Submitted to the Annals of Applied Probability

ASYMPTOTIC ANALYSIS FOR EXTREME EIGENVALUES OF PRINCIPAL MINORS OF RANDOM MATRICES

BY T. TONY CAI *, TIEFENG JIANG†, AND XIAOOU LI†

University of Pennsylvania and University of Minnesota†*

Consider a standard white Wishart matrix with parameters n and p. Motivated by applications in high-dimensional statistics and signal processing, we perform asymptotic analysis on the maxima and minima of the eigenvalues of all the $m \times m$ principal minors, under the asymptotic regime that n, p, m go to infinity. Asymptotic results concerning extreme eigenvalues of principal minors of real Wigner matrices are also obtained. In addition, we discuss an application of the theoretical results to the construction of compressed sensing matrices, which provides insights to compressed sensing in signal processing and high dimensional linear regression in statistics.

1. Introduction. Random matrix theory is traditionally focused on the spectral analysis of eigenvalues and eigenvectors of a single random matrix. See, for example, Bai and Silverstein (2010); Bryc et al. (2006); Diaconis and Evans (2001); Dyson (1962a,b,c); Jiang (2004a,b); Johnstone (2001, 2008); Mehta (2004); Tracy and Widom (1994, 1996, 2000); Wigner (1955, 1958). It has been proved to be a powerful tool in many fields including high-dimensional statistics, quantum physics, electrical engineering, and number theory.

The laws of large numbers and the limiting distributions for the extreme eigenvalues of the Wishart matrices are now well known, see, e.g., Bai (1999) and Johnstone (2001, 2008). Let $X = X_{n\times p}$ be a random matrix with i.i.d. $N(0,1)$ entries and let $W = X^\intercal X$. Let $\lambda_1(W) \geq \cdots \geq \lambda_p(W)$ be the eigenvalues of W. The limiting distribution of the largest eigenvalue $\lambda_1(W)$ satisfies, for $n, p \to \infty$ with $n/p \to \gamma$,

$$\mathbb{P}\Big(\frac{\lambda_1(W) - \mu_n}{\sigma_n} \leq x\Big) \to F_1(x)$$

where $\mu_n = (\sqrt{n-1} + \sqrt{p})^2$ and $\sigma_n = (\sqrt{n-1} + \sqrt{p})(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}})^{1/3}$ and $F_1(x)$ is the distribution function of the Tracy-Widom law of type I. The results for the smallest eigenvalue $\lambda_p(W)$ can be found in, e.g., Edelman

MSC 2010 subject classifications: Primary 60B20,60F99; secondary 60K35

Keywords and phrases: random matrix, extremal eigenvalues, maximum of random variables, minimum of random variables

(1988) and Bai and Yin (1993). These results have also been extended to generalized Wishart matrices, i.e., the entries of X are i.i.d. but not necessarily normally distributed, in, e.g., Bai and Silverstein (2010); Péché (2009); Tao and Vu (2010). For limiting theory for Toeplitz minors and its connection with random matrices, see, for example, Bump and Diaconis (2002); Johansson (1990); Szeg (1939); Tracy and Widom (2002).

Motivated by applications in high-dimensional statistics and signal processing, we study in this paper the extreme eigenvalues of the principal minors of a Wishart matrix W. Write $X = (x_{ij})_{n\times p} = (x_1, \cdots, x_p)$. Let $S = \{i_1, \cdots, i_k\} \subset \{1, 2, \cdots, p\}$ with the size of S being k and $X_S = (x_{i_1}, \cdots, x_{i_k})$. Then $W_S = X_S^\intercal X_S$ is a $k \times k$ principal minor of W. Denote by $\lambda_1(W_S) \geq \cdots \geq \lambda_k(W_S)$ the eigenvalues of W_S in descending order. The quantities of interest are the largest and the smallest eigenvalues of all the $k \times k$ principal minors of W in the setting that n, p, and k are large but k relatively smaller than $\min\{n, p\}$. More specifically, we are interested in the properties of the maximum of the eigenvalues of all $k \times k$ minors:

$$\lambda_{\max}(k) = \max_{1\leq i\leq k, S\subset\{1,\ldots,p\}, |S|=k} \lambda_i(W_S) \tag{1.1}$$

and the minimum of the eigenvalues of all $k \times k$ minors:

$$\lambda_{\min}(k) = \min_{1\leq i\leq k, S\subset\{1,\ldots,p\}, |S|=k} \lambda_i(W_S), \tag{1.2}$$

where $|S|$ denotes the cardinality of the set S.

This is a problem of significant interest in its own right, and it has important applications in statistics and engineering. Before we establish the properties for the extreme eigenvalues $\lambda_{\max}(k)$ and $\lambda_{\min}(k)$, of the $k \times k$ principal minors of a Wishart matrix W, we first discuss an application in signal processing and statistics, namely the construction of the compressed sensing matrix, as the motivation for our study. The properties of the extreme eigenvalues $\lambda_{\max}(k)$ and $\lambda_{\min}(k)$ can also be used in other applications, including testing for the covariance structure of a high-dimensional Gaussian distribution, which is an important problem in statistics.

1.1. *Construction of Compressed Sensing Matrices.* Compressed sensing, which aims to develop efficient data acquisition techniques that allow accurate reconstruction of highly undersampled sparse signals, has received much attention recently in several fields, including signal processing, applied mathematics, and statistics. The development of the compressed sensing theory also provides crucial insights into inference for high dimensional linear regression in statistics. It is now well understood that the constrained ℓ_1

minimization method provides an effective way for recovering sparse signals. See, e.g., Candes and Tao (2005, 2007), Donoho (2006), and Donoho et al. (2006). More specifically, in compressed sensing, one observes (X, y) with

$$y = X\beta + z$$

where $y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$ with n being much smaller than p, $\beta \in \mathbb{R}^p$ is a sparse signal of interest, and $z \in \mathbb{R}^n$ is a vector of measurement errors. One wishes to recover the unknown sparse signal $\beta \in \mathbb{R}^p$ based on (X, y) using an efficient algorithm.

Since the number of measurements n is much smaller than the dimension p, without structural assumptions, the signal β is under-determined, even in the noiseless case. A usual assumption in compressed sensing is that β is sparse and one of the most commonly used frameworks for sparse signal recovery is the *Restricted Isometry Property* (RIP). See Candes and Tao (2005). A vector is said to be k-sparse if $|\text{supp}(v)| \leq k$, where $\text{supp}(v) = \{i : v_i \neq 0\}$ is the support of v. In compressed sensing, the RIP requires subsets of certain cardinality of the columns of X to be close to an orthonormal system. For an integer $1 \leq k \leq p$, define the restricted isometry constant $\delta_k \geq 0$ to be the smallest number such that for all k-sparse vectors β,

$$(1 - \delta_k)\|\beta\|_2^2 \leq \|X\beta\|_2^2 \leq (1 + \delta_k)\|\beta\|_2^2. \tag{1.3}$$

There are a variety of sufficient conditions on the RIP for the exact/stable recovery of k-sparse signals. For example, Cai et al. (2009, 2010); Cai and Zhang (2013); Candes (2008); Mo and Li (2011) provided sufficient conditions on δ_k or δ_{2k} for the exact recovery of all k-sparse signals. Dallaporta and De Castro (2019) studied the RIP condition from a deviation inequality point of view under the assumption that X is random.

A sharp condition was established in Cai and Zhang (2014) and a conjecture was proved in Zhang and Li (2018). Let

$$b_*(t) = \begin{cases} \frac{t}{4-t} & 0 < t < \frac{4}{3} \\ \sqrt{\frac{t-1}{t}} & t \geq \frac{4}{3} \end{cases}. \tag{1.4}$$

For any given $t > 0$, the condition $\delta_{tk} < b_*(t)$ guarantees the exact recovery of all k sparse signals in the noiseless case through the constrained ℓ_1 minimization

$$\hat{\beta} = \arg\min\{\|\gamma\|_1 : y = X\gamma,\ \gamma \in \mathbf{R}^p\}.$$

Moreover, for any $\varepsilon > 0$, $\delta_{tk} < b_*(t) + \varepsilon$ is not sufficient to guarantee the exact recovery of all k-sparse signals for large k. In addition, the conditions

$\delta_{tk} < b_*(t)$ is also shown to be sufficient for stable recovery of approximately sparse signals in the noisy case.

One of the major goals of compressed sensing is the construction of the measurement matrix $X_{n\times p}$, with the number of measurements n as small as possible relative to p, such that all k-sparse signals can be accurately recovered. Deterministic construction of large measurement matrices that satisfy the RIP is known to be difficult. Instead, random matrices are commonly used. Certain random matrices have been shown to satisfy the RIP conditions with high probability. See, e.g., Baraniuk et al. (2008). When the measurement matrix X is a Gaussian matrix with i.i.d. $N(0, \frac{1}{n})$ entries, for any given t, the condition $\delta_{tk} < b_*(t)$ is equivalent to that the extreme eigenvalues, $\lambda_{\max}(tk)$ and $\lambda_{\min}(tk)$, of the $tk \times tk$ principal minors of the Wishart matrix $W = X^\intercal X$ satisfy

$$1 - b_*(t) < \lambda_{\min}(tk) \leq \lambda_{\max}(tk) < 1 + b_*(t).$$

Hence the condition (1.3) can be viewed as a condition on $\lambda_{\min}(tk)$ and $\lambda_{\max}(tk)$ as defined in (1.1) and (1.2), respectively.

1.2. *Main results and organization of the paper.* In this paper, we investigate the asymptotic behavior of the extreme eigenvalues $\lambda_{\max}(m)$ and $\lambda_{\min}(m)$ defined in (1.1) and (1.2). We also consider the extreme eigenvalues of a related Wigner matrix. We then discuss the application of the results in the construction of compressed sensing matrices.

The rest of the paper is organized as follows. Section 2 describes the precise setting of the problem. The main results are stated in Section 3. The proofs of the main theorems are given in Section 4. The proofs of all the supporting lemma are given in the Appendix. The proof strategy for the main results is given in Section 4.1.

2. Problem settings. In this paper, we consider a white Wishart matrix $W = (w_{ij})_{1\leq i,j\leq p} = X^\intercal X$, where $X = (x_{ij})_{1\leq i\leq n, 1\leq j\leq p}$ and x_{ij} are independent $N(0,1)$-distributed random variables. For $S \subset \{1, ..., p\}$, set the principal minor $W_S = (w_{ij})_{i,j\in S}$. For an $m\times m$ symmetric matrix A, let $\lambda_1(A)$ and $\lambda_m(A)$ denote the largest and the smallest eigenvalues of A, respectively. Let

$$T_{m,n,p} = \max_{S\subset\{1,\ldots,p\}, |S|=m} \lambda_1(W_S), \tag{2.1}$$

and $|S|$ denotes the cardinality of the set S. We also define

$$V_{m,n,p} = \min_{S\subset\{1,\ldots,p\}, |S|=m} \lambda_m(W_S). \tag{2.2}$$

Of interest is the asymptotic behavior of $T_{m,n,p}$ and $V_{m,n,p}$ when both n and p grow large.

Notice W_{ij} is the sum of n independent and identically distributed (i.i.d.) random variables. By the standard central limit theorem, for given $i \geq 1$ and $j \geq 1$, we have

$$\frac{w_{ij} - n}{\sqrt{n}} \Longrightarrow N(0,2) \text{ if } i = j, \text{ and } \frac{w_{ij}}{\sqrt{n}} \Longrightarrow N(0,1) \text{ if } i \neq j,$$

as $n \to \infty$, where we use "$\Longrightarrow$" to indicate convergence in distribution. Motivated by this limiting distribution, we also consider the Wigner matrix $\tilde{W} = (\tilde{w}_{ij})_{1 \leq i,j \leq p}$, which is a symmetric matrix whose upper triangular entries are independent Gaussian variables with the following distribution

$$\tilde{w}_{ij} \sim \begin{cases} N(0,2) \text{ if } i = j; \\ N(0,1) \text{ if } i < j. \end{cases} \tag{2.3}$$

For $S \subset \{1, ..., p\}$, set $\tilde{W}_S = (\tilde{w}_{ij})_{i,j \in S}$. We will work on the corresponding statistics

$$\tilde{T}_{m,p} = \max_{S \subset \{1,...,p\}, |S| = m} \lambda_1(\tilde{W}_S) \tag{2.4}$$

and

$$\tilde{V}_{m,p} = \min_{S \subset \{1,...,p\}, |S| = m} \lambda_m(\tilde{W}_S). \tag{2.5}$$

In this paper, we study asymptotic results regarding the four statistics $T_{m,n,p}$, $V_{m,n,p}$, $\tilde{V}_{m,p}$ and $\tilde{T}_{m,p}$. In addition, we will also discuss the extension to random matrices with non-Gaussian entries in Remark 4.

3. Main results.

Throughout the paper, we will let $n \to \infty$ and let $p = p_n \to \infty$ with a speed depending on n. The following technical assumptions will be used in our main results.

Assumption 1. The integer $m \geq 2$ is fixed and $\log p = o(n^{1/2})$; or $m \to \infty$ with

$$m = o\left(\min\left\{\frac{(\log p)^{1/3}}{\log\log p}, \frac{n^{1/4}}{(\log n)^{3/2}(\log p)^{1/2}}\right\}\right). \tag{3.1}$$

Notice the second part of Assumption 1 implies that $\log p = o(n^{1/2}(\log n^{-3}))$. It says the population dimension p can be very large and it can be as large as

$\exp\{o(n^{1/2}/\log n^3)\}$. This assumption is used in the analysis of $T_{m,n,p}$ and $V_{m,n,p}$. The requirement $m = o((\log p)^{1/3}/\log\log p)$ is used in the last step in (4.55). The second part of the condition $m = o(n^{1/4}(\log n)^{-3/2}(\log p)^{-1/2})$ is needed in a few places including (4.52). The key scales $(\log p)^{1/3}$ and $n^{1/4}$ in condition (3.1) are tight, the terms of lower order $\log\log p$ and $(\log n)^{3/2}$ can be improved to be relatively smaller.

The next assumption is needed for studying the properties of $\tilde{V}_{m,p}$ and $\tilde{T}_{m,p}$.

Assumption 2. The integer m satisfies that

$$m \geq 2 \text{ is fixed, or } m \to \infty \text{ with } m = o\Big(\frac{(\log p)^{1/3}}{\log\log p}\Big). \tag{3.2}$$

This condition is the same as the first part of (3.1). We start with asymptotic results for $T_{m,n,p}$ in (2.1) and $V_{m,n,p}$ in (2.2).

Theorem 1 *Suppose Assumption 1 in (3.1) holds. Recall $T_{m,n,p}$ defined as in (2.1). Then,*

$$Z_n := \frac{T_{m,n,p} - n}{\sqrt{n}} - 2\sqrt{m\log p} \to 0$$

in probability as $n \to \infty$. Furthermore,

$$\lim_{n\to\infty} \mathbb{E}\left[e^{\alpha|Z_n|}\mathbf{1}_{\{|Z_n|\geq\delta\}}\right] = 0 \tag{3.3}$$

for all $\alpha > 0$ and $\delta > 0$.

Remark 1 Suppose Assumption 1 in (3.1) holds. Recall $V_{m,n,p}$ defined as in (2.2). Similar to the proof of Theorem 1 it can be shown that

$$Z_n' := \frac{V_{m,n,p} - n}{\sqrt{n}} + 2\sqrt{m\log p} \to 0$$

in probability as $n \to \infty$, and furthermore,

$$\lim_{n\to\infty} \mathbb{E}\left[e^{\alpha|Z_n'|}\mathbf{1}_{\{|Z_n'|\geq\delta\}}\right] = 0 \tag{3.4}$$

for all $\alpha > 0$ and $\delta > 0$. For reasons of space, we omit the details here.

We now turn to the asymptotic analysis for $\tilde{T}_{m,p}$ and $\tilde{V}_{m,p}$.

Theorem 2 *Suppose Assumption 2 in (3.2) is satisfied. Recall $\tilde{T}_{m,p}$ defined as in (2.4). Then,*

$$\tilde{Z}_p := \tilde{T}_{m,p} - 2\sqrt{m \log p} \to 0$$

in probability as $n \to \infty$. Furthermore,

$$\lim_{p\to\infty} \mathbb{E}\left[e^{\alpha|\tilde{Z}_p|}\mathbf{1}_{\{|\tilde{Z}_p|\geq\delta\}}\right] = 0 \tag{3.5}$$

for all $\alpha > 0$ and $\delta > 0$.

Remark 2 Suppose Assumption 2 in (3.2) is satisfied. Review $\tilde{W} = (\tilde{w}_{ij})_{1\leq i,j\leq p}$ above (2.3), we know $\tilde{W}$ and $-\tilde{W}$ have the same distribution. Let $\tilde{V}_{m,p}$ be defined as in (2.5). It follows that $-\tilde{T}_{m,p}$ and $\tilde{V}_{m,p}$ have the same distribution. Then, by Theorem 2,

$$\tilde{Z}'_p := \tilde{V}_{m,p} + 2\sqrt{m \log p} \to 0$$

in probability as $n \to \infty$. Furthermore,

$$\lim_{p\to\infty} \mathbb{E}\left[e^{\alpha|\tilde{Z}'_p|}\mathbf{1}_{\{|\tilde{Z}'_p|\geq\delta\}}\right] = 0 \tag{3.6}$$

for all $\alpha > 0$ and $\delta > 0$.

The following remark provides a further explanation of the convergence results in (3.3)–(3.6).

Remark 3 Equation (3.3) has the following implications, whose rigorous justification is given in Section 4.

1. $\lim_{n\to\infty} \mathbb{E}\left[e^{\alpha|Z_n|}\right] = 1$ for all $\alpha > 0$;
2. $\lim_{n\to\infty} \mathbb{E}(|Z_n|^\alpha) = 0$ for all $\alpha > 0$;
3. $\lim_{n\to\infty} \mathrm{Var}(Z_n) = 0$.

We now elaborate on the above results. First, the moment generating function of $|Z_n|$ exists and is close to 1 when n is large. As a result, $|Z_n|$ has a sub-exponential tail probability for large n. Second, Z_n converges to 0 in L_q for all $q > 0$. Third, the variance of Z_n vanishes for large n, indicating that $\mathrm{Var}(T_{m,n,p}) = o(n)$ as $n \to \infty$. Overall, we can see (3.3) is stronger than the typical convergence in probability. This provides information on the behavior of the tail probability. Similar interpretations can also be made for (3.4), (3.5) and (3.6), respectively.

3.1. *Extensions.* In this section, we discuss extensions of Theorems 1 and 2. Similar extensions can also be made to Remarks 1 and 2. They are omitted for the clarity of presentation.

First, we point out that Theorems 1 and 2 still hold if we replace the size-m principal minors by the principal minors with the size no larger than m in the definition of $\tilde{T}_{m,p}$ and $T_{m,n,p}$, by the eigenvalue interlacing theorem [see, e.g., Horn and Johnson (2012)]. We then have the following corollary.

Corollary 1 *Define* $\hat{T}_{m,n,p} = \max_{S\subset\{1,\ldots,p\},|S|\leq m}\lambda_1(W_S)$ *and* $\hat{T}_{m,p} = \max_{S\subset\{1,\ldots,p\},|S|\leq m}\lambda_1(\tilde{W}_S)$. *Then, Theorems 1 and 2 still hold if "*$T_{m,n,p}$*" and "*$\tilde{T}_{m,p}$*" are replaced by "*$\hat{T}_{m,n,p}$*" and "*$\hat{T}_{m,p}$*", respectively.*

Next, we extend Theorem 2 to allow other values of variance for the Wigner matrix. Here, we assume that the matrix $\tilde{W}$ to have the following distribution, instead of that in (2.3). For some $\eta \geq 0$,

$$\tilde{w}_{ij} \sim \begin{cases} N(0,\eta) \text{ if } i=j; \\ N(0,1) \text{ if } i<j. \end{cases} \tag{3.7}$$

In addition, assume that $\tilde{W}$ is symmetric and $\tilde{w}_{ij}$ are independent for $i \leq j$. Note that if $\eta = 2$, then the above distribution is the same as that defined in (2.3). For $\tilde{W}$ defined in (3.7), we consider the statistic $\tilde{T}_{m,p}$. The following law of large numbers is obtained.

Theorem 3 *Suppose* $p \to \infty$ *and that Assumption 2 in (3.2) is satisfied. In addition, assume* $\tilde{W}$ *has the distribution as in (3.7) with* $0 \leq \eta \leq 2$. *Then,*

$$\frac{\tilde{T}_{m,p}}{\sqrt{[4(m-1)+2\eta]\log p}} \to 1 \tag{3.8}$$

in probability as $n \to \infty$.

The law of large numbers in Theorem 3 is proved by bounding the limsup and the liminf of the left hand side of (3.8), and show that the two bounds match. If $\eta > 2$, we are not able to match the two bounds due to technical reasons. We leave it as future work.

Remark 4 A related open question is whether Theorem 1 can be extended to non-Gaussian x_{ij} for the Wishart distribution. We conjecture that with certain assumptions on the moments of x_{ij} and under the asymptotic regime that n is sufficiently large compared to $\log p$ and m, and $\frac{\mathrm{Var}(x_{11}^2)}{\mathrm{Var}(x_{11}x_{12})} \leq 2$, the asymptotic behavior of $\frac{T_{m,n,p}-n}{\sqrt{n}}$ will be similar to that of $\tilde{T}_{m,p}$ as is discussed in Theorem 3. We leave this question for future research, because it requires development of some technical tools that are beyond the scope of the current paper.

Some special cases for this question have been answered in the literature for Wishart matrices with non-Gaussian entries. For example, if $m = 2$, and x_{ij} follows an asymmetric Rademacher distribution $\mathbb{P}(x_{ij} = 1) = p$ and $\mathbb{P}(x_{ij} = -1) = 1 - p$, then it is easy to check

$$W_{\{i,j\}} = \begin{pmatrix} n & \sum_{k=1}^n x_{ki}x_{kj} \\ \sum_{k=1}^n x_{ki}x_{kj} & n \end{pmatrix}$$

and $\lambda_1(W_{\{i,j\}}) = n + |\sum_{k=1}^n x_{ki}x_{kj}|$. As a result, we can see that $T_{m,n,p} = \max_{1\le i<j\le p} \lambda_1(W_{\{i,j\}}) = n + \max_{1\le i<j\le p} |\sum_{k=1}^n x_{ki}x_{kj}|$. Analysis on similar quantities has been studied extensively in the literature including Cai et al. (2013); Cai and Jiang (2012); Fan et al. (2018); Jiang (2004a); Li et al. (2010, 2012); Li and Rosalsky (2006); Shao and Zhou (2014); Zhou (2007). The limiting distributions of $T_{m,n,p}$ are the Gumbel distribution.

3.2. *Application to Construction of Compressed Sensing Matrices.* The main results given above have direct implications for the construction of compressed sensing matrix $X_{n\times p}$ whose entries are i.i.d. $N(0, \frac{1}{n})$. As discussed in the introduction, the goal is to construct the measurement matrix X with the number of measurements n as small as possible relative to p, such that k-sparse signals β can be accurately recovered. For any given t, the RIP framework guarantees accurate recover of all k-sparse signals β if the extreme eigenvalues, $\lambda_{\max}(tk)$ and $\lambda_{\min}(tk)$, of the $tk \times tk$ principal minors of the Wishart matrix $W = X^\intercal X$ satisfy

$$1 - b_*(t) < \lambda_{\min}(tk) \le \lambda_{\max}(tk) < 1 + b_*(t) \tag{3.9}$$

where $b_*(t)$ is given in (1.4).

By setting $m = tk$, $\lambda_{\max}(tk) = T_{m,n,p}/n$, and $\lambda_{\min}(tk) = V_{m,n,p}/n$, it follows from Theorems 1 and Remark 1 that, under Assumption 1 in (3.1),

$$\lambda_{\max}(tk) = 1 + 2\sqrt{\frac{tk\log p}{n}}(1 + o_p(1))$$

and

$$\lambda_{\min}(tk) = 1 - 2\sqrt{\frac{tk\log p}{n}}(1 + o_p(1)).$$

On the other hand, Assumption 1 implies that $\sqrt{\frac{m\log p}{n}} = \sqrt{\frac{tk\log p}{n}} = o(1)$. So the above asymptotic approximation gives $\lambda_{\max}(tk) = 1 + o_p(1)$ and

$\lambda_{\min}(tk) = 1 + o_p(1)$, and hence (3.9) is satisfied. That is, Assumption 1 guarantees the exact recovery of all k sparse signals in the noiseless case through the constrained ℓ_1 minimization as explained in (1.3) and (1.4).

4. Technical Proofs. Throughout the proof, as mentioned earlier, we will let $n \to \infty$ and $p = p_n \to \infty$; the integer $m \geq 2$ is either fixed or $m = m_n \to \infty$. The following notation will be adopted. We write $a_n = O(b_n)$ if there is a constant κ independent of n, p and m (unless otherwise indicated) such that $|a_n| \leq \kappa b_n$. Moreover, we write $a_n = o(b_n)$, if there is a sequence c_n independent of n, p and m such that $c_n \to 0$ and $|a_n| \leq c_n b_n$. Define $\xi_p = \log\log\log p$. This is a sequence growing to infinity with a very slow speed compared to n and p.

This section is organized as follows. We first introduce the main steps in proving Theorems 1 and 2 in Section 4.1. In Section 4.2, we present the proofs for Theorems 1-3, Corollary 1, and Remark 3. The proofs for all technical lemmas are given in the Appendix. For reader's convenience, we list the content of each section below.

Section 4.1. The Strategy of the Proofs for Theorems 1 and 2.
Section 4.2. Proof of the results in Section 3.
  Section 4.2.1. Proof of Theorem 2.
  Section 4.2.2. Proof of Theorem 1.
  Section 4.2.3. Proofs of Theorem 3 and Remark 3.

4.1. *The Strategy of the Proofs for Theorems 1 and 2.* We first explain the proof strategy for Theorem 2 and then explain that for Theorem 1, since Wigner matrices have simpler structure than Wishart matrices. The proof of Theorem 2 consists of three steps. The first step is to find an upper bound on the right tail probability $\mathbb{P}(\tilde{T}_{m,p} \geq 2\sqrt{m\log p}+t)$ for $t \geq \delta$. Our method here is to first develop a moderate deviation bound of $\mathbb{P}(\lambda_1(\tilde{W}_S) \geq 2\sqrt{m\log p}+t)$ for each $S \subset \{1, ..., p\}$ and $|S| = m$, and then use the union bound to control $\mathbb{P}(\tilde{T}_{m,p} \geq 2\sqrt{m\log p} + t)$. The second step is to find an upper bound on the left tail probability $\mathbb{P}(\tilde{T}_{m,p} \leq 2\sqrt{m\log p} - t)$ for $t \geq \delta$. Our approach is to construct a sequence of events $E_{p,m}$ with high probability, such that when $E_{p,m}$ occurs, there exists $S \subset \{1, ..., p\}$ satisfying $|S| = m$ and $\lambda_1(\tilde{W}_S) \geq 2\sqrt{m\log p} - t$. The third step is to combine the left and right tail bounds obtained from the previous two steps to show (3.5).

The proof of Theorem 1 is based on a similar strategy to that of Theorem 2. A new and key ingredient is to control the approximation speed of the Wishart matrix to the Wigner matrix (after normalization). Change-of-measure arguments are used to quantify the approximation speed in the moderate deviation domain.

We point out that the proof for the asymptotic lower bound of $\tilde{T}_{m,p}$ in this paper is different from the standard technique for analyzing the maximum/minimum statistic for a large random matrix (see, e.g. Jiang (2004a)). In particular, the proof in Jiang (2004a) employs the Chen-Stein's Poisson approximation method [see, e.g., Arratia et al. (1990)] and the asymptotic independence. However, this method does not fit our problem. For this reason, new technique are developed and, in particular, we *construct* an event on which $\tilde{T}_{m,p}$ achieves the asymptotic lower bound.

4.2. *Proof of the results in Section 3.* As mentioned earlier, Wigner matrices have simpler structure than Wishart matrices. Thus, we first present the proof of Theorem 2, followed by the proof of Theorem 1. At the end of the section, the proofs of Corollary 1, Theorem 3 and Remark 3 are presented.

In each proof we will need auxiliary results. To make the proof clearer, we place the proofs of the auxiliary results in the Appendix. Sometimes a statement or a formula holds as n is sufficiently large. We will not say "as n is sufficiently large" if the context is apparent.

4.2.1. *Proof of Theorem 2.* To prove Theorem 2, we need the following two key results.

Proposition 1 *Suppose Assumption 2 in (3.2) is satisfied. Recall $\tilde{T}_{m,p}$ defined as in (2.4). Then,*

$$\lim_{p\to\infty}\sup_{t\ge\delta} e^{\alpha t}t^2\mathbb{P}\left(\tilde{T}_{m,p}\ge 2\sqrt{m\log p}+t\right)=0$$

for every $\alpha>0$ and every $\delta>0$.

Proposition 2 *Suppose Assumption 2 in (3.2) is satisfied. Recall $\tilde{T}_{m,p}$ defined as in (2.4). Then,*

$$\lim_{p\to\infty}\sup_{t\ge\delta} e^{\alpha t}t^2\mathbb{P}\left(\tilde{T}_{m,p}\le 2\sqrt{m\log p}-t\right)=0$$

for every $\alpha>0$ and every $\delta>0$.

Another auxiliary lemma is need. Its proof is put in the Appendix.

Lemma 1 *Let $Z\ge 0$ be a random variable with $\mathbb{E}[e^{\alpha Z}]<\infty$ for all $\alpha>0$. Then*

$$\mathbb{E}\big[e^{\alpha Z}\mathbf{1}_{\{Z\ge\delta\}}\big]=e^{\alpha\delta}\mathbb{P}(Z\ge\delta)+\alpha\int_\delta^\infty e^{\alpha t}\mathbb{P}(Z>t)dt$$

for every $\alpha>0$ and every $\delta>0$.

Proof of Theorem 2. By Propositions 1 and 2, we have

$$\lim_{p\to\infty}\sup_{t\geq\delta} e^{\alpha t}t^2\mathbb{P}\left(\left|\tilde{T}_{m,p}-2\sqrt{m\log p}\right|\geq\delta\right)=0$$

for any $\alpha>0$ and $\delta>0$. Consequently, for given $\alpha>0$, there exists a sequence of positive numbers $a_p\to 0$ such that

$$e^{\alpha t}t^2\mathbb{P}\left(\left|\tilde{T}_{m,p}-2\sqrt{m\log p}\right|\geq t\right)\leq a_p \tag{4.1}$$

for all $t\geq\delta$ as p is sufficiently large. Now we estimate

$$\mathbb{E}(e^{\alpha|\tilde{T}-2\sqrt{m\log p}|}\mathbf{1}_{\{|\tilde{T}-2\sqrt{m\log p}|\geq\delta\}}).$$

Apply Lemma 1 to $Z_{p,m}=|\tilde{T}-2\sqrt{m\log p}|$, we see

$$\begin{aligned}&\mathbb{E}\left[e^{\alpha|\tilde{T}_{m,p}-2\sqrt{m\log p}|}\mathbf{1}_{|\tilde{T}_{m,p}-2\sqrt{m\log p}|\geq\delta}\right]\\ =&e^{\alpha\delta}\mathbb{P}(Z_{p,m}\geq\delta)+\int_\delta^\infty e^{\alpha t}\mathbb{P}(Z_{p,m}\geq t)dt.\end{aligned}$$

According to (4.1), the above display can be bounded from above by $\delta^{-2}a_p+a_p\int_\delta^\infty t^{-2}dt$, which tends to 0 as $p\to\infty$. The proof is then complete. ■

Now we proceed to prove Propositions 1 and 2.

Proof of Proposition 1. For any $t>0$, we have from the definition of $\tilde{T}_{m,p}$ that

$$\begin{aligned}\mathbb{P}\left(\tilde{T}_{m,p}\geq 2\sqrt{m\log p}+t\right)\leq&\sum_{S\subset\{1,\ldots,p\},|S|=m}\mathbb{P}\Big(\lambda_1(\tilde{W}_S)\geq 2\sqrt{m\log p}+t\Big)\\ \leq& p^m\mathbb{P}\Big(\lambda_1(\tilde{W}_{\{1,\ldots,m\}})\geq 2\sqrt{m\log p}+t\Big),\end{aligned} \tag{4.2}$$

where the first inequality is due to the union bound and the definition of $\tilde{T}_{m,p}=\max_{S\subset\{1,\ldots,p\},|S|=m}\lambda_1(\tilde{W}_S)$, and in the second inequality we use the fact that W_S are identically distributed for all different S with $|S|=m$. The following result enables us to bound the last probability.

Lemma 2 *Let $\tilde{W}_{\{1,\ldots,m\}}$ be defined as above (2.4) with $S=\{1,...,m\}$. Then there is a constant $\kappa>0$ such that*

$$P\Big(\lambda_1(W_{\{1,\ldots,m\}})\geq x \text{ or } \lambda_m(W_{\{1,\ldots,m\}})\leq -x\Big)\leq e^{-(x^2/4)+\kappa m\log x}$$

for all $x>4\sqrt{m}$ and all $m\geq 2$.

Taking $x := 2\sqrt{m\log p} + t$ in the above lemma, we know $x > 4\sqrt{m}$ as n is large enough, and hence

$$\begin{aligned}
&\log\left[e^{\alpha t}p^m\mathbb{P}\left(\lambda_1(\tilde{W}_{\{1,\dots,m\}}) \geq 2\sqrt{m\log p}+t\right)\right]\\
\leq&\alpha t + m\log p - \frac{1}{4}\left(2\sqrt{m\log p}+t\right)^2 + \kappa m\log\left(2\sqrt{m\log p}+t\right)\\
=&\alpha t - t\sqrt{m\log p} - \frac{1}{4}t^2\\
&\quad + \kappa m\log\left(2\sqrt{m\log p}\right) + \kappa m\log\left(1+\frac{t}{2\sqrt{m\log p}}\right).
\end{aligned}$$

Note that $-\frac{1}{4}t^2 \leq 0$, $\kappa m\log(2\sqrt{m\log p}) = O(m\log\log p)$, and $\kappa m\log(1+\frac{t}{2\sqrt{m\log p}}) = O(\frac{\sqrt{m}}{\sqrt{\log p}}t) < t$ as p is sufficiently large. Thus, the above inequality further implies

$$\begin{aligned}
&\log\left[e^{\alpha t}p^m\mathbb{P}\left(\lambda_1(\tilde{W}_{\{1,\dots,m\}}) \geq 2\sqrt{m\log p}+t\right)\right]\\
\leq\ & -\frac{t}{2}\sqrt{m\log p} + O\left(m\log\log p\right)
\end{aligned} \tag{4.3}$$

uniformly for all $t \geq 0$ as p sufficiently large, where $\alpha > 0$ is fixed. With the above inequality, we complete the proof. ■

Proof of Proposition 2. Recall $\xi_p = \log\log\log p$. The proof will be evidently finished if the following two limits hold. For each $\alpha > 0$ and each $\delta > 0$,

$$\lim_{p\to\infty}\sup_{\delta\leq t\leq 2\sqrt{m\log p}-m\xi_p} e^{\alpha t}t^2\mathbb{P}\left(\tilde{T}_{m,p} \leq 2\sqrt{m\log p}-t\right) = 0 \tag{4.4}$$

and

$$\lim_{p\to\infty}\sup_{t\geq 2\sqrt{m\log p}-m\xi_p} e^{\alpha t}t^2\mathbb{P}\left(\tilde{T}_{m,p} \leq 2\sqrt{m\log p}-t\right) = 0. \tag{4.5}$$

We now verify the above two limits.

The proof of (4.4). Recall

$$\lambda_1(A) \geq \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k}a_{ij} \tag{4.6}$$

for any $k\times k$ square and symmetric matrix $A = (a_{ij})_{1\leq i,j\leq k}$, where $\lambda_1(A)$ is the largest eigenvalue of A.

For each $S \subset \{1, ..., p\}$ such that $|S| = m$ and $\tilde{W}_S = (\tilde{W}_{ij})$, set

$$\tilde{A}_S = \left\{ \tilde{W}_{ij} \geq (1 - \varepsilon_{m,p,t}) \sqrt{\frac{4 \log p}{m}} \text{ for all } i, j \in S \text{ and } i \leq j \right\}, \tag{4.7}$$

where $\varepsilon_{m,p,t} := (4m \log p)^{-1/2} t$. If $0 < t \leq 2\sqrt{m \log p} - m\xi_p$ then $0 < \varepsilon_{m,p,t} < 1$. According to (4.6), if there exists $S_0 \subset \{1, ..., p\}$ such that $|S_0| = m$ and $\tilde{A}_{S_0}$ occurs, then $\tilde{T}_{m,p} \geq \lambda_1(\tilde{W}_{S_0}) \geq m(1 - \varepsilon_{m,p,t}) \sqrt{\frac{4 \log p}{m}} = 2\sqrt{m \log p} - t$. Define $\tilde{Q}_{m,p} = \sum_{S \subset \{1,...,p\} \colon |S| = m} \mathbf{1}_{\tilde{A}_S}$, where $\mathbf{1}_{\tilde{A}_S}$ is the indicator function of $\tilde{A}_S$. Then,

$$\mathbb{P}\left(\tilde{T}_{m,p} < 2\sqrt{m \log p} - t\right) \leq \mathbb{P}\left(\tilde{Q}_{m,p} = 0\right). \tag{4.8}$$

For any random variable Y with $\mathbb{E}Y > 0$ and $\mathbb{E}(Y^2) < \infty$, we have

$$\mathbb{P}(Y \leq 0) \leq \mathbb{P}((Y - \mathbb{E}Y)^2 \geq (\mathbb{E}Y)^2) \leq \frac{Var(Y)}{(\mathbb{E}Y)^2}. \tag{4.9}$$

Applying this inequality to $\tilde{Q}_{m,p}$, we obtain

$$\mathbb{P}\left(\tilde{Q}_{m,p} = 0\right) = \mathbb{P}\left(\tilde{Q}_{m,p} \leq 0\right) \leq \frac{Var(\tilde{Q}_{m,p})}{(\mathbb{E}\tilde{Q}_{m,p})^2}. \tag{4.10}$$

We proceed to find a lower bound on $\mathbb{E}\tilde{Q}_{m,p}$ and an upper bound on $Var(\tilde{Q}_{m,p})$ in two steps.

Step 1: the estimate of $\mathbb{E}\tilde{Q}_{m,p}$. Note that $\mathbf{1}_{\tilde{A}_S}$ are identically (not independently) distributed Bernoulli variables for different S with success rate $\mathbb{P}(\tilde{A}_{\{1,...,m\}})$. Thus, we have

$$\mathbb{E}\tilde{Q}_{m,p} = \binom{p}{m} \mathbb{P}(\tilde{A}_{S_0}), \tag{4.11}$$

where we choose $S_0 = \{1, ..., m\}$ with a bit abuse of notation. For convenience, write $\tau_{m,p,t} = (1 - \varepsilon_{m,p,t}) \sqrt{\frac{4 \log p}{m}} = \sqrt{\frac{4 \log p}{m}} - \frac{t}{m}$. Since the upper triangular entries of $\tilde{W}$ are independent Gaussian variables, we have from (4.7) that

$$\mathbb{P}(\tilde{A}_{S_0}) = \prod_{k=1}^{m} \mathbb{P}\left(\tilde{W}_{kk} \geq \tau_{m,p,t}\right) \prod_{1 \leq i < j \leq m} \mathbb{P}\left(\tilde{W}_{ij} \geq \tau_{m,p,t}\right).$$

Recall that $\tilde{W}_{kk} \sim N(0,2)$ and $\tilde{W}_{ij} \sim N(0,1)$ for $i \neq j$. Hence

$$\mathbb{P}(\tilde{A}_{S_0}) = \bar{\Phi}\left(\frac{1}{\sqrt{2}}\tau_{m,p,t}\right)^m \bar{\Phi}\left(\tau_{m,p,t}\right)^{\frac{m(m-1)}{2}}, \tag{4.12}$$

where $\bar{\Phi}(z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw$. It is well known that

$$\log \bar{\Phi}(x) = -\frac{x^2}{2} - \log(x) - \log\sqrt{2\pi} + o(1) \tag{4.13}$$

as $x \to \infty$. Recall the assumption that $t \le 2\sqrt{m\log p} - m\xi_p$, so $\tau_{m,p,t} = \sqrt{\frac{4\log p}{m}} - \frac{t}{m} \ge \xi_p \to \infty$. Thus, by (4.12) and (4.13),

$$\begin{aligned}\log \mathbb{P}(\tilde{A}_{S_0}) = &-\frac{1}{2}m \cdot \frac{1}{2} \cdot \tau_{m,p,t}^2 - m\log\frac{\tau_{m,p,t}}{\sqrt{2}} \\ &- \frac{1}{2}\cdot\frac{m(m-1)}{2}\left(\tau_{m,p,t}\right)^2 - \frac{m(m-1)}{2}\log\left(\tau_{m,p,t}\right) + O(m^2).\end{aligned}$$

Note that $1 > 1-\varepsilon_{m,p,t} \ge \xi_p\sqrt{\frac{m}{4\log p}}$ since $0 < t \le 2\sqrt{m\log p} - m\xi_p$. It follows that $|\log(1-\varepsilon_{m,p,t})| = O\left(\log\sqrt{\frac{\log p}{m\xi_p^2}}\right) = O(\log\log p)$. Also, $\log\sqrt{\frac{4\log p}{m}} = O(\log\log p)$. As a result, from the definition of $\tau_{m,p,t}$ we have $\tau_{m,p,t}^2 = (1-\varepsilon_{m,p,t})^2 \cdot \frac{4\log p}{m}$ and $\log\left(\tau_{m,p,t}\right) = O(\log\log p)$. It follows that

$$\log \mathbb{P}(\tilde{A}_{S_0}) \;=\; -(1-\varepsilon_{m,p,t})^2 m\log p + O(m^2\log\log p). \tag{4.14}$$

Combining this with (4.11), we see

$$\log(\mathbb{E}\tilde{Q}_{m,p}) = \log\binom{p}{m} - (1-\varepsilon_{m,p,t})^2 m\log p + O(m^2\log\log p). \tag{4.15}$$

To control $\binom{p}{m}$, we need the next result, which will be proved in the Appendix.

Lemma 3 *For all $m \ge p \ge 1$, $m\log p - m\log m \le \log\binom{p}{m} \le m\log p + m - m\log m$.*

Using the above lemma, (4.15), and note that $m\log m = O(m^2\log\log p)$, we have

$$\log(\mathbb{E}\tilde{Q}_{m,p}) = \left[1-(1-\varepsilon_{m,p,t})^2\right] m\log p + O(m^2\log\log p). \tag{4.16}$$

Step 2: the estimate of $Var(\tilde{Q}_{m,p})$. Recalling $\tilde{Q}_{m,p} = \sum_{S\subset\{1,\ldots,p\}:\,|S|=m} \mathbf{1}_{\tilde{A}_S}$, we have

$$
\begin{aligned}
Var(\tilde{Q}_{m,p}) &= \mathbb{E}\tilde{Q}^2_{m,p} - (\mathbb{E}\tilde{Q}_{m,p})^2 \\
&= \sum_{S_1,S_2\subset\{1,..,p\},|S_1|=|S_2|=m} \mathbb{P}(\tilde{A}_{S_1}\cap\tilde{A}_{S_2}) - (\mathbb{E}\tilde{Q}_{m,p})^2. \qquad (4.17)
\end{aligned}
$$

Note that $\mathbb{P}(\tilde{A}_{S_1}\cap\tilde{A}_{S_2})$ is determined by $|S_1\cap S_2|$ and m. By (4.7),

(4.18)

$$
\begin{aligned}
&\sum_{S_1,S_2\subset\{1,..,p\},|S_1|=|S_2|=m} \mathbb{P}\left(\tilde{A}_{S_1}\cap\tilde{A}_{S_2}\right) \\
=&\sum_{l=0}^{m}\sum_{|S_1\cap S_2|=l,|S_1|=|S_2|=m} \mathbb{P}\left(\tilde{A}_{S_1}\cap\tilde{A}_{S_2}\right) \\
=&\frac{p!}{m!m!(p-2m)!}\mathbb{P}\left(\tilde{A}_{\{1,\ldots,m\}}\right)^2 + \binom{p}{m}\mathbb{P}\left(\tilde{A}_{\{1,\ldots,m\}}\right) \\
&+\sum_{l=1}^{m-1}\frac{p!}{l!(m-l)!(m-l)!(p-2m+l)!}\mathbb{P}\left(\tilde{A}_{\{1,\ldots,m\}}\cap\tilde{A}_{\{1,\ldots,l,m+1,\ldots,2m-l\}}\right).
\end{aligned}
$$

On the other hand, $\mathbb{E}\tilde{Q}_{m,p} = \binom{p}{m}P\big(\tilde{A}_{\{1,\ldots,m\}}\big)$ and hence

$$
(\mathbb{E}\tilde{Q}_{m,p})^2 = \frac{p!}{m!m!(p-2m)!}P\left(\tilde{A}_{\{1,\ldots,m\}}\right)^2\cdot\frac{p!(p-2m)!}{(p-m)!^2}. \qquad (4.19)
$$

Combining (4.17), (4.18) and (4.19), we arrive at $Var(\tilde{Q}_{m,p}) = (\mathbb{E}\tilde{Q}_{m,p})^2\big(\frac{(p-m)!^2}{p!(p-2m)!} - 1\big) + \mathbb{E}\tilde{Q}_{m,p} + \sum_{l=1}^{m-1}\frac{p!}{l!(m-l)!(m-l)!(p-2m+l)!}\mathbb{P}\big(\tilde{A}_{\{1,\ldots,m\}}\cap\tilde{A}_{\{1,\ldots,l,m+1,\ldots,2m-l\}}\big)$. Observe that $\frac{p!}{(p-2m+l)!} = p(p-1)\cdots(p-2m+l-1) \le p^{2m-l}$ and $\frac{1}{l!(m-l)!(m-l)!} \le 1$. It follows that

(4.20)

$$
\begin{aligned}
Var(\tilde{Q}_{m,p}) \le&\mathbb{E}\tilde{Q}_{m,p} + (\mathbb{E}\tilde{Q}_{m,p})^2\big(\frac{(p-m)!^2}{p!(p-2m)!} - 1\big) \\
&+ m\max_{l=1,\ldots,m-1} p^{2m-l}\mathbb{P}\big(\tilde{A}_{\{1,\ldots,m\}}\cap\tilde{A}_{\{1,\ldots,l,m+1,\ldots,2m-l\}}\big).
\end{aligned}
$$

Similar to (4.12) we have

(4.21)

$$
\mathbb{P}\Big(\tilde{A}_{\{1,\ldots,m\}}\cap\tilde{A}_{\{1,\ldots,l,m+1,\ldots,2m-l\}}\Big) =\bar{\Phi}\Big(\frac{1}{\sqrt{2}}\tau_{m,p,t}\Big)^{2m-l}\bar{\Phi}\big(\tau_{m,p,t}\big)^{\frac{m(m-1)}{2}\cdot 2-\frac{l(l-1)}{2}}.
$$

Again, we find an approximation for the above display by using (4.13) and simplifying it. We arrive at $\log \mathbb{P}\left(\tilde{A}_{\{1,\dots,m\}} \cap \tilde{A}_{\{1,\dots,l,m+1,\dots,2m-l\}}\right) \le -\frac{1}{m}(2m^2-l^2)(1-\varepsilon_{m,p,t})^2 \log p + O(m^2 \log\log p)$. Therefore, for the last term in (4.20), we see

$$\begin{aligned}&\log\left[mp^{2m-l}\mathbb{P}\big(\tilde{A}_{\{1,\dots,m\}} \cap \tilde{A}_{\{1,\dots,l,m+1,\dots,2m-l\}}\big)\right]\\ \le&\big[2m-l-\frac{1}{m}(1-\varepsilon_{m,p,t})^2(2m^2-l^2)\big]\log p + O(m^2\log\log p).\end{aligned} \tag{4.22}$$

The following lemma enables us to evaluate the coefficient of $\log p$.

Lemma 4 *For any* $0<\varepsilon<1$ *and* $m\ge 2$, $\max_{l=1,\dots,m-1}\big\{(2m-l)-\frac{2m^2-l^2}{m}(1-\varepsilon)^2\big\} = (2m-1)-\big(2m-\frac{1}{m}\big)(1-\varepsilon)^2 = 2m\big[1-(1-\varepsilon)^2\big] - \big[1-\frac{1}{m}(1-\varepsilon)^2\big]$.

Applying the above lemma to (4.22), we get

$$\begin{aligned}&m\max_{l=1,\dots,m-1} p^{2m-l}\mathbb{P}\Big(\tilde{A}_{\{1,\dots,m\}} \cap \tilde{A}_{\{1,\dots,l,m+1,\dots,2m-l\}}\Big)\\ \le&\exp\Big\{2m\big[1-(1-\varepsilon_{m,p,t})^2\big]\log p\\ &-\big[1-\frac{1}{m}(1-\varepsilon_{m,p,t})^2\big]\log p + O(m^2\log\log p)\Big\}.\end{aligned}$$

This inequality together with (4.16) implies that

$$\begin{aligned}&\big(\mathbb{E}\tilde{Q}_{m,p}\big)^{-2} m\max_{l=1,\dots,m-1} p^{2m-l}\mathbb{P}\big(\tilde{A}_{\{1,\dots,m\}} \cap \tilde{A}_{\{1,\dots,l,m+1,\dots,2m-l\}}\big)\\ \le&\exp\Big\{-\big[1-\frac{1}{m}(1-\varepsilon_{m,p,t})^2\big]\log p + O(m^2\log\log p)\Big\}.\end{aligned} \tag{4.23}$$

Combining the above display with (4.20), we arrive at

$$\begin{aligned}\frac{Var(\tilde{Q}_{m,p})}{\big(\mathbb{E}\tilde{Q}_{m,p}\big)^2} \le&\exp\Big\{-\big[1-\frac{1}{m}(1-\varepsilon_{m,p,t})^2\big]\log p + O(m^2\log\log p)\Big\}\\ &+\big(\mathbb{E}\tilde{Q}_{m,p}\big)^{-1} + \frac{(p-m)!^2}{p!(p-2m)!} - 1.\end{aligned}$$

It is not hard to show that for all integers $p\ge m\ge 1$ satisfying $2m<p$,

$$\frac{(p-m)!^2}{p!(p-2m)!} < 1. \tag{4.24}$$

Therefore,

$$\frac{Var(\tilde{Q}_{m,p})}{(\mathbb{E}\tilde{Q}_{m,p})^2} \le \exp\Big\{ -\big[1-\frac{1}{m}(1-\varepsilon_{m,p,t})^2\big]\log p + O(m^2\log\log p)\Big\} + \big(\mathbb{E}\tilde{Q}_{m,p}\big)^{-1}. \tag{4.25}$$

We now study the last two terms one by one. For $m \ge 2$,

$$-\big[1-\frac{1}{m}(1-\varepsilon_{m,p,t})^2\big]\log p + O(m^2\log\log p) \le -\frac{1}{4}\log p \tag{4.26}$$

for n sufficiently large under Assumption 2 in (3.2). Recalling $\varepsilon_{m,p,t} = (4m\log p)^{-1/2}t$, we see from (4.16) that

$$\begin{aligned} \log\big(\mathbb{E}\tilde{Q}_{m,p}\big)^{-1} &= -\big[1-(1-\varepsilon_{m,p,t})^2\big] m\log p + O(m^2\log\log p) \\ &\le -\frac{t}{2}\sqrt{m\log p} + O(m^2\log\log p). \end{aligned} \tag{4.27}$$

Combining (4.25), (4.26) and (4.27), we arrive at

$$\frac{Var(\tilde{Q}_{m,p})}{(\mathbb{E}\tilde{Q}_{m,p})^2} \le \exp\Big\{-\frac{t}{2}\sqrt{m\log p} + O(m^2\log\log p)\Big\} + \exp\Big\{-\frac{1}{4}\log p\Big\}.$$

This together with (4.8) and (4.10) yields $\mathbb{P}\big(\tilde{T}_{m,p} \le 2\sqrt{m\log p} - t\big) \le \exp\big\{-\frac{t}{2}\sqrt{m\log p} + O(m^2\log\log p)\big\} + \frac{1}{p^{1/4}}$ uniformly for all $\delta \le t \le 2\sqrt{m\log p} - m\xi_p$. Consequently, we get (4.4).

The proof of (4.5). For any $S \subset \{1,...,p\}$ with $|S| = m$, write $\tilde{W}_S = (\tilde{W}_{ij})_{i,j\in S}$. Note that $\lambda_1(\tilde{W}_S) \ge \max_{i\in S}\tilde{W}_{ii}$. Thus, $\tilde{T}_{m,p} \ge \max\limits_{S\subset\{1,...,p\},|S|=m}\lambda_1(\tilde{W}_S) \ge \max\limits_{1\le i\le p}\tilde{W}_{ii}$. As a result,

$$\begin{aligned} \mathbb{P}\Big(\tilde{T}_{m,p} \le 2\sqrt{m\log p} - t\Big) &\le \mathbb{P}\Big(\max_{1\le i\le p}\tilde{W}_{ii} \le 2\sqrt{m\log p} - t\Big) \\ &= \Phi\Big(\sqrt{2m\log p} - \frac{1}{\sqrt{2}}t\Big)^p, \end{aligned}$$

where the function $\Phi(z) = \int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{s^2}{2}}ds$ for $z \in \mathbb{R}$. To proceed, we discuss two scenarios: $2\sqrt{m\log p} - m\xi_p \le t \le 4\sqrt{m\log p}$ and $t > 4\sqrt{m\log p}$. For

$2\sqrt{m\log p} - m\xi_p \le t \le 4\sqrt{m\log p}$, we have

$$\begin{aligned}\Phi\Big(\sqrt{2m\log p} - \frac{t}{\sqrt{2}}\Big)^p \le & \Phi\Big(\sqrt{2m\log p} - \frac{2\sqrt{m\log p} - m\xi_p}{\sqrt{2}}\Big)^p \\ = & \exp\Big\{p\log\big(1 - \bar{\Phi}(\frac{m\xi_p}{\sqrt{2}})\big)\Big\} \\ \le & \exp\Big\{-p\bar{\Phi}(\frac{m\xi_p}{\sqrt{2}})\Big\},\end{aligned}$$

where $\bar{\Phi}(z) = 1 - \Phi(z)$ for any $z \in \mathbb{R}$ and the inequality $\log(1-x) \le -x$ for any $x < 1$ is used in the last step. Note $\bar{\Phi}\left(\frac{1}{\sqrt{2}}m\xi_p\right) = (1 + o(1))\frac{1}{\sqrt{4\pi}m\xi_p}e^{-\frac{m^2\xi_p{}^2}{4}}$ and $p^{0.1}(\xi_p)^{-1}e^{-\frac{m^2\xi_p{}^2}{4}} \to \infty$ since $\xi_p = \log\log\log p$. Thus,

$$\Phi\Big(\sqrt{2m\log p} - \frac{t}{\sqrt{2}}\Big)^p \le \exp\Big\{-p^{0.9}m\Big\}, \tag{4.28}$$

for sufficiently large p. This further implies

$$\lim_{p\to\infty} \sup_{2\sqrt{m\log p} - m\xi_p \le t \le 4\sqrt{m\log p}} e^{\alpha t}t^2\mathbb{P}\left(\tilde{T}_{m,p} \le 2\sqrt{m\log p} - t\right) = 0 \tag{4.29}$$

for any $\alpha > 0$. Note that $\Phi(-x) = \bar{\Phi}(x) \le \frac{1}{\sqrt{2\pi}\,x}e^{-x^2/2} \le e^{-x^2/2}$ for any $x \ge 1$. Then, for the other scenario where $t \ge 4\sqrt{m\log p}$, we have $\Phi\big(\sqrt{2m\log p} - \frac{t}{\sqrt{2}}\big)^p \le \Phi\big(-\frac{t}{2\sqrt{2}}\big)^p \le \exp\big\{-\frac{pt^2}{16}\big\}$ as n is large enough. Thus,

$$\lim_{p\to\infty} \sup_{t \ge 4\sqrt{m\log p}} e^{\alpha t}t^2\mathbb{P}\left(\tilde{T}_{m,p} \le 2\sqrt{m\log p} - t\right) = 0 \tag{4.30}$$

for any $\alpha > 0$. Joining (4.29) and (4.30), we see (4.5). This completes the whole proof. ∎

4.2.2. *Proof of Theorem 1.* To prove Theorem 1, we need the following two propositions.

Proposition 3 *Suppose Assumption 1 in (3.1) holds. Recall $T_{m,n,p}$ defined as in (2.1). Then,* $\lim_{n\to\infty}\sup_{t\ge\delta} e^{\alpha t}t^2\mathbb{P}\left(\frac{1}{\sqrt{n}}(T_{m,n,p} - n) \ge 2\sqrt{m\log p} + t\right) = 0$ *for any $\alpha > 0$ and $\delta > 0$.*

Proposition 4 *Suppose Assumption 1 in (3.1) holds. Recall $T_{m,n,p}$ defined as in (2.1). Then,* $\lim_{n\to\infty}\sup_{t\ge\delta} e^{\alpha t}t^2\mathbb{P}\left(\frac{1}{\sqrt{n}}(T_{m,n,p} - n) \le 2\sqrt{m\log p} - t\right) = 0$ *for any $\alpha > 0$ and $\delta > 0$.*

Proof of Theorem 1. Similar to the proof of Theorem 2, it is sufficient to prove (3.3). By the same argument as in the proof of Theorem 2, with the upper bound for $\mathbb{P}(\frac{T_{m,n,p}-n}{\sqrt{n}} \geq 2\sqrt{m\log p}+t)$ given in Proposition 3 and the upper bound for $\mathbb{P}(\frac{T_{m,n,p}-n}{\sqrt{n}} \leq 2\sqrt{m\log p}-t)$ for $t>\delta$ given in Proposition 4, we get (3.3). ■

In the following we start to prove Propositions 3 and 4.

Proof of Proposition 3. Without loss of generality, we assume $\delta < 1$ since the expectation in (3.3) is monotonically decreasing in δ.

Let $W_{\{1,\dots,m\}}$ be as W_S above (2.1) with $S=\{1,2,\cdots,m\}$. Analogous to (4.2), we have

$$\begin{aligned}&\mathbb{P}\Big(\frac{1}{\sqrt{n}}(T_{m,n,p}-n)\geq 2\sqrt{m\log p}+t\Big)\\ \leq& p^m\mathbb{P}\Big(\frac{1}{\sqrt{n}}(\lambda_1(W_{\{1,\dots,m\}})-n)\geq 2\sqrt{m\log p}+t\Big).\end{aligned}\tag{4.31}$$

We now bound the last probability. Since the above tail probability involve moderate bound and large deviation bound for different ranges of t, we will discuss three different cases and use different proof strategies. Recall $\xi_p=\log\log\log p$. Set

$$\omega_n=\Big(\frac{m}{\log p}\Big)^{1/2}\xi_p\log n.\tag{4.32}$$

The three cases are: (1) $t>\frac{\delta\sqrt{n}}{100}$, (2) $\delta\vee\omega_n\leq t\leq\frac{\delta\sqrt{n}}{100}$, and (3) $\delta\leq t<\delta\vee\omega_n$. They cover all situations for $t\geq\delta$. For the first two cases, the upper bound is based on the next lemma, which gives a moderate deviation bound for the spectrum of $\frac{1}{\sqrt{n}}W_{\{1,\dots,m\}}$ from the identity matrix I_m.

Lemma 5 *There exists a constant $\kappa>0$ such that for all n,p,m, $r\geq 1$, $0<d<1/2$ and $y>2dmr$, we have*

$$\begin{aligned}&\mathbb{P}\Big(\frac{\lambda_1(W_{\{1,\dots,m\}})-n}{n}\geq y\Big)\\ \leq& 2\cdot\exp\Big\{-nI(1+y-2dmr)+\kappa m\log\frac{1}{d}\Big\}+2\cdot e^{-mnI(r)}\end{aligned}\tag{4.33}$$

and

$$\begin{aligned}&\mathbb{P}\Big(\frac{\lambda_m(W_{\{1,\dots,m\}})-n}{n}\leq -y\Big)\\ \leq& 2\cdot\exp\Big\{-nI(1-y+2dmr)+\kappa m\log\frac{1}{d}\Big\}+2\cdot e^{-mnI(r)}\end{aligned}\tag{4.34}$$

where $I(s)=\frac{1}{2}(s-1-\log s)$ for $s>0$ and $I(s)=\infty$ for $s\leq 0$.

Case 1: $t > \frac{\delta\sqrt{n}}{100}$**.** Let $\alpha > 0$ be given. Choose $r = \max(2, 1 + \frac{80\alpha t}{mn})$, $d = \min(\frac{1}{2}, \frac{t}{4m\sqrt{nr}})$, and $y = \frac{2\sqrt{m\log p}+t}{\sqrt{n}}$ in Lemma 5. The choice of r, d, and y satisfies that $2dmr \le \frac{t}{2\sqrt{n}}$ and hence $y - 2dmr \ge \frac{2\sqrt{m\log p}}{\sqrt{n}} + \frac{t}{2\sqrt{n}}$. Set $z = \frac{2\sqrt{m\log p}}{\sqrt{n}} + \frac{t}{2\sqrt{n}}$. Notice that $I(s)$ from Lemma 5 is increasing for $s \ge 1$. Then, by the lemma,

$$\begin{aligned} & t^2 e^{\alpha t} p^m \mathbb{P}\Big(\frac{\lambda_1(W_{\{1,\dots,m\}}) - n}{\sqrt{n}} \ge 2\sqrt{m\log p} + t\Big) \\ \le & 2\cdot\exp\Big\{-\frac{n}{2}[z - \log(1+z)] + \kappa m \log\frac{1}{d} + \alpha t + 2\log t + m\log p\Big\} \\ & + 2\cdot\exp\Big\{-\frac{1}{2}(r - 1 - \log r)mn + \alpha t + 2\log t + m\log p\Big\}. \end{aligned} \tag{4.35}$$

The following lemma says that both of the last two terms go to zero.

Lemma 6 *Suppose Assumption 1 in (3.1) holds. Let $\alpha > 0$ and $\delta > 0$ be given. For $r = \max(2, 1+\frac{80\alpha t}{mn})$, $d = \min(\frac{1}{2}, \frac{t}{4m\sqrt{nr}})$ and $z = \frac{2\sqrt{m\log p}}{\sqrt{n}} + \frac{t}{2\sqrt{n}}$, we have*

$$\lim_{n\to\infty} \sup_{t > \frac{\delta\sqrt{n}}{100}} \exp\Big\{-\frac{n}{2}[z - \log(1+z)] + \kappa m\log\frac{1}{d} + \alpha t + 2\log t + m\log p\Big\} = 0, \tag{4.36}$$

$$\lim_{n\to\infty} \sup_{t > \frac{\delta\sqrt{n}}{100}} \exp\Big\{-\frac{1}{2}(r - 1 - \log r)mn + \alpha t + 2\log t + m\log p\Big\} = 0. \tag{4.37}$$

Combining (4.31), (4.35)-(4.37), we conclude

$$\lim_{n\to\infty} \sup_{t > \frac{\delta\sqrt{n}}{100}} t^2 e^{\alpha t}\mathbb{P}\left(\frac{1}{\sqrt{n}}(T_{m,n,p} - n) \ge 2\sqrt{m\log p} + t\right) = 0. \tag{4.38}$$

Case 2: $\delta \vee \omega_n \le t \le \frac{\delta\sqrt{n}}{100}$**.** Review ω_n in (4.32). Now we choose $r = 2$, $d = \frac{t}{8m\sqrt{n}} < \frac{1}{2}$ and $y = \frac{2\sqrt{m\log p}+t}{\sqrt{n}}$. Then $y > \frac{t}{2\sqrt{n}} = 2dmr$. By (4.33),

$$\begin{aligned} & t^2 e^{\alpha t} p^m \mathbb{P}\Big(\frac{\lambda_1(W_{\{1,\dots,m\}}) - n}{\sqrt{n}} \ge 2\sqrt{m\log p} + t\Big) \\ \le & 2\cdot\exp\Big\{-\frac{n}{2}[z - \log(1+z)] + \kappa m \log\frac{1}{d} + \alpha t + 2\log t + m\log p\Big\} \\ & + 2\cdot\exp\Big\{-\frac{1}{2}(1 - \log 2)mn + \alpha t + 2\log t + m\log p\Big\} \end{aligned} \tag{4.39}$$

where $z := y - 2dmr = \frac{2\sqrt{m\log p}}{\sqrt{n}} + \frac{t}{2\sqrt{n}}$. The last two terms are analyzed in the next lemma.

Lemma 7 *Suppose Assumption 1 in* (3.1) *holds. Let* ω_n *be as in* (4.32)*. For* $\delta \vee \omega_n \le t \le \frac{\delta\sqrt{n}}{100}$*,* $z = \frac{2\sqrt{m\log p}}{\sqrt{n}} + \frac{t}{2\sqrt{n}}$ *and* $d = \frac{t}{8m\sqrt{n}}$*, we have*

$$\begin{aligned} & \exp\Big\{-\frac{n}{2}[z-\log(1+z)] + \kappa m \log\frac{1}{d} + \alpha t + 2\log t + m\log p\Big\} \\ \le & \exp\Big\{-\frac{1}{4}t\sqrt{m\log p}\Big\} \end{aligned} \tag{4.40}$$

as n *is sufficiently large. In addition, as* $n \to \infty$*,*

$$\begin{aligned} & -\frac{1}{2}(1-\log 2)mn + \alpha t + 2\log t + m\log p \\ = & -\frac{1}{2}(1-\log 2)[1+o(1)]mn. \end{aligned} \tag{4.41}$$

Joining (4.39)-(4.41), we obtain

$$\lim_{n\to\infty} \sup_{\delta\vee\omega_n \le t \le \frac{\delta\sqrt{n}}{100}} p^m t^2 e^{\alpha t} \mathbb{P}\Big(\frac{\lambda_1(W_{\{1,\dots,m\}}) - n}{\sqrt{n}} \ge 2\sqrt{m\log p} + t\Big) = 0, \tag{4.42}$$

which together with (4.31) implies that

$$\lim_{n\to\infty} \sup_{\delta\vee\omega_n \le t \le \frac{\delta\sqrt{n}}{100}} t^2 e^{\alpha t} \mathbb{P}\left(\frac{1}{\sqrt{n}}(T_{m,n,p} - n) \ge 2\sqrt{m\log p} + t\right) = 0.$$

This completes our analysis for Case 2. By using the same argument as obtaining (4.42), we have the following limit, which will be used later on.

$$\lim_{n\to\infty} \sup_{\delta\vee\omega_n \le t \le \frac{\delta\sqrt{n}}{100}} t^2 e^{\alpha t} p^m \mathbb{P}\left(\frac{\lambda_m(W_{\{1,\dots,m\}}) - n}{\sqrt{n}} \le -2\sqrt{m\log p} - t\right) = 0. \tag{4.43}$$

We next study Case 3.

Case 3: $\delta \le t < \delta \vee \omega_n$. Note that this case is only possible if $n \ge \exp\{((\log p)/m)^{1/2}{\xi_p}^{-1}\delta\}$. We point out that Lemma 5 is not a suitable approach for bounding the tail probability in this case because the term $m\log(1/d)$, which cannot be easily controlled, will dominate the other terms in the error bound for very large n. Instead, we will use another approach to obtain an upper bound of $\mathbb{P}\left(\lambda_1(W_{\{1,\dots,m\}}) \ge 2\sqrt{m\log p} + t\right)$. The main step here is to quantify the approximation of the extreme eigenvalue of a Wishart matrix to that of a Wigner matrix. We will analyze their density functions and leverage them with the results in the proof of Theorem 2.

Let $\mu = (\mu_1, ..., \mu_m)$ be the order statistics of the eigenvalues of $W_{\{1,...,m\}}$ such that $\mu_1 > \mu_2 > ... > \mu_m$. Write $\nu = (\nu_1, ..., \nu_m)$ with $\nu_i = (\mu_i - n)/\sqrt{n}$. Let $\tilde{W}_{\{1,...,m\}} = (\tilde{w}_{ij})_{1\le i,j\le m}$ where $\tilde{w}_{ij}$'s are as in (2.3). Let the eigenvalues of $\tilde{W}_{\{1,...,m\}}$ be $\lambda_1 > ... > \lambda_m$. Set $\lambda = (\lambda_1, ..., \lambda_m)$. Intuitively, the law of ν is close to that of λ when n is large. The next lemma quantifies the approximation speed. Review $\|x\|_\infty = \max_{1\le i\le m} |x_i|$ for any $x = (x_1, \cdots, x_m) \in \mathbb{R}^m$.

Lemma 8 *Let $g_{n,m}(\cdot)$ be the density function of ν, and let $h_m(\cdot)$ be the density function of λ. Assume $m^3 = o(n)$. Then,*

$$\begin{aligned} & \log g_{n,m}(v) - \log h_m(v) \\ = \;& o(1) + O\left(m^2 n^{-1/2}\|v\|_\infty + m^2 n^{-1}\|v\|_\infty^2 + mn^{-1/2}\|v\|_\infty^3\right) \end{aligned}$$

for all $v \in \mathbb{R}^m$ with $\|v\|_\infty \le \frac{2}{3}\sqrt{n}$.

Let $r_{m,n} = 2\sqrt{m\log p} + \omega_n$, where ω_n is as in (4.32). Then for t such that $\delta \le t \le \omega_n$,

$$\begin{aligned} & \mathbb{P}\Big(\frac{1}{\sqrt{n}}\big(\lambda_1(W_{\{1,...,m\}}) - n\big) \ge 2\sqrt{m\log p} + t\Big) \\ \le & \mathbb{P}\Big(\frac{1}{\sqrt{n}}\big(\lambda_1(W_{\{1,...,m\}}) - n\big) \ge 2\sqrt{m\log p} + t, \max_{1\le i\le m} |\nu_i| \le r_{m,n}\Big) \\ & \qquad + \mathbb{P}\big(\max_{1\le i\le m} |\nu_i| > r_{m,n}\big). \end{aligned} \tag{4.44}$$

There are three probabilities above, denote the second one by H_n. For H_n, we use the change-of-measure argument. In fact,

$$\begin{aligned} H_n = & \int_{v_1\ge 2\sqrt{m\log p}+t, \|v\|_\infty \le r_{m,n}} g_{n,m}(v)dv \\ = & \int_{v_1\ge 2\sqrt{m\log p}+t, \|v\|_\infty \le r_{m,n}} \exp\{\log g_{n,m}(v) - \log h_m(v)\} h_m(v)dv \\ = & \exp\Big\{o(1) + O(m^2 n^{-1/2} r_{m,n}) + O(m^2 n^{-1} r_{m,n}^2) + O(mn^{-1/2} r_{m,n}^3)\Big\} \\ & \qquad \cdot \int_{v_1\ge 2\sqrt{m\log p}+t, \|v\|_\infty \le r_{m,n}} h_m(v)dv \\ \le & 2\cdot \exp\Big\{O(m^2 n^{-1/2} r_{m,n}) + O(m^2 n^{-1} r_{m,n}^2) + O(mn^{-1/2} r_{m,n}^3)\Big\} \\ & \qquad \cdot \mathbb{P}\left(\lambda_1(\tilde{W}_{\{1,...m\}}) \ge 2\sqrt{m\log p} + t\right). \end{aligned}$$

Recall the definition of ω_n in (4.32) and note that $r_{m,n} > \sqrt{m\log p}$ and $m = o(n)$. It is not hard to verify that $O(m^2 n^{-1/2} r_{m,n}) + O(m^2 n^{-1} r_{m,n}^2) +$

$O(mn^{-1/2}r_{m,n}^3) = o(\sqrt{m\log p})$, where Assumption 1 from (3.1) is used. Therefore, $H_n \le \exp\big(o\big(\sqrt{m\log p}\,\big)\big)\cdot\mathbb{P}\left(\lambda_1(\tilde{W}_{\{1,\dots m\}}) \ge 2\sqrt{m\log p}+t\right)$. Note that $t \le e^t$ for all $t > 0$. It follows from (4.3) that $\sup\limits_{t\ge\delta}\big\{p^m e^{\alpha t}t^2 H_n\big\} \le \sup\limits_{t\ge\delta}\exp\big\{-\frac{1}{2}t\sqrt{m\log p}+O(m\log\log p)+o(\sqrt{m\log p}\,)\big\} = o(1)$ under Assumption 1. by the fact $t\ge\delta$ and Assumption 1. Combining this with (4.44), we have

$$\begin{aligned}(4.45)\qquad &\sup_{\delta\le t\le\delta\vee\omega_n}\Big\{p^m e^{\alpha t}t^2\cdot\mathbb{P}\Big(\frac{\lambda_1(W_{\{1,\dots,m\}})-n}{\sqrt{n}}\ge 2\sqrt{m\log p}+t\Big)\Big\}\\ &\le o(1)+p^m e^{\alpha\omega_n+2\log\omega_n}\cdot\mathbb{P}\big(\max_{1\le i\le m}|\nu_i|\ge r_{m,n}\big).\end{aligned}$$

We next analyze $\mathbb{P}\big(\max_{1\le i\le m}|\nu_i|\ge r_{m,n}\big)$. Recall $r_{m,n} = 2\sqrt{m\log p}+\omega_n$, where ω_n is as in (4.32). Recall that we only discuss Case 3 when $\delta\le t<\delta\vee\omega_n$, and this is only meaningful when $\omega_n>\delta$. Thus, $\delta\vee\omega_n=\omega_n\le\frac{\sqrt{n}\delta}{100}$. Thus, from (4.42) we have

$$(4.46)\qquad \lim_{n\to\infty}p^m e^{\alpha\omega_n+2\log\omega_n}\mathbb{P}\left(\frac{\lambda_1(W_{\{1,\dots,m\}})-n}{\sqrt{n}}\ge r_{m,n}\right)=0.$$

By (4.43),

$$(4.47)\qquad \lim_{n\to\infty}p^m e^{\alpha\omega_n+2\log\omega_n}\mathbb{P}\left(\frac{\lambda_m(W_{\{1,\dots,m\}})-n}{\sqrt{n}}\le -r_{m,n}\right)=0.$$

Since $\max_{1\le i\le m}|\nu_i| = \max(\nu_1,-\nu_m)$, by combining (4.46) and (4.47), we see that $\lim_{n\to\infty}p^m e^{\alpha\omega_n+2\log\omega_n}\mathbb{P}\left(\max_{1\le i\le m}|\nu_i|\ge r_{m,n}\right)=0$. Combining this with (4.45), we further have

$$(4.48)\qquad \lim_{n\to\infty}\sup_{\delta\le t\le\delta\vee\omega_n}p^m e^{\alpha t}t^2\mathbb{P}\Big(\frac{1}{\sqrt{n}}(\lambda_1(W_{\{1,\dots,m\}})-n)\ge 2\sqrt{m\log p}+t\Big)=0.$$

This completes our analysis for Case 3.

Now, we combine (4.38), (4.42) and (4.48), and arrive at

$$\lim_{n\to\infty}\sup_{t\ge\delta}p^m e^{\alpha t}t^2\mathbb{P}\left(\frac{1}{\sqrt{n}}(\lambda_1(W_{\{1,\dots,m\}})-n)\ge 2\sqrt{m\log p}+t\right)=0.$$

This and (4.31) conclude $\lim_{n\to\infty}\sup_{t\ge\delta}e^{\alpha t}t^2\mathbb{P}\Big(\frac{1}{\sqrt{n}}(T_{m,n,p}-n)\ge 2\sqrt{m\log p}+t\Big)=0$. ■

Proof of Proposition 4. Noticing the expectation in (3.3) is non-increasing in δ. Without loss of generality, we assume $\delta < 1$. Here we discuss two scenarios that are similar to those in the proof of Theorem 2. They are 1) $\delta \le t \le 2\sqrt{m\log p} - m\xi_p$ and 2) $t > 2\sqrt{m\log p} - m\xi_p$, where $\xi_p = \log\log\log p$.
Scenario 1: $\delta \le t \le 2\sqrt{m\log p} - m\xi_p$. Similar to the proof of Theorem 2, we define the event A_S as follows. For each $S \subset \{1, ..., p\}$ with $|S| = m$, set $A_S = \big\{\frac{1}{\sqrt{n}}(W_{kk} - n) \ge \tau_{m,p,t}, \frac{W_{ij}}{\sqrt{n}} \ge \tau_{m,p,t}$ for all $i, j, k \in S$ and $i < j\big\}$, where $\tau_{m,p,t} = (1 - \varepsilon_{m,p,t})\sqrt{\frac{4\log p}{m}}$ and $\varepsilon_{m,p,t} = (4m\log p)^{-1/2}t$. We also define

$$Q_{m,n,p} = \sum_{S\subset\{1,...,p\}:\,|S|=m} \mathbf{1}_{A_S}. \tag{4.49}$$

Similar to the discussion between (4.6) and (4.10) in the proof of Theorem 2, we have

$$\mathbb{P}\big(T_{m,n,p} \le 2\sqrt{m\log p} - t\big) \le \frac{Var(Q_{m,n,p})}{\mathbb{E}(Q_{m,n,p})^2}. \tag{4.50}$$

In the rest of the discussion under Scenario 1, we will develop a lower bound for $\mathbb{E}(Q_{m,n,p})$ and an upper bound for $Var(Q_{m,n,p})$ in two steps.

Step 1: the estimate of $\mathbb{E}(Q_{m,n,p})$. For a $m \times m$ symmetric matrix M, we use $\|M\|$ to denote its spectral norm. Set $S_0 = \{1, 2, \cdots, m\}$. Review ω_n in (4.32). Since $\{\mathbf{1}_{A_S}; S \subset \{1, ..., p\}$ with $|S| = m\}$ are identically distributed, we have

$$\mathbb{E}(Q_{m,n,p}) = \binom{p}{m}\mathbb{P}(A_{S_0}) \ge \binom{p}{m}\mathbb{P}(A_{S_0} \cap \mathcal{L}_{m,n,p}), \tag{4.51}$$

where we let $s_{m,n,p} = \max\big\{10\sqrt{m\log p},\ 2\sqrt{m\log p} + \omega_n\big\}$ and define $\mathcal{L}_{m,n,p} := \big\{\frac{\|W_{\{1,\dots,m\}} - nI_m\|}{\sqrt{n}} \le s_{m,n,p}\big\}$. It is easy to check that Assumption 1 in (3.1) implies

$$\frac{s_{m,n,p}}{\sqrt{n}} \to 0 \quad \text{and} \quad \frac{\sqrt{m}s^3_{m,n,p}}{\sqrt{n\log p}} \to 0. \tag{4.52}$$

Similar to Lemma 8, we need the following lemma, which quantifies the speed that a Wishart matrix converges to a Wigner matrix. The difference is that Lemma 8 provides a log-likelihood ratio bound for the eigenvalues of random matrices, while the following Lemma 9 gives such a bound for all the entries jointly. Both lemmas are needed in the proof as neither can be directly derived from the other.

Write $W_{\{1,\dots,m\}}$ for W_S above (2.1) with $S=\{1,2,\cdots,m\}$. Review that the Wigner matrix $\tilde{W}_{\{1,\dots,m\}}=(\tilde{w}_{ij})_{m\times m}$, where $\tilde{w}_{ij}$'s are as in (2.3).

Lemma 9 *Let $f_{m,n}(w)$ be the density function of $\frac{1}{\sqrt{n}}(W_{\{1,\dots,m\}}-nI_m)$ and $\tilde{f}_m(w)$ be the density function of $\tilde{W}_{\{1,\dots,m\}}$. If $m^3=o(n)$, then*

$$\begin{aligned}&\log f_{m,n}(w)-\log \tilde{f}_m(w)\\ =\ & o(1)+O\left(m^2n^{-1/2}\|w\|+m^2n^{-1}\|w\|^2+mn^{-1/2}\|w\|^3\right)\end{aligned}$$

for all $m\times m$ symmetric matrix w with $\|w\|\le \frac{2}{3}\sqrt{n}$.

Remark 5 We note that the approximation of the Wigner matrix by the Whishart matrices has been studied in the literature. For example, Doumerc (2002) studied the convergence theory and its connection with Brownian percolation when m is fixed and $n\to\infty$. Also, see (Chapter 3 Anderson, 1962) for the asymptotic joint distribution of the sample mean and sample covariance of i.i.d. multivariate Gaussian observations. Different from these existing results, Lemma 9 quantifies the convergence speed for growing n and m in terms of the log-likelihood ratio.

Below, we combine the above Lemma 9 and some change of measure arguments to obtain a lower bound of $\mathbb{P}(A_{S_0}\cap\mathcal{L}_{m,n,p})$. Define a non-random set $B_{m,p}=\{w_{ij}:w_{ij}\ge\tau_{m,p,t},1\le i\le j\le m\}$. By the first limit from (4.52), $s_{m,n,p}\le\frac{2}{3}\sqrt{n}$. Therefore, from Lemma 9 we have

$$\begin{aligned}&\mathbb{P}\left(A_{S_0}\cap\mathcal{L}_{m,n,p}\right)\\ =&\int_{w\in B_{m,p},\|w\|\le s_{m,n,p}}\tilde{f}_m(w)\cdot\exp\{\log f_{m,n}(w)-\log\tilde{f}_m(w)\}dw\\ =&\exp\left\{o(1)+O\big(\frac{m^2s_{m,n,p}}{\sqrt{n}}+\frac{m^2s_{m,n,p}^2}{n}+\frac{ms_{m,n,p}^3}{\sqrt{n}}\big)\right\}\\ &\qquad\qquad\cdot\mathbb{P}(\tilde{A}_{\{1,\dots,m\}}\cap\tilde{\mathcal{L}}_{m,n,p})\\ \ge&\frac{1}{2}\cdot\exp\left\{O\big(\frac{m^2s_{m,n,p}}{\sqrt{n}}+\frac{m^2s_{m,n,p}^2}{n}+\frac{ms_{m,n,p}^3}{\sqrt{n}}\big)\right\}\\ &\qquad\qquad\cdot\left[\mathbb{P}(\tilde{A}_{\{1,\dots,m\}})-\mathbb{P}(\tilde{\mathcal{L}}_{m,n,p}^c)\right],\end{aligned}$$

where $\tilde{A}_{\{1,\dots,m\}}$ is as in (4.7) with $S=\{1,\cdots,m\}$ and $\tilde{\mathcal{L}}_{m,n,p}=\{\|\tilde{W}_{\{1,\dots,m\}}\|\le s_{m,n,p}\}$. Under Assumption 1 in (3.1), evidently $\frac{m}{s_{m,n,p}^2}\to0$ and $\frac{m}{\sqrt{n}\,s_{m,n,p}}\to0$. This implies $\frac{m^2s_{m,n,p}}{\sqrt{n}}+\frac{m^2s_{m,n,p}^2}{n}+\frac{ms_{m,n,p}^3}{\sqrt{n}}=O\big(\frac{ms_{m,n,p}^3}{\sqrt{n}}\big)$. Thus, we have

$$\mathbb{P}\left(A_{S_0}\cap\mathcal{L}_{m,n,p}\right)\ge\frac{1}{2}\cdot e^{O(ms_{m,n,p}^3/\sqrt{n})}\left\{\mathbb{P}(\tilde{A}_{\{1,\dots,m\}})-\mathbb{P}(\tilde{\mathcal{L}}_{m,n,p}^c)\right\}. \quad (4.53)$$

Obviously, $\mathbb{E}(Q_{m,n,p}) = \binom{p}{m}\mathbb{P}(A_{\{1,...,m\}})$. Recalling $\tilde{A}_{\{1,...,m\}}$ and $\tilde{Q}_{m,p}$ as in (4.7) and $\tilde{Q}_{m,p} = \sum_{S\subset\{1,...,p\}:\,|S|=m} \mathbf{1}_{\tilde{A}_S}$, respectively, we see that $\mathbb{E}(\tilde{Q}_{m,p}) = \binom{p}{m}\mathbb{P}(\tilde{A}_{\{1,...,m\}})$. Thus, we further have from (4.51) and (4.53) that

$$\mathbb{E}(Q_{m,n,p}) \geq \frac{1}{2}\cdot e^{O(ms^3_{m,n,p}/\sqrt{n})}\Big\{\mathbb{E}(\tilde{Q}_{m,p}) - \binom{p}{m}\mathbb{P}(\tilde{\mathcal{L}}^c_{m,n,p})\Big\}. \tag{4.54}$$

To further obtain a lower bound of the above expression, we analyze each term on the right-hand side. Recall the definition of $\varepsilon_{m,p,t}$ below (4.7), we know $\varepsilon_{m,p,t} \in (0,1)$. By (4.16),

$$\begin{aligned}\mathbb{E}(\tilde{Q}_{m,p}) &\geq \exp\left\{\varepsilon_{m,p,t} m\log p + O(m^2\log\log p)\right\}\\ &\geq \exp\left\{\frac{\delta}{4}\sqrt{m\log p}\right\}\end{aligned} \tag{4.55}$$

where the condition $m = o((\log p)^{1/3}/\log\log p)$ from Assumption 1 in (3.1) is essentially used in the last step. Now,

$$\begin{aligned}\binom{p}{m}\mathbb{P}\left(\tilde{\mathcal{L}}^c_{m,n,p}\right) &\leq p^m\mathbb{P}\left(\|\tilde{W}_{\{1,...,m\}}\| \geq s_{m,n,p}\right)\\ &= 2p^m\mathbb{P}\left(\lambda_1(\tilde{W}_{\{1,...,m\}}) \geq s_{m,n,p}\right),\end{aligned} \tag{4.56}$$

where the fact that $\tilde{W}_{\{1,..,m\}}$ and $-\tilde{W}_{\{1,...,m\}}$ have the same distribution is used in the last step. The following lemma help us estimate the last probability.

Lemma 10 *[Lemma 4.1 from Jiang and Li (2015)] Let $\tilde{W}_{\{1,...,m\}}$ be defined by $\tilde{W}_S$ above (2.4) with $S = \{1,...,m\}$. Then there is a constant $\kappa > 0$ such that $\mathbb{P}\big(\lambda_1(\tilde{W}_{\{1,...,m\}}) \geq x$ or $\lambda_m(\tilde{W}_{\{1,...,m\}}) \leq -x\big) \leq \kappa\cdot e^{-\frac{x^2}{4}+\kappa\sqrt{m}x}$ for all $x > 0$ and all $m \geq 2$.*

By letting $x = s_{m,n,p}$ in Lemma 10, we have $\mathbb{P}\big(\lambda_1(\tilde{W}_{\{1,...,m\}}) \geq s_{m,n,p}\big) \leq \exp\big\{-\frac{s^2_{m,n,p}}{4} + \kappa\sqrt{m}s_{m,n,p}\big\}$. Combining this with (4.56), we arrive at

$$\binom{p}{m}\mathbb{P}(\tilde{\mathcal{L}}^c_{m,n,p}) \leq 2\cdot\exp\Big\{m\log p - \frac{s^2_{m,n,p}}{4} + \kappa\sqrt{m}s_{m,n,p}\Big\}.$$

Since $s_{m,n,p} \geq 10\sqrt{m\log p}$, we know $m\log p - \frac{1}{4}s^2_{m,n,p} \leq -\frac{6}{25}s^2_{m,n,p}$. Moreover, $\sqrt{m}s_{m,n,p} = o(s^2_{m,n,p})$. Consequently,

$$\binom{p}{m}\mathbb{P}(\tilde{\mathcal{L}}^c_{m,n,p}) \leq \exp\Big\{-\big(\frac{6}{25} + o(1)\big)s^2_{m,n,p}\Big\}. \tag{4.57}$$

Comparing the above inequality with (4.55), we obtain $\binom{p}{m}\mathbb{P}\big(\tilde{\mathcal{L}}^c_{m,n,p}\big) = o(1) = o(\mathbb{E}(\tilde{Q}_{m,p}))$. This result, combined with (4.54), gives

$$\mathbb{E}(Q_{m,n,p}) \geq \frac{1}{3} \cdot e^{O(ms^3_{m,n,p}/\sqrt{n})}\mathbb{E}(\tilde{Q}_{m,p}), \tag{4.58}$$

which joint with (4.16) concludes

$$\begin{aligned}\mathbb{E}(Q_{m,n,p}) \quad &\geq \quad \frac{1}{3}\exp\big\{[1-(1-\varepsilon_{m,p,t})^2]m\log p \\ &\qquad\qquad + O(m^2\log\log p + mn^{-1/2}s^3_{m,n,p})\big\}.\end{aligned} \tag{4.59}$$

This completes our analysis for $\mathbb{E}(Q_{m,n,p})$.

Step 2: the estimate of $Var(Q_{m,n,p})$. Replacing "$\tilde{A}_S$" in (4.7) with "A_S" in (4.49), and using the same argument as obtaining (4.20), we have from (4.24) that

$$\begin{aligned}Var(Q_{m,n,p}) \leq \quad & \mathbb{E}(Q_{m,n,p}) \\ & + m\max_{l=1,\dots,m-1} p^{2m-l}\mathbb{P}\Big(A_{\{1,\dots,m\}} \cap A_{\{1,\dots,l,m+1,\dots,2m-l\}}\Big).\end{aligned} \tag{4.60}$$

Now we bound the last term above. Review $\mathcal{L}_{2m,n,p}$ below (4.51). Trivially,

$$\begin{aligned}&\mathbb{P}\left(A_{\{1,\dots,m\}} \cap A_{\{1,\dots,l,m+1,\dots,2m-l\}}\right) \\ \leq&\mathbb{P}\left(A_{\{1,\dots,m\}} \cap A_{\{1,\dots,l,m+1,\dots,2m-l\}} \cap \mathcal{L}_{2m,n,p}\right) + \mathbb{P}(\mathcal{L}^c_{2m,n,p}).\end{aligned} \tag{4.61}$$

By (4.57), we know $mp^{2m}\mathbb{P}(\mathcal{L}^c_{2m,n,p}) = o(1)$. Let $f_{2m,n}(w)$ be the density function of $\frac{1}{\sqrt{n}}(W_{\{1,\dots,2m\}} - nI_{2m})$ and $\tilde{f}_{2m}(w)$ be the density function of $\tilde{W}_{\{1,\dots,2m\}}$. Recall $A_S = \big\{\frac{1}{\sqrt{n}}(W_{kk}-n) \geq \tau_{m,p,t}, \frac{W_{ij}}{\sqrt{n}} \geq \tau_{m,p,t}$ for all $i,j,k \in S$ and $i<j\big\}$. Define (non-random) set

$$B_S = \big\{(w_{ij})_{i,j\in S};\, w_{ij} \geq \tau_{m,p,t} \text{ for all } i,j \in S \text{ with } i \leq j\big\}.$$

Then,

$$\begin{aligned}&\mathbb{P}\left(A_{\{1,\dots,m\}} \cap A_{\{1,\dots,l,m+1,\dots,2m-l\}} \cap \mathcal{L}_{2m,n,p}\right) \\ = \quad & \int_B \tilde{f}_{2m}(w)\cdot\exp\{\log f_{2m,n}(w) - \log\tilde{f}_{2m}(w)\}dw\end{aligned}$$

where $B := B_{\{1,\cdots,m\}} \cap B_{\{1,\dots,l,m+1,\dots,2m-l\}} \cap \{\|w\| \leq s_{2m,n,p}\}$. By Lemma 9 and by a change-measure argument similar to the one getting (4.53), we see

$$\begin{aligned}&\mathbb{P}\left(A_{\{1,\dots,m\}} \cap A_{\{1,\dots,l,m+1,\dots,2m-l\}} \cap \mathcal{L}_{2m,n,p}\right) \\ \leq& 2\cdot e^{O(ms^3_{m,n,p}/\sqrt{n})}\mathbb{P}(\tilde{A}_{\{1,\dots,m\}} \cap \tilde{A}_{\{1,\dots,l,m+1,\dots,2m-l\}}).\end{aligned} \tag{4.62}$$

The benefit of the above step is transferring the probability on the Wishart matrix to that on the Wigner matrix up to a certain error. Combining (4.61)-(4.62), we have $mp^{2m-l}\mathbb{P}\big(A_{\{1,\dots,m\}}\cap A_{\{1,\dots,l,m+1,\dots,2m-l\}}\big)\le 2\cdot e^{O(ms^3_{m,n,p}/\sqrt{n})}\cdot mp^{2m-l}\mathbb{P}(\tilde{A}_{\{1,\dots,m\}}\cap\tilde{A}_{\{1,\dots,l,m+1,\dots,2m-l\}})+o(1)$. Combining this with (4.60), we have $Var(Q_{m,n,p})\le\mathbb{E}(Q_{m,n,p})+2\cdot e^{O(s^3_{m,n,p}/\sqrt{n})}\cdot G_m+o(1)$, where

$$G_m:=\max_{l=1,\dots,m-1}\left\{mp^{2m-l}\mathbb{P}(\tilde{A}_{\{1,\dots,m\}}\cap\tilde{A}_{\{1,\dots,l,m+1,\dots,2m-l\}})\right\}.$$

Thus,

$$\begin{aligned}\frac{Var(Q_{m,n,p})}{\mathbb{E}(Q_{m,n,p})^2}\le&e^{O(ms^3_{m,n,p}/\sqrt{n})}\big[\mathbb{E}(Q_{m,n,p})\big]^{-2}\cdot G_m\\&+\big[\mathbb{E}(Q_{m,n,p})\big]^{-1}+o\Big(\big[\mathbb{E}(Q_{m,n,p})\big]^{-2}\Big).\end{aligned}\tag{4.63}$$

According to (4.23) and (4.58), the first term on the right-hand side of the above inequality is no more than

$$9\cdot\exp\left\{-\Big[1-\frac{1}{m}(1-\varepsilon_{m,p,t})^2\Big]\log p+O\Big(m^2\log\log p+\frac{ms^3_{m,n,p}}{\sqrt{n}}\Big)\right\}.$$

Notice that $1-\varepsilon_{m,p,t}\le 1$ and $m\ge 2$ and $O(m^2\log\log p+mn^{-1/2}s^3_{m,n,p})=o(\log p)$. Thus, the above display further implies

$$e^{O(s^3_{m,n,p}/\sqrt{n})}\big(\mathbb{E}(Q_{m,n,p})\big)^{-2}\cdot G_m\le\exp\left\{-\Big(\frac{1}{2}+o(1)\Big)\log p\right\}.$$

We next study the last two terms from (4.63).

By the condition $m=o((\log p)^{1/3}/\log\log p)$ from Assumption 1 in (3.1) and the second limit in (4.52),

$$O\left(\frac{ms^3_{m,n,p}}{\sqrt{n}}+m^2\log\log p\right)=o\big(\sqrt{m\log p}\,\big).\tag{4.64}$$

Recall $\varepsilon_{m,p,t}:=(4m\log p)^{-1/2}t$. It is readily seen that $[1-(1-\varepsilon_{m,p,t})^2]m\log p\ge\varepsilon_{m,p,t}m\log p\ge\frac{t}{2}\sqrt{m\log p}$. Consequently, it is known from (4.59) that $\mathbb{E}(Q_{m,n,p})\ge\frac{1}{3}\cdot\exp\big\{\frac{t}{2}\sqrt{m\log p}\big\}$ uniformly over $\delta\le t\le 2\sqrt{m\log p}-m\xi_p$. Therefore, we conclude from (4.59) and (4.64) that

$$\big(\mathbb{E}(Q_{m,n,p})\big)^{-1}+o\Big(\big(\mathbb{E}(Q_{m,n,p})\big)^{-2}\Big)\le 3\cdot\exp\Big\{-\Big(\frac{1}{2}+o(1)\Big)t\sqrt{m\log p}\Big\}.\tag{4.65}$$

Combining (4.63)-(4.65), we see

$$\begin{aligned}&\frac{Var(Q_{m,n,p})}{\mathbb{E}(Q_{m,n,p})^2}\\ \le&\exp\Big\{-\Big(\frac{1}{2}+o(1)\Big)\log p\Big\}+3\cdot\exp\Big\{-\Big(\frac{1}{2}+o(1)\Big)t\sqrt{m\log p}\Big\}.\end{aligned}$$

By (4.50) and the above inequality,

$$\begin{aligned}&\mathbb{P}\left(T_{m,n,p}\le 2\sqrt{m\log p}-t\right)\\ \le&\exp\Big\{-\Big(\frac{1}{2}+o(1)\Big)\log p\Big\}+3\cdot\exp\Big\{-\Big(\frac{1}{2}+o(1)\Big)t\sqrt{m\log p}\Big\}.\end{aligned}$$

Finally, from the inequality $t^2\le 2e^t$ we have that

$$\limsup_{n\to\infty}\sup_{\delta\le t\le 2\sqrt{m\log p}-m\xi_p} e^{\alpha t}t^2\mathbb{P}\left(T_{m,n,p}\le 2\sqrt{m\log p}-t\right)=0 \tag{4.66}$$

for any $\alpha>0$ and $\delta>0$.

Scenario 2: $t>2\sqrt{m\log p}-m\xi_p$**.** Review (2.1). By the fact that $\lambda_1(M)\ge\max_{1\le i\le m}M_{ii}$ for any non-negative definite matrix $M=(M_{ij})_{m\times m}$, we have $T_{m,n,p}\ge\max_{S\subset\{1,\dots,p\},|S|=m}\lambda_1(W_S)\ge\max_{1\le i\le p}W_{ii}$, where $W_{ii}=\sum_{j=1}^n x_{ji}^2$ and $\{W_{ii};\,1\le i\le m\}$ are i.i.d. random variables. Thus, by independence,

$$\mathbb{P}\Big(\frac{T_{m,n,p}-n}{\sqrt{n}}\le 2\sqrt{m\log p}-t\Big)\le\mathbb{P}\Big(\frac{W_{11}-n}{\sqrt{n}}\le 2\sqrt{m\log p}-t\Big)^p. \tag{4.67}$$

Note that $W_{11}=\sum_{j=1}^n x_{j1}^2$ is a sum of i.i.d. random variables with $Var(x_{11})=2$ and $\mathbb{E}(x_{11}^6)<\infty$. We discuss two situations: $2\sqrt{m\log p}-m\xi_p\le t\le 4\sqrt{m\log p}$ and $t\ge 4\sqrt{m\log p}$.

Assuming $2\sqrt{m\log p}-m\xi_p\le t\le 4\sqrt{m\log p}$ for now. Recalling $\Phi(x)=(2\pi)^{-1/2}\int_{-\infty}^x e^{-t^2/2}\,dt$, we get from the Berry-Essen Theorem that

$$\begin{aligned}\mathbb{P}\Big(\frac{W_{11}-n}{\sqrt{n}}\le 2\sqrt{m\log p}-t\Big)\le&\Phi\Big(\sqrt{2m\log p}-\frac{t}{\sqrt{2}}\Big)+\frac{\kappa}{\sqrt{n}}\\ \le&2\cdot\max\Big\{\Phi\big(\sqrt{2m\log p}-\frac{t}{\sqrt{2}}\big),\frac{\kappa}{\sqrt{n}}\Big\}\end{aligned}$$

for some constant $\kappa>0$. Combine the above inequalities with (4.28) to see

$$\mathbb{P}\Big(\frac{T_{m,n,p}-n}{\sqrt{n}}\le 2\sqrt{m\log p}-t\Big)\le 2\cdot\max\Big\{e^{-mp^{0.9}},e^{-\frac{p\log n}{2}(1+o(1))}\Big\}.$$

By (3.1), $\sqrt{m\log p} \le \log p$. It is easy to check
(4.68)
$$\lim_{n\to\infty} \sup_{2\sqrt{m\log p}-m\xi_p \le t\le 4\sqrt{m\log p}} e^{\alpha t}t^2\mathbb{P}\left(\frac{T_{m,n,p}-n}{\sqrt{n}} \le 2\sqrt{m\log p}-t\right)=0.$$

We proceed to the second situation: $t\ge 4\sqrt{m\log p}$. In this case, $2\sqrt{m\log p}-t\le -2\sqrt{m\log p}$. By Lemma 1 from Laurent and Massart (2000), $\mathbb{P}(W_{11}-n\le -2\sqrt{nx})\le e^{-x}$ for any $x>0$. Thus, $\mathbb{P}\big(\frac{W_{11}-n}{\sqrt{n}}\le 2\sqrt{m\log p}-t\big)\le \exp\big\{-\big(\frac{t}{2}-\sqrt{m\log p}\big)^2\big\}\le \exp\big\{-\frac{t^2}{16}\big\}$. This inequality and (4.67) yield $\mathbb{P}\big(\frac{T_{m,n,p}-n}{\sqrt{n}}\le 2\sqrt{m\log p}-t\big)\le \exp\big\{-\frac{pt^2}{16}\big\}$. Consequently,

$$\begin{aligned}&\sup_{t\ge 4\sqrt{m\log p}} e^{\alpha t}t^2\mathbb{P}\left(\frac{T_{m,n,p}-n}{\sqrt{n}}\le 2\sqrt{m\log p}-t\right)\\ &\le \sup_{t\ge 4\sqrt{m\log p}}\exp\left\{-\frac{pt^2}{16}+\alpha t+2\log t\right\}\\ &\le \exp\left\{-mp(\log p)(1+o(1))\right\}.\end{aligned}$$

Hence,

$$\lim_{n\to\infty}\sup_{t\ge 4\sqrt{m\log p}} e^{\alpha t}t^2\mathbb{P}\left(\frac{T_{m,n,p}-n}{\sqrt{n}}\le 2\sqrt{m\log p}-t\right)=0. \tag{4.69}$$

By collecting (4.66), (4.68) and (4.69) together, we arrive at

$$\lim_{n\to\infty}\sup_{t\ge\delta} e^{\alpha t}t^2\mathbb{P}\left(\frac{T_{m,n,p}-n}{\sqrt{n}}\le 2\sqrt{m\log p}-t\right)=0.$$

The proof is completed. ■

4.2.3. *Proofs of Theorem 3 and Remark 3.* The following lemma serves the proof of Theorem 3. Its own proof is placed in Appendix.

Lemma 11 *Let $\tilde{W}=\tilde{W}_{m\times m}$ be as defined in (3.7) with $0\le\eta\le 2$. Then*

$$\mathbb{P}(\lambda_1(\tilde{W})\ge x)\le \frac{m^{1.5}\log m}{\delta^m}\cdot\exp\Big\{-\frac{(x-2r\delta)^2}{2m^{-1}(\eta-2)+4}\Big\}+2\cdot e^{-r^2/8}$$

for all $r\ge 4m$, $\delta\in(0,1)$ and $x>2r\delta+1$.

Proof of Theorem 3. For any $0<\varepsilon<1$, we first show that

$$\mathbb{P}\Big(\lambda_1(\tilde{W}_{\{1,\dots,m\}})\ge(1+\varepsilon)\sqrt{\{4m+2(\eta-2)\}\log p}\Big)=o(p^{-m}) \tag{4.70}$$

by using Lemma 11. To do so, set $x = (1+\varepsilon)\sqrt{[4m+2(\eta-2)]\log p}$, $r = \sqrt{128 m \log p}$ and $\delta = (8r)^{-1}\varepsilon\sqrt{[4m+2(\eta-2)]\log p}$. Rewrite δ such that $\delta = \big(\frac{1}{64}\sqrt{\frac{2m+\eta-2}{m}}\big)\varepsilon$. It is easy to check that the coefficient of ε is always sitting in $[1/64, 2/64]$ for any $m \geq 2$ and $\eta \in [0,2]$. This, the fact that $\sup_{k\geq 2}(k^{1.5}\log k)\cdot\delta^k < \infty$ and the definition of r lead to

$$\frac{m^{1.5}\log m}{\delta^m} = O(\varepsilon^{-2m}) \text{ and } e^{-r^2/8} = o(p^{-6m}). \tag{4.71}$$

We can see that $x - r\delta = \left(1+\frac{7}{8}\varepsilon\right)\cdot\sqrt{[4m+2(\eta-2)]\log p}$. It follows that $-\frac{(x-2r\delta)^2}{2m^{-1}(\eta-2)+4} \leq -\frac{(1+\frac{\varepsilon}{2})^2[4m+2(\eta-2)]\log p}{2m^{-1}(\eta-2)+4}$, and thus $\exp\big\{-\frac{(x-2r\delta)^2}{2m^{-1}(\eta-2)+4}\big\} \leq p^{-[1+(\varepsilon/2)]^2 m}$. This and (4.71) implies (4.70). Consequently,

$$\begin{aligned}&\mathbb{P}\left(\tilde{T}_{m,p} \geq (1+\varepsilon)\sqrt{[4m+2(\eta-2)]\log p}\right)\\ \leq& p^m\mathbb{P}\left(\lambda_1(\tilde{W}_{\{1,\dots,m\}}) \geq (1+\varepsilon)\sqrt{[4m+2(\eta-2)]\log p}\right) \to 0.\end{aligned}$$

To complete the proof, it is enough to check that

$$\mathbb{P}\left(\tilde{T}_{m,p} < (1-\varepsilon)\sqrt{[4m+2(\eta-2)]\log p}\right) \to 0 \tag{4.72}$$

for each $\varepsilon \in (0,1)$. For notational simplicity, let $K_m = 4m + 2(\eta-2)$ and $\tau_{m,p} = \log p / K_m$. Similar to the proof of Theorem 2, define $\tilde{A}_S = \big\{\tilde{W}_{ii} \geq 2(1-\varepsilon)\eta\sqrt{\tau_{m,p}},\ \tilde{W}_{ij} \geq 4(1-\varepsilon)\sqrt{\tau_{m,p}}$ for all $i,j \in S$ and $i \leq j\big\}$ for each $S \subset \{1,\dots,p\}$ with $|S| = m$. We next compute $\mathbb{P}(\tilde{A}_{S_0})$ and $\mathbb{P}(\tilde{A}_{S_0}\cap\tilde{A}_{S_1})$, respectively, where $S_0 = \{1,\dots,m\}$ and $S_1 = \{1,\dots,l,m+1,\dots,2m-l\}$. By independence,

$$\begin{aligned}\mathbb{P}\left(\tilde{A}_{S_0}\right) &= \prod_{i=1}^m \mathbb{P}\left(\tilde{W}_{ii} \geq 2(1-\varepsilon)\eta\sqrt{\tau_{m,p}}\right)\cdot\\ &\qquad \prod_{1\leq i<j\leq m}\mathbb{P}\left(\tilde{W}_{ij} \geq 4(1-\varepsilon)\sqrt{\tau_{m,p}}\right).\end{aligned}$$

Since $\tilde{W}_{ii} \sim N(0,\eta)$ and $\tilde{W}_{ij} \sim N(0,1)$ for all $i \neq j$, we further have

$$\mathbb{P}\left(\tilde{A}_{S_0}\right) = \bar{\Phi}\left(2(1-\varepsilon)\sqrt{\eta\tau_{m,p}}\right)^m\bar{\Phi}\left(4(1-\varepsilon)\sqrt{\tau_{m,p}}\right)^{\frac{m(m-1)}{2}},$$

where $\bar{\Phi}(x) = (2\pi)^{-1/2}\int_x^\infty e^{-t^2/2}\,dt$ for $x \in \mathbb{R}$. Similar to (4.21),

$$\begin{aligned}&\mathbb{P}\left(\tilde{A}_{S_0}\cap\tilde{A}_{S_1}\right)\\ =& \bar{\Phi}\left(2(1-\varepsilon)\sqrt{\eta\tau_{m,p}}\right)^{2m-l}\bar{\Phi}\left(4(1-\varepsilon)\sqrt{\tau_{m,p}}\right)^{\frac{m(m-1)}{2}\cdot 2-\frac{l(l-1)}{2}}.\end{aligned} \tag{4.73}$$

From (4.13), $\log\bar{\Phi}(x) = -\frac{x^2}{2} - \log(x) - \log\sqrt{2\pi} + o(1)$ as $x \to \infty$. Then,

$$\log \mathbb{P}(\tilde{A}_{S_0}) = -2m(1-\varepsilon)^2\eta\tau_{m,p} - 4m(m-1)(1-\varepsilon)^2\tau_{m,p} + R_{m,p},$$

where

$$\begin{aligned}R_{m,p} :=& - m\log\left[2(1-\varepsilon)\sqrt{\eta\tau_{m,p}}\right] - m\log\sqrt{2\pi}\\ &- \frac{m(m-1)}{2}\cdot\log\left[4(1-\varepsilon)\sqrt{\tau_{m,p}}\right] - \frac{m(m-1)}{2}\cdot\log\sqrt{2\pi} + o(m^2).\end{aligned}$$

Notice $-2m(1-\varepsilon)^2\eta\tau_{m,p} - 4m(m-1)\cdot(1-\varepsilon)^2\tau_{m,p} = -(1-\varepsilon)^2 m\log p$. Similar to (4.14), we obtain that $R_{m,p} = O(m^2\log\log p)$. Thus, $\log\mathbb{P}\big(\tilde{A}_{S_0}\big) = -(1-\varepsilon)^2 m\log p + O(m^2\log\log p)$. By the same argument as obtaining (4.16), we see

$$\log\left[\binom{p}{m}\mathbb{P}\big(\tilde{A}_{S_0}\big)\right] = [1-(1-\varepsilon)^2]m\log p + O(m^2\log\log p). \tag{4.74}$$

In particular the above goes to infinity as $n\to\infty$. By (4.13) and (4.73),

$$\begin{aligned}&\log\mathbb{P}\left(\tilde{A}_{S_0}\cap\tilde{A}_{S_1}\right)\\ =& -2(2m-l)(1-\varepsilon)^2\eta\tau_{m,p} - 8\left[m(m-1) - \frac{1}{2}l(l-1)\right](1-\varepsilon)^2\tau_{m,p}\\ &+ O(m^2\log\log p).\end{aligned}$$

The right hand side above without the term "$O(m^2\log\log p)$" is identical to $-2(1-\varepsilon)^2 m\log p + (1-\varepsilon)^2(\log p)\cdot K_m^{-1}\cdot(4l-4+2\eta)l$. Thus,

$$\begin{aligned}&\log\left[mp^{2m-l}\mathbb{P}\left(\tilde{A}_{S_0}\cap\tilde{A}_{S_1}\right)\right]\\ =&[2m-l-2(1-\varepsilon)^2m]\log p + (1-\varepsilon)^2(\log p)\cdot K_m^{-1}\cdot l(4l-4+2\eta)\\ &+ O(m^2\log\log p)\\ =&(\log p)\left\{2\left[1-(1-\varepsilon)^2\right]m + \left[-1+(1-\varepsilon)^2\cdot K_m^{-1}\cdot(4l-4+2\eta)\right]l\right\}\\ &+ O(m^2\log\log p).\end{aligned} \tag{4.75}$$

Let us take a closer look at the above display. For $1\le l\le m-1$, $K_m^{-1}(4l-4+2\eta) = \frac{4(l-1)+2\eta}{4(m-1)+2\eta} < 1$. Thus, for $1\le l\le m-1$ and $0<\varepsilon<1$, $\left[-1+(1-\varepsilon)^2K_m^{-1}(4l-4+2\eta)\right]l \le [-1+(1-\varepsilon)^2]\cdot l\le -\varepsilon$. Combining this with (4.75), we obtain that

$$\log\left[mp^{2m-l}\mathbb{P}\left(\tilde{A}_{S_0}\cap\tilde{A}_{S_1}\right)\right] \le 2\left[1-(1-\varepsilon)^2\right]m\log p - \varepsilon\log p + O(m^2\log\log p) \tag{4.76}$$

uniformly for $1 \leq l \leq m-1$. Define $\tilde{Q}_p = \sum_{S\subset\{1,...,p\},|S|=m} \mathbf{1}_{\tilde{A}_S}$. From (4.74), $\mathbb{E}\tilde{Q}_p = \binom{p}{m}\mathbb{P}\left(\tilde{A}_{S_0}\right) \to \infty$ as $n \to \infty$. Moreover, we see from (4.74) and (4.76) that $\big(\mathbb{E}\tilde{Q}_p\big)^{-2} \max_{1\leq l\leq m-1} mp^{2m-l}\mathbb{P}\big(\tilde{A}_{S_0}\cap\tilde{A}_{S_1}\big) \leq \exp\{-\varepsilon \log p + O(m^2 \log\log p)\} \to 0$. By (4.24) and a similar argument to (4.20), we get $Var(\tilde{Q}_p) \leq \mathbb{E}\tilde{Q}_p + \max_{1\leq l\leq m-1} mp^{2m-l}\mathbb{P}\big(\tilde{A}_{S_0}\cap\tilde{A}_{S_1}\big)$. This and the analysis imply $\frac{Var(\tilde{Q}_p)}{(\mathbb{E}\tilde{Q}_p)^2} \to 0$. As a result, $\lim_{p\to\infty}\mathbb{P}(\tilde{Q}_p = 0) = 0$ by (4.9). According to (4.6), if there exists $S_0 \subset \{1, ..., p\}$ such that $|S_0| = m$ and $\tilde{A}_{S_0}$ occurs, then $\tilde{T}_{m,p} \geq \lambda_1(\tilde{W}_{S_0}) \geq \frac{1}{m}\big(2m(1-\varepsilon)\eta\sqrt{\tau_{m,p}} + 4m(m-1)(1-\varepsilon)\sqrt{\tau_{m,p}}\big) = (1-\varepsilon)\sqrt{[4m+2(\eta-2)]\log p}$. Therefore,

$$\mathbb{P}\left(\tilde{T}_{m,p} < (1-\varepsilon)\sqrt{\{4m+2(\eta-2)\}\log p}\right) \leq \mathbb{P}\left(\tilde{Q}_p = 0\right) \to 0.$$

This implies (4.72). The proof is finished. ■

Proof of Remark 3. These results are direct consequences of the following lemma, whose proof is given in Appendix B. ■

Lemma 12 *Let $\{Z_p\}_{p\geq 1}$ be a sequence of non-negative random variables. Consider the following statements.*

(i) $\lim_{p\to\infty} \mathbb{E}\left[e^{\alpha Z_p}\mathbf{1}_{\{Z_p\geq\delta\}}\right] = 0$ *for all* $\alpha > 0$ *and* $\delta > 0$.

(ii) $\lim_{p\to\infty} \mathbb{E}(e^{\alpha Z_p}) = 1$ *for all* $\alpha > 0$.

(iii) $\lim_{p\to\infty} \mathbb{E}(Z_p^\alpha) = 0$ *for all* $\alpha > 0$.

(iv) $\lim_{p\to\infty} \mathbb{P}(Z_p \geq \delta) = 0$ *for all* $\delta > 0$.

(v) $\lim_{p\to\infty} \mathrm{Var}(Z_p) = 0$ *for all* $\alpha > 0$.

Then, (i)$\iff$ (ii) $\implies$ (iii) $\implies$ (iv) and (v). Here, "A $\iff$ B" means two statements A and B are equivalent, and A $\implies$ B means statement A implies statement B.

Acknowledgment. The research of Tony Cai was supported in part by NSF grants DMS-1712735 and DMS-2015259 and NIH grants R01-GM129781 and R01-GM123056. Tiefeng Jiang is partially supported by NSF Grant DMS-1406279. Xiaoou Li is partially supported by NSF Grant DMS-1712657.

References.

Anderson, T. W. (1962). *An introduction to multivariate statistical analysis.* Wiley New York.

Arratia, R., Goldstein, L., and Gordon, L. (1990). Poisson approximation and the chen-stein method. *Statistical Science*, pages 403–424.

Bai, Z. D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica*, 9(3):611–662.

Bai, Z. D. and Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices*, volume 20. Springer.

Bai, Z. D. and Yin, Y. Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *The Annals of Probability*, 21(3):1275–1294.

Baraniuk, R., Davenport, M., DeVore, R., and Wakin, M. (2008). A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263.

Bryc, W., Dembo, A., and Jiang, T. (2006). Spectral measure of large random hankel, markov and toeplitz matrices. *The Annals of Probability*, pages 1–38.

Bump, D. and Diaconis, P. (2002). Toeplitz minors. *Journal of Combinatorial Theory, Series A*, 97(2):252–271.

Cai, T. T., Fan, J., and Jiang, T. (2013). Distributions of angles in random packing on spheres. *The Journal of Machine Learning Research*, 14(1):1837–1864.

Cai, T. T. and Jiang, T. (2012). Phase transition in limiting distributions of coherence of high-dimensional random matrices. *Journal of Multivariate Analysis*, 107:24–39.

Cai, T. T., Wang, L., and Xu, G. (2009). Shifting inequality and recovery of sparse signals. *IEEE Transactions on Signal processing*, 58(3):1300–1308.

Cai, T. T., Wang, L., and Xu, G. (2010). New bounds for restricted isometry constants. *IEEE Transactions on Information Theory*, 56(9):4388–4394.

Cai, T. T. and Zhang, A. (2013). Sharp RIP bound for sparse signal and low-rank matrix recovery. *Applied and Computational Harmonic Analysis*, 35(1):74–93.

Cai, T. T. and Zhang, A. (2014). Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. *IEEE Transactions on Information Theory*, 60(1):122–132.

Candes, E. and Tao, T. (2005). Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215.

Candes, E. and Tao, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6):2313–2351.

Candes, E. J. (2008). The restricted isometry property and its implications for compressed sensing. *Comptes rendus mathematique*, 346(9-10):589–592.

Dallaporta, S. and De Castro, Y. (2019). Sparse recovery from extreme eigenvalues deviation inequalities. *ESAIM: Probability and Statistics*, 23:430–463.

Diaconis, P. and Evans, S. (2001). Linear functionals of eigenvalues of random matrices. *Transactions of the American Mathematical Society*, 353(7):2615–2633.

Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306.

Donoho, D. L., Elad, M., and Temlyakov, V. N. (2006). Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6–18.

Doumerc, Y. (2002). An asymptotic link between LUE and GUE and its spectral interpretation. *arXiv preprint math/0204175*.

Dyson, F. J. (1962a). Statistical theory of the energy levels of complex systems. i. *Journal of Mathematical Physics*, 3(1):140–156.

Dyson, F. J. (1962b). Statistical theory of the energy levels of complex systems. ii. *Journal of Mathematical Physics*, 3(1):157–165.

Dyson, F. J. (1962c). Statistical theory of the energy levels of complex systems. iii. *Journal of Mathematical Physics*, 3(1):166–175.

Edelman, A. (1988). Eigenvalues and condition numbers of random matrices. *SIAM*

Journal on Matrix Analysis and Applications, 9(4):543–560.

Fan, J., Shao, Q.-M., and Zhou, W.-X. (2018). Are discoveries spurious? distributions of maximum spurious correlations and their applications. *Annals of Statistics*, 46(3):989.

Fey, A., van der Hofstad, R., and Klok, M. J. (2008). Large deviations for eigenvalues of sample covariance matrices, with applications to mobile communication systems. *Advances in Applied Probability*, 40(4):1048–1071.

Horn, R. A. and Johnson, C. R. (2012). *Matrix Analysis*. Cambridge university press.

James, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *The Annals of Mathematical Statistics*, 35(2):475–501.

Jiang, T. (2004a). The asymptotic distributions of the largest entries of sample correlation matrices. *Annals of Applied Probability*, pages 865–880.

Jiang, T. (2004b). The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā: The Indian Journal of Statistics*, pages 35–48.

Jiang, T. and Li, D. (2015). Approximation of Rectangular Beta-Laguerre Ensembles and Large Deviations. *Journal of Theoretical Probability*, 28(3):804–847.

Johansson, K. (1990). On Szego's asymptotic formula for Toeplitz determinants and generalizations.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, pages 295–327.

Johnstone, I. M. (2008). Multivariate analysis and jacobi ensembles: Largest eigenvalue, tracy–widom limits and rates of convergence. *The Annals of Statistics*, 36(6):2638–2716.

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, pages 1302–1338.

Li, D., Liu, W.-D., and Rosalsky, A. (2010). Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probability Theory and Related Fields*, 148(1-2):5–35.

Li, D., Qi, Y., and Rosalsky, A. (2012). On jiang's asymptotic distribution of the largest entry of a sample correlation matrix. *Journal of Multivariate Analysis*, 111:256–270.

Li, D. and Rosalsky, A. (2006). Some strong limit theorems for the largest entries of sample correlation matrices. *The Annals of Applied Probability*, 16(1):423–447.

Mehta, M. L. (2004). *Random matrices*, volume 142. Elsevier, 3 edition.

Mo, Q. and Li, S. (2011). New bounds on the restricted isometry constant δ_{2k}. *Applied and Computational Harmonic Analysis*, 31(3):460–468.

Muirhead, R. J. (2009). *Aspects of Multivariate Statistical Theory*, volume 197. John Wiley & Sons.

Péché, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probability Theory and Related Fields*, 143(3-4):481–516.

Rogers, C. (1963). Covering a sphere with spheres. *Mathematika*, 10(2):157–164.

Shao, Q.-M. and Zhou, W.-X. (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *The Annals of Probability*, 42(2):623–648.

Szeg, G. (1939). *Orthogonal polynomials*, volume 23. American Mathematical Soc.

Tao, T. and Vu, V. (2010). Random matrices: Universality of esds and the circular law. *The Annals of Probability*, pages 2023–2065.

Tracy, C. A. and Widom, H. (1994). Level spacing distributions and the bessel kernel. *Communications in Mathematical Physics*, 161(2):289–309.

Tracy, C. A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177(3):727–754.

Tracy, C. A. and Widom, H. (2000). The distribution of the largest eigenvalue in the gaussian ensembles: β= 1, 2, 4. In *Calogero–Moser–Sutherland Models*, pages 461–472.

Springer.

Tracy, C. A. and Widom, H. (2002). On the limit of some Toeplitz-like determinants. *SIAM journal on matrix analysis and applications*, 23(4):1194–1196.

Wigner, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, pages 548–564.

Wigner, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, pages 325–327.

Zhang, R. and Li, S. (2018). A proof of conjecture on restricted isometry property constants δ_{tk} $(0 < t < 4/3)$. *IEEE Transactions on Information Theory*, 64(3):1699–1705.

Zhou, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Transactions of the American Mathematical Society*, 359(11):5345–5363.

Appendix

There are two sections in this part. In Appendix A we derive some results on Gamma functions, which will be used later on. The material in this part is independent of previous sections. In Appendix B we will prove the lemmas appeared in earlier sections.

APPENDIX A: AUXILIARY RESULTS ON GAMMA FUNCTIONS

Recall the Gamma function $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$ for $x > 0$.

Lemma A.1 *Let*

$$c_{m,n} = m!2^{-nm/2}\prod_{j=1}^{m}\frac{\Gamma(3/2)}{\Gamma(1+j/2)\Gamma((n-m+j)/2)};$$
$$C(m,n) = n^{m/2}e^{-nm/2}n^{m(n-m+1)/2-m}n^{m(m-1)/4}c_{m,n};$$
$$c_m = m!2^{-m}2^{-m(m-1)/4}\pi^{-m/2}\prod_{j=1}^{m}\frac{\Gamma(3/2)}{\Gamma(1+j/2)}.$$

If $m^3 = o(n)$, *then* $\log C(n,m) - \log c_m = o(1)$ *as* $n \to \infty$.

Proof of Lemma A.1. Easily,

$$\begin{aligned}\frac{C(m,n)}{c_m} &= \frac{n^{m/2}e^{-nm/2}n^{m(n-m+1)/2-m}n^{m(m-1)/4}\cdot 2^{-nm/2}}{2^{-m}2^{-m(m-1)/4}\pi^{-m/2}\cdot\prod_{j=1}^{m}\Gamma((n-m+j)/2)}\\ &= n^{m(2n-m-1)/4}e^{-mn/2}2^{(m-2n+3)m/4}\pi^{m/2}e^{-J_n}\end{aligned}$$

where

$$J_n := \log\prod_{j=0}^{m-1}\Gamma((n-j)/2) = m\log\Gamma\Big(\frac{n}{2}\Big) + \log\prod_{j=0}^{m-1}\frac{\Gamma((n-j)/2)}{\Gamma(n/2)}. \tag{A.1}$$

Then

$$\begin{aligned}&\log\frac{C(m,n)}{c_m}\\ &= \frac{m}{4}\big(2n-m-1\big)\log n - \frac{1}{2}mn + \frac{m}{4}(m-2n+3)\log 2 \\ &\quad + \frac{m}{2}\log\pi - J_n \qquad \text{(A.2)}\\ &= \frac{m}{4}\big(2n-m-1\big)\log\frac{n}{2} - \frac{1}{2}mn + \frac{m}{2}\log(2\pi) - J_n. \qquad \text{(A.3)}\end{aligned}$$

Write $\log \frac{\Gamma(x+b)}{\Gamma(x)} = (x+b)\log(x+b) - x\log x - b + \delta(x,b)$. By Lemma 5.1 from Jiang and Qi (2015), there exists a constant $C > 0$ free of x and b such that $|\delta(x,b)| \leq C \cdot \frac{b^2+|b|x+1}{x^2}$ for all $x \geq 10$ and $|b| \leq x/2$. It is easy to see that $\sum_{j=0}^{m-1} \left|\delta\Big(\frac{n}{2}, -\frac{j}{2}\Big)\right| \leq C' \cdot \frac{m^3+nm^2+m}{n^2}$ where C' is a constant free of m and n. This implies that

$$\log \prod_{j=0}^{m-1} \frac{\Gamma((n-j)/2)}{\Gamma(n/2)} = O\Big(\frac{m^2}{n}\Big) + \sum_{j=0}^{m-1} \Big(\frac{n-j}{2}\log\frac{n-j}{2} - \frac{n}{2}\log\frac{n}{2} + \frac{j}{2}\Big)$$

as $n \to \infty$. Write

$$\frac{n-j}{2}\log\frac{n-j}{2} - \frac{n}{2}\log\frac{n}{2} = \frac{n}{2}\log\Big(1-\frac{j}{n}\Big) - \frac{j}{2}\log\frac{n}{2} - \frac{j}{2}\log\Big(1-\frac{j}{n}\Big).$$

Easily, $\log(1-\frac{j}{n}) = -\frac{j}{n} + O(\frac{m^2}{n^2})$ as $n \to \infty$ uniformly for all $1 \leq j \leq m$. Hence,

$$\begin{aligned} & \sum_{j=0}^{m-1} \Big(\frac{n-j}{2}\log\frac{n-j}{2} - \frac{n}{2}\log\frac{n}{2} + \frac{j}{2}\Big) \\ = \;& O\Big(\frac{m^2}{n}\Big) - \frac{1}{4}m(m-1)\log\frac{n}{2} + \frac{(m-1)m(2m-1)}{12n} \\ = \;& -\frac{1}{4}m(m-1)\log\frac{n}{2} + O\Big(\frac{m^3}{n}\Big) \end{aligned}$$

as $n \to \infty$. In summary,

$$\log \prod_{j=0}^{m-1} \frac{\Gamma((n-j)/2)}{\Gamma(n/2)} = -\frac{1}{4}m(m-1)\log\frac{n}{2} + O\Big(\frac{m^3}{n}\Big)$$

as $n \to \infty$. On the other hand, by the Stirling formula,

$$m\log\Gamma\Big(\frac{n}{2}\Big) = \frac{(n-1)m}{2}\log\frac{n}{2} - \frac{1}{2}mn + \frac{m}{2}\log(2\pi) + O\Big(\frac{m}{n}\Big)$$

as $n \to \infty$. From (A.1) and the above two assertions we see

$$J_n = \frac{m}{4}\big(2n-m-1\big)\log\frac{n}{2} - \frac{1}{2}mn + \frac{m}{2}\log(2\pi) + o(1) \tag{A.4}$$

as $n \to \infty$, which together with (A.2) proves the lemma. ■

Lemma A.2 *Let*

$$\Gamma_m\Big(\frac{n}{2}\Big) = \pi^{m(m-1)/4}\prod_{j=1}^{m}\Gamma\Big(\frac{n-j+1}{2}\Big);$$
$$A(m,n) = n^{m(m+1)/4+m(n-m-1)/2}e^{-nm/2}2^{-\frac{nm}{2}}/\Gamma_m(n/2);$$
$$B(m) = (2\pi)^{-m(m+1)/4}2^{-m/2}.$$

If $m^3 = o(n)$, *then* $\log A(m,n) - \log B(m) \to 0$ *as* $n \to \infty$.

Proof of Lemma A.2. Observe

$$\begin{aligned}
&\log\frac{A(m,n)}{B(m)}\\
= \ & \Big[\frac{1}{4}m(m+1)+\frac{1}{2}m(n-m-1)\Big]\log n - \frac{1}{2}mn\\
&-\frac{1}{2}m(n-1)\log 2 + \frac{1}{4}m(m+1)\log(2\pi) - \log\Gamma_m\Big(\frac{n}{2}\Big).
\end{aligned} \tag{A.5}$$

By definition, $\Gamma_m\big(\frac{n}{2}\big) = \pi^{m(m-1)/4}\prod_{j=0}^{m-1}\Gamma\big(\frac{n-j}{2}\big)$. From (A.1) and (A.4), we see that

$$\begin{aligned}
&\log\Gamma_m\Big(\frac{n}{2}\Big)\\
= \ & \frac{1}{4}m(m-1)\log\pi + \frac{m}{4}\big(2n-m-1\big)\log\frac{n}{2} - \frac{1}{2}mn + \frac{m}{2}\log(2\pi) + o(1)\\
= \ & \frac{m}{4}\big(2n-m-1\big)\log n - \frac{1}{2}mn - \frac{1}{2}m(n-1)\log 2 + \frac{1}{4}m(m+1)\log(2\pi) + o(1).
\end{aligned}$$

By comparing this identity with (A.5), we conclude $\log\frac{A(m,n)}{B(m)} \to 0$. ■

APPENDIX B: PROOFS OF LEMMAS

The following result is based on a slight modification of the second inequality of (4.8) from Jiang and Li (2015) and a care taken by noticing that the version of the Wigner matrix here is $\sqrt{2}$ times of the version there.

Proof of Lemma 2. Review the proof of Lemma 4.1 from Jiang and Li (2015). Notice that the version of the Wigner matrix here is $\sqrt{2}$ times of the version there. From the second inequality in (4.8) in the paper, there is a positive constant C not depending on m such that $\mathbb{P}\big(\lambda_1(\tilde{W}_{\{1,\dots,m\}}) \geq x \text{ or } \lambda_m(\tilde{W}_{\{1,\dots,m\}}) \leq -x\big) \leq C\cdot\exp\big\{-\frac{x^2}{4}+Cm\log x+Cm\big\}$ for all $x > 4\sqrt{m}$ and all $m \geq 2$. Since the right side of this inequality is increasing in C, without loss of generality, we assume $C > 1$. It is easy to see $\log C \leq Cm \leq$

$Cm \log x$ under the assumption that $x > 4\sqrt{m}$. By taking $\kappa = 3C$ we get the desired conclusion. ∎

Proof of Lemma 3. Note that $\binom{p}{m} = \frac{p\cdot(p-1)\cdot\ldots\cdot(p-m+1)}{m!}$ and $\frac{p-l}{m-l} \geq \frac{p}{m}$ for $l \geq 0$. Thus, $\binom{p}{m} \geq \left(\frac{p}{m}\right)^m$. On the other hand, by the Sterling formula, $m! \geq \sqrt{2\pi} m^{m+1/2} e^{-m} > m^m e^{-m}$. Therefore, $\binom{p}{m} \leq \frac{p^m}{m!} < \frac{p^m}{m^m e^{-m}}$. Combining the two inequalities, we complete the proof. ∎

Proof of Lemma 4. Write $g(x) := (2m - x) - \frac{2m^2 - x^2}{m}(1-\varepsilon)^2 = 2m\big[1 - (1-\varepsilon)^2\big] + \frac{(1-\varepsilon)^2}{m}x^2 - x$ for $1 \leq x \leq m-1$. Obviously, $g(x)$ is a convex function. This leads to that $\max_{1\leq l\leq m-1} g(l) = g(1) \vee g(m-1)$. It is trivial to check that $g(1) \geq g(m-1)$. The first identity is thus obtained. The second identity follows from the first one. ∎

Proof of Lemma 1. Note that

$$\mathbb{E}\big(e^{\alpha Z}\mathbf{1}_{\{Z\geq\delta\}}\big) = \mathbb{E}\Big[\int_{-\infty}^{Z} \alpha e^{\alpha t} dt \cdot \mathbf{1}\{Z \geq \delta\}\Big] = \mathbb{E}\Big[\int_{-\infty}^{\infty} \alpha e^{\alpha t} dt \cdot \mathbf{1}\{Z \geq \delta \vee t\}\Big] dt.$$

Use the Fubini Theorem to see $\mathbb{E}(e^{\alpha Z}\mathbf{1}_{\{Z\geq\delta\}}) = \int_{-\infty}^{\delta} \alpha e^{\alpha t} dt \mathbb{P}(Z \geq \delta) + \alpha \int_{\delta}^{\infty} e^{\alpha t}\mathbb{P}(Z \geq t)dt = e^{\alpha\delta}\mathbb{P}(Z \geq \delta) + \alpha \int_{\delta}^{\infty} e^{\alpha t}\mathbb{P}(Z \geq t)dt$. ∎

Proof of Lemma 5. The technique to be used here is similar to that from Fey et al. (2008), where the large deviations for the extreme eigenvalues of Wishart matrices are developed. Thus we will omit the repetitive details and only state the main steps. To ease notation, we write $U_n = \frac{1}{n} W_{\{1,\ldots,m\}}$, and we use $\lambda_m(U_n)$ to denote its smallest eigenvalue. The event $\big\{\frac{\lambda_1(W_{\{1,\ldots,m\}})-n}{\sqrt{n}} \geq \sqrt{n}y\big\}$ is equal to $\{\lambda_1(U_n) \geq 1 + y\}$ and $\big\{\frac{\lambda_m(W_{\{1,\ldots,m\}})-n}{\sqrt{n}} \leq -\sqrt{n}y\big\}$ is equal to $\{\lambda_m(U_n) \leq 1-y\}$. We start to bound $\mathbb{P}(\lambda_m(U_n) \leq 1-y)$. Since $\lambda_m(W_{\{1,\ldots,m\}}) \geq 0$, we assume $y \in (0,1)$ without loss of generality.

Note that $\lambda_m(U_n) = \min_{v:\|v\|=1} v^\intercal U_n v$ and the sphere $S^{m-1} = \{v \in \mathbb{R}^m : \|v\| = 1\}$ can be covered by $\cup B(v^{(i)}, d)$ for some $v^{(1)}, \ldots, v^{(N_d)} \in S^{m-1}$. Here, we use $B(v,d)$ to denote an open ball centered around v with radius d. It is straightforward to verify that for any $v \in S^{m-1}$, there always exists $j \in \{1, \ldots, N_d\}$ such that

$$|v^\intercal U_n v - v^{(j)\intercal} U_n v^{(j)}| \leq 2\lambda_1(U_n)d. \tag{B.1}$$

Therefore, by considering $\{\lambda_1(U_n) \geq rm\}$ occurs or not, we have

$$\begin{aligned} & \mathbb{P}(\lambda_m(U_n) \leq 1-y) \\ = \; & \mathbb{P}(\min_{v:\|v\|=1} v^\intercal U_n v \leq 1-y) \\ \leq \; & N_d \cdot \sup_{v:\|v\|=1} \mathbb{P}(v^\intercal U_n v \leq 1 - y + 2dmr) + \mathbb{P}(\lambda_1(U_n) \geq mr) \qquad \text{(B.2)} \end{aligned}$$

for all $r > 0$. We next analyze N_d, $\mathbb{P}(v^\intercal U_n v \le 1-y+2dmr)$ and $\mathbb{P}(\lambda_1(U_n) \ge mr)$ separately.

We start with N_d, which is the minimum number of balls with the radius d required to cover S^{m-1}. By a result from Rogers (1963) we see $N_d = O\big(m^{1.5}(\log m)d^{-m}\big)$ for all $0 < d < 1/2$ and $m \ge 1$. As a result,

$$\log N_d = O\Big(m \log \frac{1}{d}\Big), \quad 0 < d < \frac{1}{2}. \tag{B.3}$$

We proceed to an upper bound for $\mathbb{P}(v^\intercal U_n v \le 1 - y + 2dmr)$. Recall that $U_n = \frac{1}{n}X^\intercal_{\cdot,[1,..,m]}X_{\cdot,[1,..,m]}$, where we use the notation $X_{\cdot,[1,..,m]} = (x_{ij})_{1\le i\le n, 1\le j\le m}$. Thus,

$$v^\intercal U_n v = \frac{1}{n}\|X_{\cdot,[1,..,m]}v\|^2 = \frac{1}{n}\sum_{i=1}^n S_{v,i}^2, \tag{B.4}$$

where we define $S_{v,i} = \sum_{l=1}^m X_{il}v_l$. Review $\|v\| = 1$. Since x_{ij}'s are standard normals, so are $\{S_{v,i};\ 1 \le i \le n\}$. By the large deviation bound for the sum of i.i.d. random variables [see, e.g., page 27 from Dembo and Zeitouni (1998)],

$$\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^n S_{v,i}^2 \in A\Big) \le 2\cdot \exp\big\{-n \inf_{x\in A} I(x)\big\} \tag{B.5}$$

where $A \subset \mathbb{R}$ is any Borel set and $I(x) = \sup_{t\in\mathbb{R}}\{tx - \log \mathbb{E}e^{tN(0,1)^2}\}$. Since $\log \mathbb{E}(e^{tN(0,1)^2}) = -\frac{1}{2}\log(1-2t)$ for $t < 1/2$, it is easy to check that

$$I(x) = \begin{cases} \frac{1}{2}(x - 1 - \log x), & \text{if } x > 0; \\ \infty, & \text{if } x \le 0. \end{cases}$$

Observe that $I(x)$ is decreasing for $x \le 1$. This together with (B.4) and (B.5) implies that

$$\mathbb{P}\left(v^\intercal U_n v \le 1 - y + 2dmr\right) \le e^{-nI(1-y+2dmr)} \tag{B.6}$$

for all $y > 2dmr$.

Now we estimate $\mathbb{P}(\lambda_1(U_n) \ge r)$ appeared in (B.2). Noting that U_n is semi-positive definite, we have $\lambda_1(U_n) \le \mathrm{trace}(U_n) \le \frac{1}{n}\sum_{i=1}^n\sum_{l=1}^m x_{il}^2$, and hence

$$\mathbb{P}\left(\lambda_1(U_n) \ge mr\right) \le \mathbb{P}\left(\frac{1}{mn}\sum_{i=1}^n\sum_{l=1}^m x_{il}^2 \ge r\right) \le e^{-mnI(r)} \tag{B.7}$$

for $r \geq 1$ by (B.5). Combining (B.3), (B.6), and (B.7), we obtain from (B.2) that

$$\begin{aligned} & \mathbb{P}(\lambda_m(U_n) \leq 1-y) \\ \text{(B.8)} \quad \leq \quad & \exp\Big\{ -nI(1-y+2dmr) + O\Big(m\log\frac{1}{d}\Big)\Big\} + \exp\{-mnI(r)\}. \end{aligned}$$

for $y > 2dmr$ and $r \geq 1$. This confirms (4.34).

To get (4.33), just notice $\lambda_1(U_n) = \max_{v:\|v\|=1} v^{\intercal}U_n v$. From (B.1) and (B.2) we see that

$$\mathbb{P}(\lambda_1(U_n) \geq 1+y) \leq N_d \cdot \sup_{v:\|v\|=1} \mathbb{P}(v^{\intercal}U_n v \geq 1+y-2dmr) + \mathbb{P}(\lambda_1(U_n) \geq mr).$$

Then (4.33) follows from similar arguments to (B.6)-(B.8). ■

Proof of Lemma 6. Review Assumption 1 in (3.1). We start with the analysis of (4.37). Here, we consider two sub-cases: $t \leq \frac{mn}{80\alpha}$ and $t > \frac{mn}{80\alpha}$. For $t \leq \frac{mn}{80\alpha}$, we have $r = \max(2, 1+\frac{80\alpha t}{mn}) = 2$ and

$$\begin{aligned} & \exp\left\{-\frac{1}{2}(r-1-\log r)mn + \alpha t + 2\log t + m\log p\right\} \\ \leq & \exp\left\{-\frac{1-\log 2}{2}mn + \frac{mn}{80} + 2\log\left(\frac{mn}{80\alpha}\right) + m\log p\right\}. \end{aligned}$$

Trivially $\frac{1-\log 2}{2} - \frac{1}{80} = 0.14\cdots > \frac{1}{10}$. Note that $\log\frac{mn}{80\alpha} = o(mn)$ and $m\log p = o(mn)$ under Assumption 1 in (3.1). It follows that

$$-\frac{1-\log 2}{2}mn + \frac{mn}{80} + 2\log\left(\frac{mn}{80\alpha}\right) + m\log p \leq -\Big[\frac{1}{10} + o(1)\Big]mn.$$

This implies

$$\lim_{n\to\infty} \sup_{\frac{\delta\sqrt{n}}{100} \leq t \leq \frac{mn}{80\alpha}} \exp\left\{-\frac{1}{2}(r-1-\log r)mn + \alpha t + 2\log t + m\log p\right\} = 0. \tag{B.9}$$

Now we consider another sub-case where $t \geq \frac{mn}{80\alpha}$. For this case, $r = 1+\frac{80\alpha t}{mn}$. It is not hard to see $r-1-\log r \geq \frac{r}{12}$ for $r \geq 2$. Apparently, $m\log p \leq \alpha t$ for $t \geq \frac{mn}{80\alpha}$ as n is sufficiently large. For this range of t, it is easy to verify that $\exp\big\{-\frac{1}{2}(r-1-\log r)mn + \alpha t + 2\log t + m\log p\big\} \leq \exp\big\{-\big(\frac{4\alpha}{3} + o(1)\big)t\big\}$. This implies

$$\lim_{n\to\infty} \sup_{t \geq \frac{mn}{80\alpha}} \exp\left\{-\frac{1}{2}(r-1-\log r)mn + \alpha t + 2\log t + m\log p\right\} = 0. \tag{B.10}$$

Combining (B.9) and (B.10), we obtain

$$\lim_{n\to\infty}\sup_{t\geq\frac{\delta\sqrt{n}}{100}}\exp\left\{-\frac{1}{2}(r-1-\log r)mn+\alpha t+2\log t+m\log p\right\}=0.$$

This completes the proof of (4.37). We next show (4.36).

Recall $z=\frac{2\sqrt{m\log p}}{\sqrt{n}}+\frac{t}{2\sqrt{n}}\geq\frac{t}{2\sqrt{n}}$. Obviously, $z>\frac{\delta}{200}$ as $t>\frac{\delta\sqrt{n}}{100}$. It is elementary to check there exists $\varepsilon>0$ such that $x-\log(1+x)\geq\varepsilon x$ for all $x>\frac{\delta}{200}$. Hence, $\frac{n}{2}[z-\log(1+z)]\geq\frac{1}{2}n\varepsilon z\geq\frac{\varepsilon}{4}\sqrt{n}\,t$ for all $t>\frac{\delta\sqrt{n}}{100}$. Reviewing $r=\max(2,1+\frac{80\alpha t}{mn})$ and $d=\min(\frac{1}{2},\frac{t}{4m\sqrt{nr}})$, we have

$$m\log\frac{1}{d}=O\big(m\log n+m\log\big(1+\frac{80\alpha t}{mn}\big)\big)=O\big(m\log n+\frac{t}{n}\big)$$

since $0<\log(1+x)<x$ for all $x>0$. Furthermore, $\alpha t+2\log t\leq 2\alpha t$ as t is sufficiently large, and $m\log p=o(mn)$ by Assumption 1 in (3.1). Consequently,

$$\begin{aligned}&\sup_{t>\frac{\delta\sqrt{n}}{100}}\exp\left\{-\frac{n}{2}(z-\log(1+z))+\kappa m\log\frac{1}{d}+\alpha t+2\log t+m\log p\right\}\\ \leq&\sup_{t>\frac{\delta\sqrt{n}}{100}}\exp\left\{-\frac{\varepsilon}{4}\sqrt{n}\,t+3\alpha t+O(m\log n)\right\}\\ =&\exp\left\{-(1+o(1))\frac{\varepsilon}{4}\sqrt{n}\,t\right\}\\ =&o(1).\end{aligned}$$

We obtain (4.36) and the proof is completed. ■

Proof of Lemma 7. Recall the assumption that $\delta\in(0,1)$. Then $z=\frac{2\sqrt{m\log p}}{\sqrt{n}}+\frac{t}{2\sqrt{n}}\leq\frac{2\sqrt{m\log p}}{\sqrt{n}}+\frac{\delta}{200}\leq 1$ as n is sufficiently large. Now,

$$z^2=\Big(\frac{2\sqrt{m\log p}}{\sqrt{n}}+\frac{t}{2\sqrt{n}}\Big)^2\geq\frac{4m\log p}{n}+\frac{2t\sqrt{m\log p}}{n}.$$

It is trivial to show that $z-\log(1+z)\geq\frac{1}{4}z^2$ for $0\leq z\leq 1$. By this inequality, we see

$$z-\log(1+z)\geq\frac{m\log p}{n}+\frac{t\sqrt{m\log p}}{2n}. \tag{B.11}$$

Now, reviewing $d=\frac{t}{8m\sqrt{n}}$ and $t\geq\delta$, we have $m\log(1/d)=O(m\log m+$

$m\log n) = O(m\log n)$. This joint with (B.11) implies that

$$\begin{aligned}&\exp\left\{-\frac{n}{2}[z-\log(1+z)]+\kappa m\log\frac{1}{d}+\alpha t+2\log t+m\log p\right\}\\ \leq&\exp\left\{-\frac{1}{4}t\sqrt{m\log p}+O(m\log n)+\alpha t+2\log t\right\}\\ =&\exp\left\{\big[-\frac{1}{4}+o(1)\big]t\sqrt{m\log p}+O(m\log n)\right\}.\end{aligned}$$

Since $t \geq (m/\log p)^{1/2}\xi_p\log n$, we know $O(m\log n) = o(t\sqrt{m\log p})$ uniformly in t. Thus,

$$\begin{aligned}&\exp\left\{-\frac{n}{2}[z-\log(1+z)]+\kappa m\log\frac{1}{d}+\alpha t+2\log t+m\log p\right\}\\ \leq&\exp\left\{-\big[\frac{1}{2}+o(1)\big]t\sqrt{m\log p}\right\}\\ \leq&\exp\left\{-\frac{1}{4}t\sqrt{m\log p}\right\}\end{aligned}$$

as n is sufficiently large. We then get (4.40). Evidently, $\sup_{\delta\vee\omega_n\leq t\leq\frac{\delta\sqrt{n}}{100}}\{\alpha t+2\log t\} = O(\sqrt{n}\,)$ as $n\to\infty$. This implies that

$$-\frac{1}{2}(1-\log 2)mn+\alpha t+2\log t+m\log p=-\frac{1-\log 2}{2}[1+o(1)]mn.$$

The assertion (4.41) is verified. ■

Proof of Lemma 8. Review the notation $W_{\{1,\ldots,m\}}$ above (2.1) with $S=\{1,\cdots,m\}$. Let $\mu_1>\ldots>\mu_m$ be the eigenvalues of $W_{\{1,\ldots,m\}}$. According to James (1964) or Muirhead (2009), $\mu=(\mu_1,\ldots,\mu_m)$ has density function

$$f_{m,n}(\mu)=c_{m,n}e^{-\frac{\sum_{i=1}^m\mu_i}{2}}\prod_{i=1}^m\mu_i^{\frac{n-m+1}{2}-1}\prod_{1\leq j<i\leq m}(\mu_j-\mu_i)I(\mu_1>\ldots>\mu_m>0),$$

where $c_{m,n}=m!2^{-nm/2}\prod_{j=1}^m\frac{\Gamma(3/2)}{\Gamma(1+(j/2))\Gamma((n-m+j)/2)}$. In addition, $\lambda=(\lambda_1,\ldots,\lambda_m)$ has density

$$h_m(\lambda)=c_me^{-\frac{1}{4}\sum_{k=1}^m\lambda_k^2}\prod_{1\leq j<i\leq m}(\lambda_j-\lambda_i)\text{ with}$$
$$c_m=m!2^{-m}2^{-m(m-1)/4}\pi^{-m/2}\prod_{j=1}^m\frac{\Gamma(3/2)}{\Gamma(1+(j/2))};$$

see, for example, Chapter 17 from Mehta (2004). Note that $\nu_i = (\mu_i - n)/\sqrt{n}$, so we can write down the expression of $g_{n,m}$ as follows.

$$
\begin{aligned}
& g_{n,m}(v) \\
=& n^{m/2} c_{m,n} \exp\left(-\frac{1}{2}\sum_{i=1}^{m}(\sqrt{n}v_i + n)\right)\prod_{i=1}^{m}(\sqrt{n}v_i + n)^{\frac{n-m+1}{2}-1} \\
& \cdot \prod_{1\le j<i\le m}(\sqrt{n}v_j - \sqrt{n}v_i) \\
=& n^{m/2} e^{-nm/2} n^{m(n-m+1)/2-m} n^{m(m-1)/4} c_{m,n} \\
& \cdot e^{-(\sqrt{n}/2)\sum_{i=1}^{m} v_i} \prod_{i=1}^{m}\left(1 + \frac{v_i}{\sqrt{n}}\right)^{\frac{n-m-1}{2}} \prod_{1\le j<i\le m}(v_j - v_i)
\end{aligned}
$$

for $v_1 > v_2 > ... > v_m > -\sqrt{n}$ and $g_{n,m}(v) = 0$, otherwise. Denote $C(m,n) = n^{m/2} e^{-nm/2} n^{m(n-m+1)/2-m} n^{m(m-1)/4} c_{m,n}$. Then,

$$
\begin{aligned}
& \log g_{n,m}(v) - \log h_m(v) \\
=& \log C(m,n) - \log c_m + \sum_{i=1}^{m}\left[-\frac{\sqrt{n}}{2}v_i + \frac{n-m-1}{2}\log\left(1 + \frac{v_i}{\sqrt{n}}\right)\right] + \frac{1}{4}\sum_{i=1}^{m} v_i^2
\end{aligned}
$$

for $v_1 > v_2 > ... > v_m > -\sqrt{n}$. By Lemma A.1 in Appendix A,

$$
\begin{aligned}
& \log g_{n,m}(v) - \log h_m(v) \\
= & o(1) + \sum_{i=1}^{m}\left[-\frac{\sqrt{n}}{2}v_i + \frac{n-m-1}{2}\log\left(1 + \frac{v_i}{\sqrt{n}}\right) + \frac{1}{4}v_i^2\right]
\end{aligned}
\tag{B.12}
$$

for $v_1 > v_2 > ... > v_m > -\sqrt{n}$. By the Taylor expansion, $\left|\log(1+x) - \left(x - \frac{x^2}{2}\right)\right| \le \sum_{i=3}^{\infty}\frac{|x|^k}{k} \le \frac{|x|^3}{3(1-|x|)}$ for all $|x| < 1$. Therefore,

$$
\left|\log(1+x) - \left(x - \frac{x^2}{2}\right)\right| \le |x|^3 \tag{B.13}
$$

for $|x| < \frac{2}{3}$. Writing $\frac{n-m-1}{2} = \frac{n}{2} - \frac{m+1}{2}$, it is easy to check

$$
\sum_{i=1}^{m}\left[-\frac{\sqrt{n}}{2}v_i + \frac{n-m-1}{2}\left(\frac{v_i}{\sqrt{n}} - \frac{v_i^2}{2n}\right) + \frac{1}{4}v_i^2\right] = -\frac{m+1}{2}\sum_{i=1}^{m}\left(\frac{v_i}{\sqrt{n}} - \frac{v_i^2}{2n}\right). \tag{B.14}
$$

Combining (B.12)-(B.14), and noting that $|v_i| \le \|v\|_\infty$ for all i, we get

$$\begin{aligned}&\log g_{n,m}(v) - \log h_m(v)\\ =&o(1) + \varpi_{m,n}\\ &+ \sum_{i=1}^{m}\left[-\frac{\sqrt{n}}{2}v_i + \frac{n-m-1}{2}\Big(\frac{v_i}{\sqrt{n}} - \frac{v_i^2}{2n}\Big) + \frac{1}{4}v_i^2\right]\\ =&o(1) + \varpi_{m,n} - \frac{m+1}{2}\sum_{i=1}^{m}\Big(\frac{v_i}{\sqrt{n}} - \frac{v_i^2}{2n}\Big)\end{aligned}$$

provided $\|v\|_\infty \le \frac{2}{3}\sqrt{n}$, where $\varpi_{m,n}$ is the error term and it is controlled by

$$|\varpi_{m,n}| \le \frac{n-m-1}{2}\cdot\frac{1}{n^{3/2}}\sum_{i=1}^{m}|v_i|^3 \le \frac{1}{n^{1/2}}\sum_{i=1}^{m}|v_i|^3.$$

By using the trivial bound that $|v_i| \le \|v\|_\infty$ for each i, we obtain the desired conclusion from the above two assertions. ■

Proof of Lemma 9. According to the density function of the Wishart distribution [see, e.g., Anderson (1962) or Muirhead (2009)], the density function for $W_{\{1,\ldots,m\}}$ is

$$f_{W,m}(V) = \frac{|V|^{(n-m-1)/2}e^{-tr(V)/2}}{\Gamma_m(\frac{n}{2})2^{mn/2}},$$

for every $m\times m$ positive definite matrix V, where $\Gamma_m(\cdot)$ is the multivariate gamma function defined by $\Gamma_m\left(\frac{n}{2}\right) = \pi^{m(m-1)/4}\prod_{j=1}^{m}\Gamma\left(\frac{n-j+1}{2}\right)$, and we write $|V|$ for the determinant of a matrix V. It is easy to see that the density function for $\frac{W_{\{1,\ldots,m\}}-nI_m}{\sqrt{n}}$ is given by

$$\begin{aligned}f_{m,n}(w) :=&(\sqrt{n}\,)^{\frac{m(m+1)}{2}}f_{W,m}(\sqrt{n}w + nI_m)\\ =&n^{m(m+1)/4}\frac{|\sqrt{n}w + nI_m|^{(n-m-1)/2}e^{-tr(\sqrt{n}w+nI_m)/2}}{\Gamma_m(\frac{n}{2})2^{mn/2}}\end{aligned}$$

for every $m\times m$ matrix w such that $w+\sqrt{n}I_m$ is positive definite. Simplifying the above display, we further have

$$f_{m,n}(w) = A(m,n)\exp\left\{\frac{n-m-1}{2}\log\left|1+\frac{w}{\sqrt{n}}\right| - \frac{1}{2}\sqrt{n}\,tr(w)\right\},$$

where $A(m,n)=n^{m(m+1)/4+m(n-m-1)/2}e^{-nm/2}2^{-\frac{nm}{2}}/\Gamma_m(n/2)$. On the other hand, $\tilde{f}_m(w)=B(m)e^{-\frac{tr(w^2)}{4}}$, where $B(m)=(2\pi)^{-m(m+1)/2}2^{-m/2}$; see, for instance, Mehta (2004). Now we consider

$$\begin{aligned}
&\log f_{m,n}(w)-\log\tilde{f}_m(w)\\
=&\log A(m,n)-\log B(m)\\
&\qquad\qquad+\frac{n-m-1}{2}\log\left|1+\frac{w}{\sqrt{n}}\right|-\frac{1}{2}\sqrt{n}\,tr(w)+\frac{1}{4}tr(w^2)\\
=&o(1)+\sum_{i=1}^{m}\left[\frac{n-m-1}{2}\log\Big(1+\frac{\lambda_i}{\sqrt{n}}\Big)-\frac{1}{2}\sqrt{n}\lambda_i+\frac{1}{4}\lambda_i^2\right]
\end{aligned}$$

for every $\lambda_i>-\sqrt{n}$ and $i=1,\cdots,m$ by Lemma A.2 in Appendix A, where $\lambda_1,\cdots\lambda_m$ are the eigenvalues of w. From (B.13) and (B.14),

$$\log f_{m,n}(w)-\log\tilde{f}_m(w)=o(1)+\varepsilon_{m,n}-\frac{m+1}{2}\sum_{i=1}^{m}\Big(\frac{\lambda_i}{\sqrt{n}}-\frac{\lambda_i^2}{2n}\Big)$$

if $\max_{1\le i\le m}|\lambda_i|\le\frac{2}{3}\sqrt{n}$, where $\varepsilon_{m,n}$ is the error term satisfying

$$|\varepsilon_{m,n}|\le\frac{n-m-1}{2}\sum_{i=1}^{m}\frac{|\lambda_i|^3}{n^{3/2}}\le\frac{m}{n^{1/2}}\max_{1\le i\le m}|\lambda_i|^3=mn^{-1/2}\|w\|^3.$$

In addition, $|\sum_{i=1}^m\lambda_i|\le m\max_{1\le i\le m}|\lambda_i|=m\|w\|$, and $\sum_{i=1}^m\lambda_i^2\le m\max_{1\le i\le m}|\lambda_i|^2=m\|w\|_2^2$, The above three assertions lead to

$$\begin{aligned}
&\log f_{m,n}(w)-\log\tilde{f}_m(w)\\
=\;&o(1)+O\Big(m^2n^{-1/2}\|w\|+m^2n^{-1}\|w\|^2+n^{-1/2}m\|w\|^3\Big)
\end{aligned}$$

provided $\|w\|=\max_{1\le i\le m}|\lambda_i|\le\frac{2}{3}\sqrt{n}$. The proof is finished. ■

Proof of Lemma 11. Let $B(v_1,\delta),..,B(v_N,\delta)$ be N balls centered around $v_1,..,v_N$, respectively, such that $\cup B(v_i,\delta)$ covers the unit sphere $\{v\in\mathbb{R}^m;\ \|v\|=1\}$. Then, for any $r>0$, by (B.1),

$$\begin{aligned}
&\mathbb{P}\big(\sup_{\|v\|=1}v^\intercal\tilde{W}_{\{1,..,m\}}v\ge x\big)\\
\text{(B.15)}\le\;&N\cdot\max_{1\le i\le N}\mathbb{P}\big(v_i^\intercal\tilde{W}_{\{1,..,m\}}v_i\ge x-2r\delta\big)+\mathbb{P}\big(\lambda_1(\tilde{W}_{\{1,...,m\}})\ge r\big).
\end{aligned}$$

According to the distribution of $\tilde{W}$, $v_i^\intercal\tilde{W}v_i\sim N(0,f(v_i))$, where $f(y):=f(y)=2+(\eta-2)\sum_{i=1}^m y_i^4$ for any $y=(y_1,\cdots,y_m)^\intercal\in\mathbb{R}^m$. In fact, for any

$y = (y_1, \cdots, y_m)^\intercal \in \mathbb{R}^m$, $y^\intercal \tilde{W} y = \sum_{i=1}^m \tilde{w}_{ii} y_i^2 + 2\sum_{i<j} \tilde{w}_{ij} y_i y_j \sim N(0, \sigma_y^2)$ such that σ_y^2 is equal to

$$\mathbb{E}\left(\sum_{i=1}^m \tilde{w}_{ii} y_i^2 + 2\sum_{i<j} \tilde{w}_{ij} y_i y_j\right)^2 = \eta \sum_{i=1}^m y_i^4 + 4\sum_{i<j} y_i^2 y_j^2 = f(y)$$

by independence. Recall $\mathbb{P}(N(0,1) \geq x) \leq e^{-x^2/2}$ for all $x \geq 1$. Thus, for $x - 2r\delta > 1$, the first term on the right side of (B.15) is bounded by

$$\begin{aligned} & N \cdot \sup_{\|y\|=1} \exp\left\{-\frac{(x-2r\delta)^2}{2(\sum_{i=1}^m y_i^4(\eta-2)+2)}\right\} \\ =& N \cdot \exp\left\{-\frac{(x-2r\delta)^2}{2\left[(\inf_{\|y\|=1}\sum_{i=1}^m y_i^4)(\eta-2)+2\right]}\right\}, \end{aligned}$$

since $0 \leq \eta \leq 2$. Observe that $\inf_{\|y\|=1} \sum_{i=1}^m y_i^4 = \frac{1}{m}$. Thus,
(B.16)

$$N \cdot \max_{1\leq i\leq N} \mathbb{P}\left(v_i^\intercal \tilde{W}_{\{1,..,m\}} v_i \geq x - 2r\delta\right) \leq N \cdot \exp\left\{-\frac{(x-2r\delta)^2}{2\left[m^{-1}(\eta-2)+2\right]}\right\}$$

if $x - 2r\delta > 1$. Now turn to estimate the last probability in (B.15). Note that

$$\lambda_1(\tilde{W}_{\{1,..,m\}}) \leq \left(tr\left(\tilde{W}^2_{\{1,..,m\}}\right)\right)^{1/2} = \left(\sum_{i=1}^m \sum_{j=1}^m \tilde{W}_{ij}^2\right)^{1/2}.$$

Note that $\sum_{i=1}^m \sum_{j=1}^m \tilde{W}_{ij}^2$ and $\eta Q_1 + 2Q_2$ have the same distribution, where $Q_1 \sim \chi^2_m$, $Q_2 \sim \chi^2_{m(m-1)/2}$ and Q_1 and Q_2 are independent. Also $\eta Q_1 + 2Q_2 \leq 2(Q_1 + Q_2) \sim 2 \cdot \chi^2_{m(m+1)/2}$ Thus, the last probability in (B.15) is dominated by

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^m \sum_{j=1}^m \tilde{W}_{ij}^2 \geq r^2\right) &\leq \mathbb{P}\left(\chi^2_{m(m+1)/2} \geq r^2/2\right) \\ &\leq \mathbb{P}\left(\frac{\chi^2_{m(m+1)/2}}{m(m+1)/2} \geq \frac{r^2}{m(m+1)}\right). \end{aligned}$$

Notice $\frac{r^2}{m(m+1)} \geq 8$ under the given condition $r \geq 4m$. Let $I(x) = \frac{1}{2}(x - 1 - \log x)$ for $x > 0$. It is easy to check that $I(8) = (7 - \log 8)/2 > 2.4$ and that

$I(x) = \frac{1}{2}(x - 1 - \log x) \geq \frac{1}{4}x$ as $x \geq 8$. By (B.5), the last probability above is no more than $2 \cdot \exp\Big(-\frac{m(m+1)}{2} I\big(\frac{r^2}{m(m+1)}\big)\Big) \leq 2 \cdot e^{-r^2/8}$. Hence,

$$\mathbb{P}\big(\lambda_1(\tilde{W}_{\{1,\dots,m\}}) \geq r\big) \leq \mathbb{P}\left(\sum_{i=1}^{m}\sum_{j=1}^{m} \tilde{W}_{ij}^2 \geq r^2\right) \leq 2 \cdot e^{-r^2/8}.$$

Combining the above display with (B.15) and (B.16), we have

$$\mathbb{P}\Big(\sup_{\|v\|=1} v^{\intercal}\tilde{W}_{\{1,..,m\}}v \geq x\Big) \leq N \cdot \exp\left\{-\frac{(x-2r\delta)^2}{2\,[m^{-1}(\eta-2)+2]}\right\} + 2 \cdot e^{-r^2/8}.$$

The desired conclusion follows since $N \leq m^{1.5}(\log m)\delta^{-m}$ (Rogers, 1963). ■

Proof of Lemma 12. (i) $\Longrightarrow$ (ii): Easily, $\mathbb{E}\big[e^{\alpha Z_p}\big] \leq \mathbb{E}\big[e^{\alpha Z_p}\mathbf{1}_{\{Z_p \geq \delta\}}\big] + e^{\alpha\delta}$. Taking $\limsup_{p\to\infty}$ on both sides and then letting $\delta \downarrow 0$, we obtain $\limsup_{p\to\infty}\mathbb{E}\big[e^{\alpha Z_p}\big] \leq 1$. On the other side, $\liminf_{p\to\infty}\mathbb{E}\big[e^{\alpha Z_p}\big] \geq 1$ since $Z_p \geq 0$. Hence, $\lim_{p\to\infty}\mathbb{E}\big[e^{\alpha Z_p}\big] = 1$.

(ii) $\Longrightarrow$ (i): For each $\beta > 0$, we know $\mathbf{1}_{\{Z_p \geq \delta\}} \leq e^{\beta(Z_p - \delta)}$. Thus,

$$\mathbb{E}[e^{\alpha Z_p}\mathbf{1}_{\{Z_p \geq \delta\}}] \leq \mathbb{E}[e^{\alpha Z_p + \beta(Z_p - \delta)}] = e^{-\beta\delta}\mathbb{E}[e^{(\alpha+\beta)Z_p}].$$

Taking $\limsup_{p\to\infty}$ on both sides and then letting $\beta \to \infty$, we obtain $\lim_{p\to\infty}\mathbb{E}[e^{\alpha Z_p}\mathbf{1}_{\{Z_p \geq \delta\}}] = 0$.

(ii) $\Longrightarrow$ (iii): First, $\mathbb{E}[Z_p^{\alpha}] = \alpha\int_0^{\infty} x^{\alpha-1}P(Z_p \geq x)\,dx$. By the Markov inequality, $\mathbb{P}(Z_p \geq x) \leq e^{-\beta x}\mathbb{E}e^{\beta Z_p}$ for all $x > 0$ and $\beta > 0$. It follows that $\mathbb{E}[Z_p^{\alpha}] \leq \alpha\left(\mathbb{E}e^{\beta Z_p}\right)\int_0^{\infty} x^{\alpha-1}e^{-\beta x}\,dx = \frac{\alpha\Gamma(\alpha)}{\beta^{\alpha}}\mathbb{E}e^{\beta Z_p}$ for all $\beta > 0$. The conclusion then follows by first letting $p \to \infty$ and then sending $\beta \to \infty$.

(iii) $\Longrightarrow$ (iv): This is a direct consequence of the Chebyshev inequality and the equality $\lim_{p\to\infty}\mathbb{E}(Z_p) = 0$.

(iii) $\Longrightarrow$ (v): Let $\alpha = 2$ in (iii), then $\limsup_p \mathrm{Var}(Z_p) \leq \lim_{p\to\infty}\mathbb{E}(Z_p^2) = 0$. ■

T. T. Cai
Department of Statistics
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104
E-mail: tcai@wharton.upenn.edu

T. Jiang
School of Statistics
University of Minnesota
Minneapolis, MN 55455
E-mail: jiang040@umn.edu

X. Li
School of Statistics
University of Minnesota
Minneapolis, MN 55455
E-mail: lixx1766@umn.edu