrt{\mathrm{Tr}|\mathcal{G}_\gamma^{(i)}|^6} \prec n^3\eta_0^{-\frac{5}{2}},$$
$$|\mathcal{I}_4|\le \big(\sum_{k,\ell}|((\mathcal{G}_\gamma^{(i)})^2)_{k\ell}|\big)^2 \le n^2\mathrm{Tr}|\mathcal{G}_\gamma^{(i)}|^4 \prec n^2\eta_0^{-\frac{5}{2}}.$$

Then it is easy to check that

$$\Big|\eta_0 \sum_{\boldsymbol{k}:d_1(\boldsymbol{k})=4} ((\mathcal{G}_\gamma^{(i)})^2)_{k_1k_2}(\mathcal{G}_\gamma^{(i)})_{k_3k_4}(\mathcal{G}_\gamma^{(i)})_{k_5k_6}\Big(\mathbb{E}\big(\prod_{j=1}^6 x_{k_j}\big)-\mathbb{E}\big(\prod_{j=1}^6 \tilde{x}_{k_j}\big)\Big)\Big| \prec n^{-1-\delta}.$$

Similarly, one can check that the above estimate holds for the cases $(a,b)=(2,1)$ or $(3,0)$. This concludes the proof of (3.7). Further, we completed the proof of Lemma 3.1. □

Using Lemma 3.1, we can now prove Theorem 1.2 and Corollary 1.4.

Proof of Theorem 1.2. Similarly to the proof of Theorem 1.1 in [17], from Lemma 3.1, one can show that

$$\begin{aligned}
&\mathbb{P}\Big(n^{\frac{2}{3}}(\lambda_1(\widetilde{S})-d_{+,c_n})\le s-n^{-\varepsilon}\Big)-n^{-\delta}\\
&\le \mathbb{P}\Big(n^{\frac{2}{3}}(\lambda_1(S)-d_{+,c_n})\le s\Big)\\
&\le \mathbb{P}\Big(n^{\frac{2}{3}}(\lambda_1(\widetilde{S})-d_{+,c_n})\le s+n^{-\varepsilon}\Big)+n^{-\delta},
\end{aligned} \tag{3.12}$$

where $\widetilde{S}$ is defined in (2.4). It is known from Theorem 2.7 of [7] that the largest eigenvalues of $\widetilde{S}$ differ from the corresponding ones of $\Xi\Xi'$ only by $O_{\prec}(\frac{1}{n})$. This together with the edge universality of the sample covariance matrix in [17] further implies

$$\begin{aligned}&\mathbb{P}\Big(n^{\frac{2}{3}}(\lambda_1(Q)-d_{+,c_n})\leq s-n^{-\varepsilon}\Big)-n^{-\delta}\\&\leq\mathbb{P}\Big(n^{\frac{2}{3}}(\lambda_1(\widetilde{S})-d_{+,c_n})\leq s\Big)\\&\leq\mathbb{P}\Big(n^{\frac{2}{3}}(\lambda_1(Q)-d_{+,c_n})\leq s+n^{-\varepsilon}\Big)+n^{-\delta},\end{aligned}\tag{3.13}$$

where Q is the Wishart matrix in Theorem 1.2. Combining (3.12) with (3.13) we can conclude the proof of Theorem 1.2.

□

Proof of Corollary 1.4. The conclusion follows directly from Theorem 1.2 and the Tracy Widom limit for $\lambda_1(Q)$. This completes the proof. □

References.

[1] Z. D. Bai, W. Zhou: Large sample covariance matrices without independence structures in columns. Statistica Sinica, 425-442. (2008)
[2] A. S. Bandeira, A. Lodhia, P. Rigollet. Marčenko-Pastur law for Kendall's tau. Electronic Communications in Probability, 22. (2017)
[3] Z. G. Bao: Tracy-Widom limit for Kendall tau. arXiv:1712.00892
[4] Z. G. Bao, G. M. Pan, W. Zhou: Tracy-Widom law for the extreme eigenvalues of sample correlation matrices, Electron. J. Probab. 17, No. 88, 1-32, (2012).
[5] Z. G. Bao, G. M. Pan, W. Zhou: Universality for the largest eigenvalue of sample covariance matrices with general population, Ann. Stat. 43(1), 382-421 (2015).
[6] Z. G. Bao, L.-C. Lin, G. M. Pan, W. Zhou. Spectral statistics of large dimensional Spearman's rank correlation matrix and its application. The Annals of Statistics, **43(6)**, 2588-2623. (2015)
[7] A. Bloemendal, A. Knowles, H.-T. Yau, J. Yin: On the principal components of sample covariance matrices. Probab. Theory and Related Fields, 164(1-2): 459-552 (2016).
[8] X. Ding, F. Yang: A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. arXiv:1607.06873.
[9] Erdős, L., Knowles, A., Yau, H.-T.: *Averaging fluctuations in resolvents of random band matrices*, Ann. Henri Poincaré **14**, 1837-1926 (2013).
[10] Erdős, L., Yau, H.-T., Yin, J.: *Bulk universality for generalized Wigner matrices*, Probability Theory and Related Fields, 1-67 (2012).
[11] Erdős, L., Yau, H.-T., Yin, J.: *Rigidity of Eigenvalues of Generalized Wigner Matrices*, Adv. Math. **229** (3), 1435-1515 (2012).
[12] I. M. Johnstone: On the distribution of the largest eigenvalue in principal components analysis. Ann. Stat., 295-327 (2001).
[13] A. Knowles, J. Yin. Anisotropic local laws for random matrices. arXiv:1410.3516.
[14] J. O. Lee, K. Schnelli. Tracy-Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. Ann. Appl. Probab., 26(6), 3786-3839 (2016).
[15] V. A. Marchenko, L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. Mathematics of the USSR-Sbornik, 1(4): 457 (1967).
[16] N. S. Pillai, J. Yin: Edge universality of correlation matrices. The Annals of Statistics, 40(3): 1737-1763 (2012).

[17] N. S. Pillai, J. Yin: Universality of covariance matrices. Ann. Appl. Probab., 24(3), 935-1001 (2014).

[18] K. Wang: Random covariance matrices: Universality of local statistics of eigenvalues up to the edge. Random Matrices: Theory and Applications, 1(01): 1150005 (2012).