The Annals of Statistics
2013, Vol. 41, No. 5, 2572–2607
DOI: 10.1214/13-AOS1154

CONVERGENCE RATES OF EIGENVECTOR EMPIRICAL SPECTRAL DISTRIBUTION OF LARGE DIMENSIONAL SAMPLE COVARIANCE MATRIX

BY NINGNING XIA[1], YINGLI QIN[2] AND ZHIDONG BAI[3]

Northeast Normal University and National University of Singapore, University of Waterloo, and Northeast Normal University and National University of Singapore

The eigenvector Empirical Spectral Distribution (VESD) is adopted to investigate the limiting behavior of eigenvectors and eigenvalues of covariance matrices. In this paper, we shall show that the Kolmogorov distance between the expected VESD of sample covariance matrix and the Marčenko–Pastur distribution function is of order $O(N^{-1/2})$. Given that data dimension n to sample size N ratio is bounded between 0 and 1, this convergence rate is established under finite 10th moment condition of the underlying distribution. It is also shown that, for any fixed $\eta > 0$, the convergence rates of VESD are $O(N^{-1/4})$ in probability and $O(N^{-1/4+\eta})$ almost surely, requiring finite 8th moment of the underlying distribution.

1. Introduction and main results. Let $\mathbf{X}_i = (X_{1i}, X_{2i}, \dots, X_{ni})^T$ and $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_N)$ be an $n \times N$ matrix of i.i.d. (independent and identically distributed) complex random variables with mean 0 and variance 1. We consider, a class of sample covariance matrices

$$\mathbf{S}_n = \frac{1}{N} \sum_{k=1}^{N} \mathbf{X}_k \mathbf{X}_k^* = \frac{1}{N} \mathbf{X}\mathbf{X}^*,$$

where $\mathbf{X}^*$ denotes the conjugate transpose of the data matrix $\mathbf{X}$. The Empirical Spectral Distribution (ESD) $F^{\mathbf{S}_n}(x)$ of $\mathbf{S}_n$ is then defined as

$$F^{\mathbf{S}_n}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(\lambda_i \le x), \tag{1.1}$$

where $\lambda_1 \le \cdots \le \lambda_n$ are the eigenvalues of $\mathbf{S}_n$ in ascending order and $\mathrm{I}(\cdot)$ is the conventional indicator function.

Received February 2013; revised June 2013.

[1]Supported by the Fundamental Research Funds for the Central Universities 11SSXT131.

[2]Supported by University of Waterloo Start-up Grant.

[3]Supported by NSF China Grant 11171057, PCSIRT Fundamental Research Funds for the Central Universities, and NUS Grant R-155-000-141-112.

MSC2010 subject classifications. Primary 15A52, 60F15, 62E20; secondary 60F17, 62H99.

Key words and phrases. Eigenvector empirical spectral distribution, empirical spectral distribution, Marčenko–Pastur distribution, sample covariance matrix, Stieltjes transform.

Marčenko and Pastur in [16] proved that with probability 1, $F^{\mathbf{S}_n}(x)$ converges weakly to the standard Marčenko–Pastur distribution $F_y(x)$ with density function

$$p_y(x) = \frac{dF_y(x)}{dx} = \frac{1}{2\pi xy}\sqrt{(x-a)(b-x)}\mathrm{I}(a \le x \le b), \tag{1.2}$$

where $a = (1-\sqrt{y})^2$ and $b = (1+\sqrt{y})^2$. Here the positive constant y is the limit of dimension to sample size ratio when both n and N tend to infinity.

In applications of asymptotic theorems of spectral analysis of large dimensional random matrices, one of the important problems is the convergence rate of the ESD. The Kolmogorov distance between the expected ESD of $\mathbf{S}_n$ and the Marčenko–Pastur distribution $F_y(x)$ is defined as

$$\Delta = \|\mathrm{E}F^{\mathbf{S}_n} - F_y\| = \sup_x \big|\mathrm{E}F^{\mathbf{S}_n}(x) - F_y(x)\big|$$

as well as the distance between two distributions $F^{\mathbf{S}_n}(x)$ and $F_y(x)$,

$$\Delta_p = \|F^{\mathbf{S}_n} - F_y\| = \sup_x \big|F^{\mathbf{S}_n}(x) - F_y(x)\big|.$$

Notice that, for any constant $C > 0$,

$$P(\Delta_p \ge C) = P\Big\{\sup_x \big|F^{\mathbf{S}_n}(x) - F_y(x)\big| \ge C\Big\} \le C^{-1}\mathrm{E}\Delta_p.$$

Thus, Δ_p measures the rate of convergence in probability.

Bai in [2, 3] firstly tackled the problem of convergence rate and established three Berry–Esseen type inequalities for the difference of two distributions in terms of their Stieltjes transforms. Götze and Tikhomirov in [11] further improved the Berry–Esseen type inequlatiy and showed the convergence rate of $F^{\mathbf{S}_n}(x)$ is $O(N^{-1/2})$ in probability under finite 8th moment condition. More recently, a sharper bound is obtained by Pillai and Yin in [18], under a stronger condition, that is, the sub-exponential decay assumption. It is shown that the difference between eigenvalues of $\mathbf{S}_n$ and the Marčenko–Pastur distribution is of order $O(N^{-1}(\log N)^{O(\log\log N)})$ in probability.

In the literature, research on limiting properties of eigenvectors of large dimensional sample covariance matrices is much less developed than that of eigenvalues, due to the cumbersome formulation of the eigenvectors. Some great achievements have been made in proving the properties of eigenvectors for large dimensional sample covariance matrices, such as [4, 19–22], and that for Wigner matrices, such as [10, 13, 25].

However, the eigenvectors of large sample covariance matrices play an important role in high-dimensional statistical analysis. In particular, due to the increasing availability of high-dimensional data, principal component analysis (PCA) has been favorably recognized as a powerful technique to reduce dimensionality. The eigenvectors corresponding to the leading eigenvalues are the directions of the principal components. Johnstone [12] proposed the spiked eigenvalue model to test the existence of principal component. Paul [17] discussed the length of the eigenvector corresponding to the spiked eigenvalue.

In PCA, the eigenvectors $(\boldsymbol{\nu}_1^0, \ldots, \boldsymbol{\nu}_n^0)$ of population covariance matrix $\boldsymbol{\Sigma}$ determine the directions in which we project the observed data and the corresponding eigenvalues $(\lambda_1^0, \ldots, \lambda_n^0)$ determine the proportion of total variability loaded on each direction of projections. In practice, the (sample) eigenvalues $(\lambda_1, \ldots, \lambda_n)$ and eigenvectors $(\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_n)$ of the sample covariance matrix $\mathbf{S}_n$ are used in PCA. In [1], Anderson has shown the following asymptotic distribution for the sample eigenvectors $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_n$ when the observations are from a multivariate normal distribution of covariance matrix $\boldsymbol{\Sigma}$ with distinct eigenvalues:

$$\sqrt{N}(\boldsymbol{\nu}_i - \boldsymbol{\nu}_i^0) \xrightarrow{d} N_n(0, \mathbf{D}_i),$$

where

$$\mathbf{D}_i = \lambda_i^0 \sum_{k=1, k\neq i}^{n} \frac{\lambda_k^0}{(\lambda_k^0 - \lambda_i^0)^2} \boldsymbol{\nu}_k^0 \boldsymbol{\nu}_k^{0T}.$$

However, this is a large sample result when the dimension n is fixed and low. In particular, if $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n$, then the eigenmatrix (matrix of eigenvectors) should be asymptotically isotropic when the sample size is large. That is, the eigenmatrix should be asymptotically Haar, under some minor moment conditions. However, when the dimension is large (increasing), the Haar property is not easy to formulate.

Motivated by the orthogonal iteration method, [15] proposed an iterative thresholding method to estimate sparse principal subspaces (spanned by the leading eigenvectors of $\boldsymbol{\Sigma}$) in high dimensional and spiked covariance matrix setting. The convergence rates of the proposed estimators are provided. By reducing the sparse PCA problem to a high-dimensional regression problem, [9] established the optimal rates of convergence for estimating the principal subspace with respect to a large collection of spiked covariance matrices. See the reference therein for more literature on sparse PCA and spiked covariance matrices.

To perform the test of existence of spiked eigenvalues, one has to investigate the null properties of the eigenmatrices, that is, when $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n$ (i.e., nonspiked). Then the eigenmatrix should be asymptotically isotropic, when the sample size is large. That is, the eigenmatrix should be asymptotically Haar. However, when the dimension is large, the Haar property is not easy to formulate. The recent development in random matrix theory can help us investigate the large dimension and large sample properties of eigenvectors. We will adopt the VESD, defined later in the paper, to characterize the asymptotical Haar property so that if the eigenmatrix is Haar, then the process defined the VESD tends to a Brownian bridge. Conversely, if the process defined by the VESD tends to a Brownian bridge, then it indicates a similarity between the Haar distribution and that of the eigenmatrix.

Therefore, studying the large sample and large dimensional results of the VESD can assist us in better examining spiked covariance matrix as assumed by [15] and [9] among many others.

Let $\mathbf{U}_n\mathbf{\Lambda}_n\mathbf{U}_n^*$ denote the spectral decomposition of $\mathbf{S}_n$, where $\mathbf{\Lambda}_n = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $\mathbf{U}_n = (u_{ij})_{n\times n}$ is a unitary matrix consisting of the corresponding orthonormal eigenvectors of $\mathbf{S}_n$. For each n, let $\mathbf{x}_n \in \mathbb{C}^n$, $\|\mathbf{x}_n\| = 1$ be nonrandom and let $\mathbf{d}_n = \mathbf{U}_n^*\mathbf{x}_n = (d_1, \ldots, d_n)^*$, where $\|\mathbf{x}_n\|$ denotes Euclidean norm of $\mathbf{x}_n$.

Define a stochastic process $X_n(t)$ by

$$X_n(t) = \sqrt{n/2}\sum_{j=1}^{[nt]}\left(|d_j|^2 - \frac{1}{n}\right), \qquad [a] \text{ denotes the greatest integer} \le a.$$

If $\mathbf{U}_n$ is Haar distributed over the orthogonal matrices, then $\mathbf{d}_n$ would be uniformly distributed over the unit sphere in $\mathbb{R}^n$, and the limiting distribution of $X_n(t)$ is a unique Brownian bridge $B(t)$ when n tends to infinity. In this paper, we use the behavior of $X_n(t)$ for all $\mathbf{x}_n$ to reflect the uniformity of $\mathbf{U}_n$. The process $X_n(t)$ is considerably important for us to understand the behavior of the eigenvectors of $\mathbf{S}_n$.

Motivated by Silverstein's ideas in [19–22], we want to examine the limiting properties of $\mathbf{U}_n$ through stochastic process $X_n(t)$. We claim that $\mathbf{U}_n$ is "asymptotically Haar distributed," which means $X_n(t)$ converges to a Brownian bridge $B(t)$. In [21], it showed that the weak convergence of $X_n(t)$ converging to a Brownian bridge $B(t)$ is equivalent to $X_n(F^{\mathbf{S}_n}(x))$ converging to $B(F_y(x))$. We therefore consider transforming $X_n(t)$ to $X_n(F^{\mathbf{S}_n}(x))$ where $F^{\mathbf{S}_n}(x)$ is the ESD of $\mathbf{S}_n$.

We define the eigenvector Empirical Spectral Distribution (VESD) $H^{\mathbf{S}_n}(x)$ of $\mathbf{S}_n$ as follows:

$$H^{\mathbf{S}_n}(x) = \sum_{i=1}^{n}|d_i|^2 \mathrm{I}(\lambda_i \le x). \tag{1.3}$$

Between $H^{\mathbf{S}_n}(x)$ in (1.3) and $F^{\mathbf{S}_n}(x)$ in (1.1), we notice that there is no difference except the coefficient associated with each indicator function such that

$$X_n\big(F^{\mathbf{S}_n}(x)\big) = \sqrt{n/2}\big(H^{\mathbf{S}_n}(x) - F^{\mathbf{S}_n}(x)\big). \tag{1.4}$$

Henceforth, the investigation of $X_n(t)$ is converted to that of the difference between two empirical distributions $H^{\mathbf{S}_n}(x)$ and $F^{\mathbf{S}_n}(x)$. The authors in [4] proved that $H^{\mathbf{S}_n}(x)$ and $F^{\mathbf{S}_n}(x)$ have the same limiting distribution, the Marčenko–Pastur distribution $F_y(x)$, where $y_n = n/N$ and $y = \lim_{n,N\to\infty} y_n \in (0, 1)$.

Before we present the main theorems, let us introduce the following notation:

$$\Delta^H = \|\mathrm{E}H^{\mathbf{S}_n} - F_{y_n}\| = \sup_x\big|\mathrm{E}H^{\mathbf{S}_n}(x) - F_{y_n}(x)\big|$$

and

$$\Delta_p^H = \|H^{\mathbf{S}_n} - F_{y_n}\| = \sup_x\big|H^{\mathbf{S}_n}(x) - F_{y_n}(x)\big|.$$

We denote $\xi_n = O_p(a_n)$ and $\eta_n = O_{\text{a.s.}}(b_n)$ if, for any $\epsilon > 0$, there exist a large positive constant c_1 and a positive random variable c_2, such that

$$P(\xi_n/a_n \ge c_1) \le \epsilon \quad \text{and} \quad P(\eta_n/b_n \le c_2) = 1,$$

respectively.

In this paper, we follow the work in [4] and establish three types of convergence rates of $H^{\mathbf{S}_n}(x)$ to $F_{y_n}(x)$ in the following theorems.

THEOREM 1.1. *Suppose that X_{ij}, $i = 1, \dots, n$, $j = 1, \dots, N$ are i.i.d. complex random variables with $\mathrm{E}X_{11} = 0$, $\mathrm{E}|X_{11}|^2 = 1$ and $\mathrm{E}|X_{11}|^{10} < \infty$. For any fixed unit vector $\mathbf{x}_n \in \mathbb{C}_1^n = \{\mathbf{x} \in \mathbb{C}^n : \|\mathbf{x}\| = 1\}$, and $y_n = n/N \le 1$, it then follows that*

$$\Delta^H = \|\mathrm{E}H^{\mathbf{S}_n} - F_{y_n}\| = \begin{cases} O(N^{-1/2}a^{-3/4}), & \text{if } N^{-1/2} \le a < 1, \\ O(N^{-1/8}), & \text{if } a < N^{-1/2}, \end{cases}$$

where $a = (1 - \sqrt{y_n})^2$ as it is defined in (1.2) and F_{y_n} denotes the Marčenko–Pastur distribution function with an index y_n.

REMARK 1.2. From the proof of Theorem 1.1, it is clear that the condition $\mathrm{E}|X_{11}|^{10} < \infty$ is required only in the truncation step in the next section. We therefore believe that the condition $\mathrm{E}|X_{11}|^{10} < \infty$ can be replaced by $\mathrm{E}|X_{11}|^8 < \infty$ in Theorems 1.6 and 1.8.

REMARK 1.3. Because the convergence rate of $\|\mathrm{E}H^{\mathbf{S}_n} - F_y\|$ depends on the convergence rate of $|y_n - y|$, we only consider the convergence rate of $\|\mathrm{E}H^{\mathbf{S}_n} - F_{y_n}\|$.

REMARK 1.4. As $a = (1 - \sqrt{y_n})^2$, we can characterize the closeness between y_n and 1 through a. In particular, when y_n is away from 1 (or $a \ge N^{-1/2}$), the convergence rate of $\|\mathrm{E}H^{\mathbf{S}_n} - F_{y_n}\|$ is $O(N^{-1/2})$, which we believe is the optimal convergence rate. This is because we observe in [4] that for an analytic function f,

$$Y_n(f) = \sqrt{n} \int f(x)\, d\big(H^{\mathbf{S}_n}(x) - F_{y_n}(x)\big) \tag{1.5}$$

converges to a Gaussian distribution. While in [6], Bai and Silverstein proved that the limiting distribution of

$$n \int f(x)\, d\big(F^{\mathbf{S}_n}(x) - F_{y_n}(x)\big)$$

is also a Gaussian distribution. We therefore conjecture that the optimal rate of $H^{\mathbf{S}_n}(x)$ should be $O(N^{-1/2})$ and $O(N^{-1})$ for $F^{\mathbf{S}_n}(x)$. Although $F^{\mathbf{S}_n}(x)$ and $H^{\mathbf{S}_n}(x)$ converge to the same limiting distribution, there exists a substantial difference between $F^{\mathbf{S}_n}(x)$ and $H^{\mathbf{S}_n}(x)$.

REMARK 1.5. Notice that two matrices $\mathbf{XX}^*$ and $\mathbf{X}^*\mathbf{X}$ share the same set of nonzero eigenvalues. However, these two matrices do not always share the same set of eigenvectors. Especially when $y_n \gg 1$, the eigenvectors of $\mathbf{S}_n$ corresponding to 0 eigenvalues can be arbitrary. As a result, the limit of $H^{\mathbf{S}_n}$ may not exist or heavily depends on the choice of unit vector $\mathbf{x}_n$. Therefore, we only consider the case of $y_n \leq 1$ in this paper and leave the case of $y_n \geq 1$ as a future research problem.

The rates of convergence in probability and almost sure convergence of the VESD are provided in the next two theorems.

THEOREM 1.6. *Under the assumptions in Theorem 1.1 except that we now only require* $\mathrm{E}|X_{11}|^8 < \infty$, *we have*

$$\Delta_p^H = \|H^{\mathbf{S}_n} - F_{y_n}\| = \begin{cases} O_p(N^{-1/4}a^{-1/2}), & \text{if } N^{-1/4} \leq a < 1, \\ O_p(N^{-1/8}), & \text{if } a < N^{-1/4}. \end{cases}$$

REMARK 1.7. As an application of Theorem 1.6, in [8] we extended the CLT of the linear spectral statistics $Y_n(f)$ established in [4] to the case where the kernel function f is continuously twice differentiable provided that the sample covariance matrix $\mathbf{S}_n$ satisfies the assumptions of Theorem 1.6. This result is useful in testing Johnstone's hypothesis when normality is not assumed.

THEOREM 1.8. *Under the assumptions in Theorem 1.6, for any* $\eta > 0$, *we have*

$$\Delta_p^H = \|H^{\mathbf{S}_n} - F_{y_n}\| = \begin{cases} O_{a.s.}(N^{-1/4+\eta}a^{-1/2}), & \text{if } N^{-1/4} \leq a < 1, \\ O_{a.s.}(N^{-1/8+\eta}), & \text{if } a < N^{-1/4}. \end{cases}$$

REMARK 1.9. In this paper, we will use the following notation:

- $\mathbf{X}^*$ denote the conjugate transpose of a matrix (or vector) $\mathbf{X}$;
- $\mathbf{X}^T$ denote the (ordinary) transpose of a matrix (or vector) $\mathbf{X}$;
- $\|\mathbf{x}\|$ denote the Euclidean norm for any vector $\mathbf{x}$;
- $\|\mathbf{A}\| = \sqrt{\lambda_{\max}(\mathbf{AA}^*)}$, the spectral norm;
- $\|F\| = \sup_x |F(x)|$ for any function F;
- $\bar{z}$ denote the conjugate of a complex number z.

The rest of the paper is organized as follows. In Section 2, we introduce the main tools used to prove Theorems 1.1, 1.6 and 1.8, including Stieltjes transform and a Berry–Esseen type inequality. The proofs of these three theorems are presented in Sections 3–6. Several important results which are repeatedly employed throughout Sections 3–6 are proved in Appendix A. Appendix B contains some existing results in the literature. Finally, preliminaries on truncation, centralization and rescaling are postponed to the last section.

2. Main tools.

2.1. *Stieltjes transform.* The Stieltjes transform is an essential tool in random matrix theory and our paper. Let us now briefly review the Stieltjes transform and some important and relevant results. For a cumulative distribution function $G(x)$, its Stieltjes transform $m_G(z)$ is defined as

$$m_G(z) = \int \frac{1}{\lambda - z} dG(\lambda), \qquad z \in \mathbb{C}^+ = \{z \in \mathbb{C}, \Im(z) > 0\},$$

where $\Im(\cdot)$ denotes the imaginary part of a complex number. The Stieltjes transforms of the ESD $F^{\mathbf{S}_n}(x)$ and the VESD $H^{\mathbf{S}_n}(x)$ are

$$m_{F^{\mathbf{S}_n}}(z) = \frac{1}{n} \operatorname{tr}(\mathbf{S}_n - z\mathbf{I}_n)^{-1}$$

and

$$m_{H^{\mathbf{S}_n}}(z) = \mathbf{x}_n^*(\mathbf{S}_n - z\mathbf{I}_n)^{-1}\mathbf{x}_n,$$

respectively. Here $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. For simplicity of notation, we use $m_n(z)$ and $m_n^H(z)$ to denote $m_{F^{\mathbf{S}_n}}(z)$ and $m_{H^{\mathbf{S}_n}}(z)$, respectively.

REMARK 2.1. Notice that although the eigenmatrix $\mathbf{U}_n$ may not be unique, the Stieltjes transform $m_n^H(z)$ of $H^{\mathbf{S}_n}$ depends on $\mathbf{S}_n$ for any $\mathbf{x}_n$ rather than $\mathbf{U}_n$.

Let $\underline{\mathbf{S}}_n = \mathbf{X}^*\mathbf{X}/N$ denote the companion matrix of $\mathbf{S}_n$. As $\mathbf{S}_n$ and $\underline{\mathbf{S}}_n$ share the same set of nonzero eigenvalues, it can be shown that Stieltjes transforms of $F^{\mathbf{S}_n}(x)$ and $F^{\underline{\mathbf{S}}_n}(x)$ satisfy the following equality:

$$\underline{m}_n(z) = -\frac{1 - y_n}{z} + y_n m_n(z), \tag{2.1}$$

where $\underline{m}_n(z)$ denotes the Stieltjes transform of $F^{\underline{\mathbf{S}}_n}(x)$. Moreover, [5] and [24] claimed that $F^{\underline{\mathbf{S}}_n}$ converges, almost surely, to a nonrandom distribution function $\underline{F}_y(x)$ with Stieltjes transform $\underline{m}(z)$ such that

$$\underline{m}(z) = -\frac{1 - y}{z} + y m_y(z), \tag{2.2}$$

where $m_y(z)$ denotes the Stieltjes transform of the Marčenko–Pastur distribution with index y. Using (6.1.4) in [7], we also obtain the relationship between two limits $m_y(z)$ and $\underline{m}(z)$ as follows:

$$m_y(z) = \frac{1}{-z(1 + \underline{m}(z))}. \tag{2.3}$$

2.2. *A Berry–Esseen type inequality.*

LEMMA 2.2. *Let $H^{\mathbf{S}_n}(x)$ and $F_{y_n}(x)$ be the VESD of $\mathbf{S}_n$ and the Marčenko–Pastur distribution with index y_n, respectively. Denote their corresponding Stieltjes transforms by $m_n^H(z)$ and $m_{y_n}(z)$, respectively. Then there exist large positive constants A, B, K_1, K_2 and K_3, such that for $A > B > 5$,*

$$\begin{aligned}\Delta^H &= \|\mathrm{E}H^{\mathbf{S}_n}(x) - F_{y_n}(x)\| \\ &\le K_1 \int_{-A}^{A} \left|\mathrm{E}m_n^H(z) - m_{y_n}(z)\right| du + K_2 v^{-1} \int_{|x|>B} \left|\mathrm{E}H^{\mathbf{S}_n}(x) - F_{y_n}(x)\right| dx \\ &\quad + K_3 v^{-1} \sup_x \int_{|t|<v} \left|F_{y_n}(x+t) - F_{y_n}(x)\right| dt,\end{aligned}$$

where $z = u + iv$ is a complex number with positive imaginary part (*i.e.*, $v > 0$).

REMARK 2.3. Lemma 2.2 can be proved using Lemma B.1. To prove Theorem 1.1, we apply Lemma 2.2. In addition, we prove Theorems 1.6 and 1.8 by replacing $\mathrm{E}H^{\mathbf{S}_n}(x)$, $\mathrm{E}m_n^H(z)$ with $H^{\mathbf{S}_n}(x)$ and $m_n^H(z)$, respectively.

3. Proof of Theorem 1.1. Under the condition of $\mathrm{E}|X_{11}|^{10} < \infty$, we can choose a sequence of η_N with $\eta_N \downarrow 0$ and $\eta_N N^{1/4} \uparrow \infty$ as $N \to \infty$, such that

$$\lim_{N\to\infty} \frac{1}{\eta_N^{10}} \mathrm{E}\big(|X_{11}|^{10}\mathrm{I}\big(|X_{11}| > \eta_N N^{1/4}\big)\big) = 0. \tag{3.1}$$

Furthermore, without loss of generality, we can assume that every $|X_{ij}|$ is bounded by $\eta_N N^{1/4}$ and has mean 0 and variance 1. See Appendix C for details on truncation, centralization and rescaling.

We introduce some notation before start proving Theorem 1.1. Throughout the paper, we use C and C_i for $i = 0, 1, 2, \ldots$ to denote positive constant numbers which are independent of N and may take different values at different appearances. Let $\mathbf{X}_j$ denote the jth column of the data matrix $\mathbf{X}$. Let $\mathbf{r}_j = \mathbf{X}_j/\sqrt{N}$ so that $\mathbf{S}_n = \sum_{j=1}^N \mathbf{r}_j\mathbf{r}_j^*$ and let

$$\begin{aligned}v_y &= \sqrt{a} + \sqrt{v} = 1 - \sqrt{y_n} + \sqrt{v}, \\ \mathbf{B}_j &= \mathbf{S}_n - \mathbf{r}_j\mathbf{r}_j^*, \\ \mathbf{A}(z) &= \mathbf{S}_n - z\mathbf{I}_n, \\ \mathbf{A}_j(z) &= \mathbf{B}_j - z\mathbf{I}_n, \\ \alpha_j(z) &= \mathbf{r}_j^*\mathbf{A}_j^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{r}_j - \frac{1}{N}\mathbf{x}_n^*\mathbf{A}_j^{-1}(z)\mathbf{x}_n, \\ \xi_j(z) &= \mathbf{r}_j^*\mathbf{A}_j^{-1}(z)\mathbf{r}_j - \frac{1}{N}\mathrm{E}\,\mathrm{tr}\,\mathbf{A}_j^{-1}(z),\end{aligned}$$

$$\hat{\xi}_j(z) = \mathbf{r}_j^*\mathbf{A}_j^{-1}(z)\mathbf{r}_j - \frac{1}{N}\operatorname{tr}\mathbf{A}_j^{-1}(z),$$
$$b(z) = \frac{1}{1+(1/N)\mathrm{E}\operatorname{tr}\mathbf{A}^{-1}(z)},$$
$$b_1(z) = \frac{1}{1+(1/N)\mathrm{E}\operatorname{tr}\mathbf{A}_1^{-1}(z)},$$
$$\beta_j(z) = \frac{1}{1+\mathbf{r}_j^*\mathbf{A}_j^{-1}(z)\mathbf{r}_j}.$$

It is easy to show that

$$\beta_j(z) - b_1(z) = -b_1(z)\beta_j(z)\xi_j(z). \tag{3.2}$$

For any $j = 1, 2, \dots, N$, we can also show that

$$\mathbf{r}_j^*\mathbf{A}^{-1}(z) = \beta_j(z)\mathbf{r}_j^*\mathbf{A}_j^{-1}(z) \tag{3.3}$$

due to the fact that

$$(\mathbf{B}_j - z\mathbf{I}_n)^{-1} - (\mathbf{B}_j + \mathbf{r}_j\mathbf{r}_j^* - z\mathbf{I}_n)^{-1} = (\mathbf{B}_j - z\mathbf{I}_n)^{-1}\mathbf{r}_j\mathbf{r}_j^*(\mathbf{B}_j + \mathbf{r}_j\mathbf{r}_j^* - z\mathbf{I}_n)^{-1}.$$

From (2.2) in [23], we can write $\underline{m}_n(z)$ in terms of $\beta_j(z)$ as follows:

$$\underline{m}_n(z) = -\frac{1}{zN}\sum_{j=1}^N \beta_j(z). \tag{3.4}$$

We proceed with the proof of Theorem 1.1:

$$\begin{aligned}
\delta &=: \mathrm{E}m_n^H(z) - m_y(z)\\
&= \mathbf{x}_n^*\big[\mathrm{E}\mathbf{A}^{-1}(z) - \big(-z\underline{m}(z)\mathbf{I}_n - z\mathbf{I}_n\big)^{-1}\big]\mathbf{x}_n\\
&= \big(z\underline{m}(z)+z\big)^{-1}\mathbf{x}_n^*\mathrm{E}\big[\big(z\underline{m}(z)+z\big)\mathbf{A}^{-1}(z) + \mathbf{I}_n\big]\mathbf{x}_n\\
&= \big(z\underline{m}(z)+z\big)^{-1}\mathbf{x}_n^*\mathrm{E}\big[\big(z\mathbf{I}_n + \mathbf{A}(z)\big)\mathbf{A}^{-1}(z) + z\underline{m}(z)\mathbf{A}^{-1}(z)\big]\mathbf{x}_n\\
&= \big(z\underline{m}(z)+z\big)^{-1}\mathbf{x}_n^*\mathrm{E}\Bigg[\sum_{j=1}^N \mathbf{r}_j\mathbf{r}_j^*\mathbf{A}^{-1}(z) + z\big(\mathrm{E}\underline{m}_n(z)\big)\mathbf{A}^{-1}(z)\\
&\qquad\qquad - z\big(\mathrm{E}\underline{m}_n(z)\big)\mathbf{A}^{-1}(z) + z\underline{m}(z)\mathbf{A}^{-1}(z)\Bigg]\mathbf{x}_n\\
&= \big(z\underline{m}(z)+z\big)^{-1}\mathbf{x}_n^*\mathrm{E}\Bigg[\sum_{j=1}^N \beta_j(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{A}_j^{-1}(z) - \big(-z\mathrm{E}\underline{m}_n(z)\big)\mathbf{A}^{-1}(z)\\
&\qquad\qquad - \big(z\mathrm{E}\underline{m}_n(z) - z\underline{m}(z)\big)\mathbf{A}^{-1}(z)\Bigg]\mathbf{x}_n
\end{aligned}$$

$$= \big(z\underline{m}(z)+z\big)^{-1}\mathbf{x}_n^*\mathrm{E}\left[\sum_{j=1}^N \beta_j(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{A}_j^{-1}(z) - \left(\frac{1}{N}\sum_{j=1}^N \mathrm{E}\beta_j(z)\right)\mathbf{A}^{-1}(z)\right]\mathbf{x}_n$$

$$- \big(z\underline{m}(z)+z\big)^{-1}\mathbf{x}_n^*\big(z\mathrm{E}\underline{m}_n(z) - z\underline{m}(z)\big)\big(\mathrm{E}\mathbf{A}^{-1}(z)\big)\mathbf{x}_n$$

$$= \big(z\underline{m}(z)+z\big)^{-1}\mathbf{x}_n^*\left[\sum_{j=1}^N \mathrm{E}\beta_j(z)\left(\mathbf{r}_j\mathbf{r}_j^*\mathbf{A}_j^{-1}(z) - \frac{1}{N}\mathrm{E}\mathbf{A}^{-1}(z)\right)\right]\mathbf{x}_n$$

$$+ m(z)\big(z\mathrm{E}\underline{m}_n(z) - z\underline{m}(z)\big)\mathrm{E}m_n^H(z)$$

$$=: \delta_1 + \delta_2,$$

where

$$\delta_1 = \big(z\underline{m}(z)+z\big)^{-1}\mathbf{x}_n^*\left[\sum_{j=1}^N \mathrm{E}\beta_j(z)\left(\mathbf{r}_j\mathbf{r}_j^*\mathbf{A}_j^{-1}(z) - \frac{1}{N}\mathrm{E}\mathbf{A}^{-1}(z)\right)\right]\mathbf{x}_n,$$

$$\delta_2 = m_y(z)\big(z\mathrm{E}\underline{m}_n(z) - z\underline{m}(z)\big)\mathrm{E}m_n^H(z).$$

LEMMA 3.1. *If*

$$|\delta_1| \le \frac{C_1}{Nvv_y}\left(\frac{1}{v_y} + \frac{\Delta^H}{v}\right)$$

holds for some constants C_0 and C_1, when $v^2 v_y \ge C_0 N^{-1}$, under the conditions of Theorem 1.1, there exists a constant C such that $\Delta^H \le Cv/v_y$.

PROOF. According to Lemma 2.2,

$$\begin{aligned}\Delta^H \le{}& K_1\int_{-A}^{A}\big|\mathrm{E}m_n^H(z) - m_y(z)\big|\,du\\ &+ K_2 v^{-1}\int_{|x|>B}\big|\mathrm{E}H^{\mathbf{S}_n}(x) - F_{y_n}(x)\big|\,dx\\ &+ K_3 v^{-1}\sup_x \int_{|t|<v}\big|F_{y_n}(x+t) - F_{y_n}(x)\big|\,dt.\end{aligned}$$

From Lemmas B.2 and A.8, we know that there exists a positive constant C, such that

$$\begin{aligned}&K_2 v^{-1}\int_{|x|>B}\big|\mathrm{E}H^{\mathbf{S}_n}(x) - F_{y_n}(x)\big|\,dx\\ &\qquad + K_3 v^{-1}\sup_x \int_{|t|<v}\big|F_{y_n}(x+t) - F_{y_n}(x)\big|\,dt \qquad (3.5)\\ &\quad\le Cv/v_y.\end{aligned}$$

Notice that

$$
\begin{aligned}
\int_{-A}^{A} & \left|\mathrm{E} m_n^H(z) - m_y(z)\right| du \\
& \le \int_{-A}^{A} |\delta_1|\, du + \int_{-A}^{A} |\delta_2|\, du \\
& \le \int_{-A}^{A} |\delta_1|\, du + \int_{-A}^{A} \left|z m_y(z)\right| \left|\mathrm{E} m_n^H(z)\right| \left|\mathrm{E} \underline{m}_n(z) - \underline{m}(z)\right| du \\
& \le \int_{-A}^{A} |\delta_1|\, du + \int_{-A}^{A} \left|z m_y(z)\right| \left|\mathrm{E} m_n^H(z) - m_y(z)\right| \left|\mathrm{E} \underline{m}_n(z) - \underline{m}(z)\right| du \\
& \quad + \int_{-A}^{A} \left|z m_y(z)\right| \left|m_y(z)\right| \left|\mathrm{E} \underline{m}_n(z) - \underline{m}(z)\right| du.
\end{aligned}
$$

Lemma A.1, (2.1) and (A.3) imply that

$$
\left|\mathrm{E} \underline{m}_n(z) - \underline{m}(z)\right| = |y_n| \left|\mathrm{E} m_n(z) - m_y(z)\right| \le \frac{C}{N v^{3/2} v_y^2}
$$

and

$$
\int_{-A}^{A} \left|\mathrm{E} \underline{m}_n(z) - \underline{m}(z)\right| du = \int_{-A}^{A} |y_n| \left|\mathrm{E} m_n(z) - m_y(z)\right| du \le C v.
$$

From (2.3) in [3], we have

$$
\begin{aligned}
z m_y(z) &= \frac{1 - y - z + \sqrt{(1 + y - z)^2 - 4y}}{2y} \\
&= -1 - \frac{1}{2\sqrt{y}} m_{\mathrm{semi}}\left(\frac{z - 1 - y}{\sqrt{y}}\right),
\end{aligned}
$$

where $m_{\mathrm{semi}}(\cdot)$ denotes the Stieltjes transform of the semicircle law, see (3.2) in [2]. Therefore $|z m_y(z)|$ is bounded by a constant, for $|m_{\mathrm{semi}}(\cdot)| \le 1$, see (3.3) in [2].

Combined with Lemma B.7, there exist constants C_2, C_3, such that

$$
\begin{aligned}
\int_{-A}^{A} & \left|\mathrm{E} m_n^H(z) - m_y(z)\right| du \\
& \le \int_{-A}^{A} |\delta_1|\, du + \frac{C_2}{N v^{3/2} v_y^2} \int_{-A}^{A} \left|\mathrm{E} m_n^H(z) - m_y(z)\right| du + \frac{C_3 v}{v_y}.
\end{aligned}
$$

Given $v^2 v_y \ge C_0 N^{-1}$, for $v_y \ge \sqrt{v}$, we have $v^{3/2} v_y^2 \ge v^2 v_y \ge C_0 N^{-1}$. For a large enough C_0 such that $C_2/C_0 \le 1/2$, we have

$$
\int_{-A}^{A} \left|\mathrm{E} m_n^H(z) - m_y(z)\right| du \le \int_{-A}^{A} |\delta_1|\, du + \frac{C v}{v_y}.
$$

As $|\delta_1| \le \frac{C_1}{Nvv_y}(\frac{1}{v_y} + \frac{\Delta^H}{v})$ and $v^2 v_y \ge C_0 N^{-1}$, if $C_0 \ge 4AC_1K_1$, we then have

$$\begin{aligned}\int_{-A}^{A} \left|\mathrm{E}m_n^H(z) - m_y(z)\right| du &\le \frac{2AC_1}{Nv^2v_y}\frac{v}{v_y} + \frac{2AC_1}{Nv^2v_y}\Delta^H + \frac{Cv}{v_y} \\ &\le \frac{\Delta^H}{2K_1} + \frac{Cv}{v_y}.\end{aligned} \tag{3.6}$$

Thus, from Lemma 2.2, equations (3.6) and (3.5), we conclude that there exists a constant C, such that

$$\Delta^H \le \frac{Cv}{v_y}.$$

The proof is complete. $\square$

To finish the proof of Theorem 1.1, we choose $v = \frac{\sqrt{C_0 N^{-1}}}{\sqrt{\sqrt{a}+N^{-1/4}}}$ such that $v^2 v_y = C_0 N^{-1} \frac{v_y}{\sqrt{a}+N^{-1/4}} \ge C_0 N^{-1}$. According to Lemma 3.1, we know that

$$\Delta^H \le \frac{Cv}{v_y} \le CN^{-1/2}\left(\sqrt{a} + N^{-1/4}\right)^{-3/2}.$$

If $\sqrt{a} < N^{-1/4}$, $\Delta^H \le CN^{-1/2}(N^{-1/4})^{-3/2} = O(N^{-1/8})$.
If $\sqrt{a} \ge N^{-1/4}$, $\Delta^H \le CN^{-1/2}(\sqrt{a})^{-3/2} = O(N^{-1/2}a^{-3/4})$.

Thus, the proof of Theorem 1.1 is complete.

4. The bound for $|\delta_1|$. In this section, we are going to show that when $v^2 v_y \ge C_0 N^{-1}$, $|\delta_1|$ is indeed bounded by $\frac{C_1}{Nvv_y}(\frac{1}{v_y} + \frac{\Delta^H}{v})$, as required by Lemma 3.1.

From δ_1, we can further write $\delta_1 = \delta_{11} + \delta_{12} + \delta_{13}$, where

$$\begin{aligned}\delta_{11} &= N\left(z\underline{m}(z) + z\right)^{-1}\mathrm{E}\left[\beta_1(z)\left(\mathbf{r}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{r}_1 - \frac{1}{N}\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\right)\right] \\ &= N\left(z\underline{m}(z) + z\right)^{-1}\mathrm{E}\left(\beta_1(z)\alpha_1(z)\right), \\ \delta_{12} &= \left(z\underline{m}(z) + z\right)^{-1}\mathrm{E}\left[\beta_1(z)\mathbf{x}_n^*(\mathbf{A}_1^{-1}(z) - \mathbf{A}^{-1}(z))\mathbf{x}_n\right], \\ \delta_{13} &= \left(z\underline{m}(z) + z\right)^{-1}\mathrm{E}\left[\beta_1(z)\mathbf{x}_n^*(\mathbf{A}^{-1}(z) - \mathrm{E}\mathbf{A}^{-1}(z))\mathbf{x}_n\right].\end{aligned}$$

According to (2.3) and Lemma B.7,

$$\left|\left(z\underline{m}(z) + z\right)^{-1}\right| = \left|-m_y(z)\right| \le \frac{C}{v_y} \tag{4.1}$$

for some constant C. Using identity (3.2) three times, we have

$$\beta_1(z) = b_1(z) - b_1^2(z)\xi_1(z) + b_1^3(z)\xi_1^2(z) - b_1^3(z)\beta_1(z)\xi_1^3(z).$$

Notice that $\mathrm{E}\alpha_1(z) = 0$ and $b_1(z)$ is bounded by a constant (due to Lemma A.3), we then have

$$
\begin{aligned}
|\delta_{11}| &\le \frac{CN}{v_y}\big|\mathrm{E}\beta_1(z)\alpha_1(z)\big| \\
&\le \frac{CN}{v_y}\big(\big|\mathrm{E}\xi_1(z)\alpha_1(z)\big| + \big|\mathrm{E}\xi_1^2(z)\alpha_1(z)\big| + \big|\mathrm{E}\beta_1(z)\xi_1^3(z)\alpha_1(z)\big|\big).
\end{aligned} \tag{4.2}
$$

Let us start with the first term in the above upper bound of $|\delta_{11}|$ as in (4.2). Note that $\mathbf{r}_1$ and $\mathbf{A}_1^{-1}(z)$ are independent. Therefore, for any integer $p > 0$ we have

$$
\mathrm{E}\big(\operatorname{tr}\mathbf{A}_1^{-1}(z) - \mathrm{E}\operatorname{tr}\mathbf{A}_1^{-1}(z)\big)\alpha_1(z) = 0
$$

and

$$
\mathrm{E}\big(\operatorname{tr}\mathbf{A}_1^{-1}(z)\big)^p\alpha_1(z) = 0.
$$

Denote $\mathbf{A}_1^{-1}(z) = (a_{ij})_{n\times n}$, $\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^* = (b_{ij})_{n\times n}$, and $\mathbf{e}_i$ be the ith canonical basis vector, that is, the n-vector whose coordinates are all 0 except that the ith coordinate is 1. Then Lemmas B.3, A.5, A.6 and the inequality $\|\mathbf{A}_1^{-1}(z)\| \le 1/v$ imply that

$$
\begin{aligned}
&\big|\mathrm{E}\xi_1(z)\alpha_1(z)\big| \\
&\quad = \big|\mathrm{E}\hat{\xi}_1(z)\alpha_1(z)\big| \\
&\quad \le \frac{C}{N^2}\mathrm{E}\left(\operatorname{tr}\big(\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar{z})\mathbf{x}_n\mathbf{x}_n^*\big) + \sum_{i=1}^n a_{ii}b_{ii}\right) \\
&\quad \le \frac{C}{N^2 v}\left\{\frac{1}{v_y} + \frac{\Delta^H}{v} + v\left(\sum_{i=1}^N \mathrm{E}\big|\mathbf{x}_n^*\mathbf{A}_1^{-1}(\bar{z})\mathbf{e}_i\big|^2 \sum_{i=1}^N \mathrm{E}|a_{ii}|^2\big|\mathbf{x}_n^*\mathbf{e}_i\big|^2\right)^{1/2}\right\} \\
&\quad \le \frac{C}{N^2 v}\left(\frac{1}{v_y} + \frac{\Delta^H}{v}\right).
\end{aligned}
$$

In the above, we use the following two results, which can be proved by applying Lemmas A.5 and A.6:

$$
\mathrm{E}|a_{11}| = \mathrm{E}\big|\mathbf{e}_1^*\mathbf{A}_1^{-1}(z)\mathbf{e}_1\big| \le C\left(\frac{1}{v_y} + \frac{\Delta^H}{v}\right),
$$

$$
\mathrm{E}|a_{11}|^2 \le \mathrm{E}\big|\mathbf{e}_1^*\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar{z})\mathbf{e}_1\big| \le \frac{C}{v}\left(\frac{1}{v_y} + \frac{\Delta^H}{v}\right).
$$

Hence, we have shown that

$$
\big|\mathrm{E}\xi_1(z)\alpha_1(z)\big| \le \frac{C}{N^2 v}\left(\frac{1}{v_y} + \frac{\Delta^H}{v}\right). \tag{4.3}
$$

Let us denote $\mathbf{X}_1^*$ the conjugate transpose of $\mathbf{X}_1$, that is, $\mathbf{X}_1^* = (\bar{X}_{11}, \bar{X}_{21}, \dots, \bar{X}_{n1})$. Then we can rewrite the second term in the upper bound of $|\delta_{11}|$ as

$$\begin{aligned}
\mathrm{E}\xi_1^2(z)\alpha_1(z) &= \frac{1}{N^3}\mathrm{E}\biggl[\biggl(\sum_{i\neq j} a_{ij}\bar{X}_{i1}X_{j1} + \sum_i \bigl(a_{ii}|X_{1i}^2| - \mathrm{E}a_{ii}\bigr)\biggr)^2 \\
&\qquad\qquad \times \biggl(\sum_{i,j} b_{ij}(\bar{X}_{i1}X_{j1} - \delta_{ij})\biggr)\biggr] \\
&= \frac{1}{N^3}\mathrm{E}\biggl[\biggl\{\biggl(\sum_{i\neq j} a_{ij}\bar{X}_{i1}X_{j1}\biggr)^2 + 2\biggl(\sum_i \bigl(a_{ii}|X_{1i}^2| - \mathrm{E}a_{ii}\bigr)\biggr) \\
&\qquad \times \biggl(\sum_{i\neq j} a_{ij}\bar{X}_{i1}X_{j1}\biggr) + \biggl(\sum_i a_{ii}|X_{1i}^2| - \mathrm{E}a_{ii}\biggr)^2\biggr\} \\
&\qquad \times \biggl\{\sum_{i\neq j} b_{ij}\bar{X}_{i1}X_{j1} + \sum_i b_{ii}\bigl(|X_{1i}^2| - 1\bigr)\biggr\}\biggr] \\
&\le \frac{C}{N^3}\mathrm{E}\biggl(\sum_i |a_{ii}^2 b_{ii}| + \sum_i |\mathrm{E}a_{ii}|^2|b_{ii}| + \sum_{i\neq j}|a_{ij}^2 b_{ii}| \\
&\qquad + \sum_{i\neq j}|a_{ij}^2 b_{ij}| + \sum_{i\neq j}|a_{ij}a_{ii}b_{jj}| + \sum_{i\neq j}|a_{ii}a_{ij}b_{ij}|\biggr) \\
&\quad + \frac{C}{N^3}\sum_{\iota,\tau}\biggl|\sum_{i\neq j\neq k}\mathrm{E}a_{ij}^{\iota}a_{jk}^{\tau}b_{ik}\biggr|,
\end{aligned}$$

where a_{ij}^{ι} and a_{ij}^{τ} denote a_{ij} or $\bar{a}_{ij}$. By following the similar proofs in establishing (4.3), we are able to show that

$$\begin{aligned}
\mathrm{E}\sum_i |a_{ii}^2 b_{ii}| &\le \frac{1}{v}\sum_i \mathrm{E}|a_{ii}b_{ii}| \\
&\le \frac{1}{v}\bigl(\mathrm{E}|a_{11}^2|\mathrm{E}\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar{z})\mathbf{x}_n\bigr)^{1/2} \\
&\le \frac{C}{v^2}\biggl(\frac{1}{v_y} + \frac{\Delta^H}{v}\biggr), \\
\mathrm{E}\sum_i |\mathrm{E}a_{ii}|^2|b_{ii}| &\le |\mathrm{E}a_{11}|^2\bigl(\mathrm{E}\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar{z})\mathbf{x}_n\bigr)^{1/2} \\
&\le \biggl[\frac{C}{v}\biggl(\frac{1}{v_y} + \frac{\Delta^H}{v}\biggr)\biggr]^{3/2}.
\end{aligned}$$

The Cauchy–Schwarz inequality implies that

$$
\begin{aligned}
\mathrm{E}\sum_{i\neq j}|a_{ij}^2 b_{ii}| &\le \mathrm{E}\Big(\sum_i |b_{ii}|\Big(\sum_j |a_{ij}^2|\Big)\Big)\\
&\le \Big(\sum_i \mathrm{E}\big|\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{e}_i\big|^2\Big)^{1/2}\Big(\sum_i |\mathbf{x}_n^*\mathbf{e}_i|^2\mathrm{E}\Big(\sum_j |a_{ij}^2|\Big)^2\Big)^{1/2}\\
&= \big(\mathrm{E}\mathbf{x}_n^*(\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar z))\mathbf{x}_n\big)^{1/2}\big(\mathrm{E}\big|\big(\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar z)_{11}\big)^2\big|\big)^{1/2}\\
&\le \frac{C}{v^{3/2}}\Big(\frac{1}{v_y}+\frac{\Delta^H}{v}\Big)^{3/2},\\
\mathrm{E}\sum_{i\neq j} a_{ij}^2 b_{ij} &\le \Big[\sum_{ij}\mathrm{E}\big|a_{ij}\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{e}_i\big|^2\sum_{ij}\mathrm{E}\big|a_{ij}\mathbf{x}_n^*\mathbf{e}_j\big|^2\Big]^{1/2}\\
&\le \Big[\sum_i \mathrm{E}\big|\big(\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar z)\big)_{ii}\big(\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{e}_i\big)^2\big|\\
&\qquad\times\sum_j \mathrm{E}\big|\big(\mathbf{A}_1^{-1}(\bar z)\mathbf{A}_1^{-1}(z)\big)_{jj}\big(\mathbf{x}_n^*\mathbf{e}_j\big)^2\big|\Big]^{1/2}\\
&\le \frac{C}{v^2}\Big(\frac{1}{v_y}+\frac{\Delta^H}{v}\Big)^{3/2},\\
\mathrm{E}\sum_{i\neq j}|a_{ii}a_{ij}b_{jj}| &\le \Big[\sum_{ij}\mathrm{E}\big|a_{ii}\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{e}_j\big|^2\sum_{ij}\mathrm{E}\big|a_{ij}\mathbf{x}_n^*\mathbf{e}_j\big|^2\Big]^{1/2}\\
&\le \Big[\sum_i \mathrm{E}\big|a_{ii}^2\big|\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar z)\mathbf{x}_n\\
&\qquad\times\sum_j \mathrm{E}\big(\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar z)\big)_{jj}\big|\mathbf{x}_n^*\mathbf{e}_j\big|^2\Big]^{1/2}\\
&\le C\frac{\sqrt{N}}{v^2}\Big(\frac{1}{v_y}+\frac{\Delta^H}{v}\Big),\\
\mathrm{E}\sum_{i\neq j}|a_{ii}a_{ij}b_{ij}| &\le \Big[\sum_{ij}\mathrm{E}\big|a_{ii}\mathbf{x}_n^*\mathbf{A}_1^{-1}\mathbf{e}_j\big|^2\sum_{ij}\mathrm{E}\big|a_{ij}\mathbf{x}_n^*\mathbf{e}_i\big|^2\Big]^{1/2}\\
&\le C\frac{\sqrt{N}}{v^2}\Big(\frac{1}{v_y}+\frac{\Delta^H}{v}\Big).
\end{aligned}
$$

Finally, we establish

$$
\sum_{\iota,\tau}\Big|\sum_{i\neq j\neq k}\mathrm{E}a_{ij}^{\iota}a_{jk}^{\tau}b_{ik}\Big| \le C\frac{\sqrt{N}}{v^2}\Big(\frac{1}{v_y}+\frac{\Delta^H}{v}\Big).
$$

By inclusive–exclusive principle and what we have just proved, it remains to show that

$$\sum_{\iota,\tau}\left|\sum_{i,j,k}\mathrm{E}a_{ij}^{\iota}a_{jk}^{\tau}b_{ik}\right| \le C\frac{\sqrt{N}}{v^2}\left(\frac{1}{v_y}+\frac{\Delta^H}{v}\right). \tag{4.4}$$

Notice that $\gamma_{ik}=\sum_j a_{ij}a_{jk}$ is the (i,k)-element of $\mathbf{A}_1^{-2}(z)$. We obtain

$$\begin{aligned}\left|\sum_{i,j,k}\mathrm{E}a_{ij}a_{jk}b_{ik}\right| &= \left|\sum_{i,k}\gamma_{ik}b_{ik}\right|\\ &\le \left(\sum_{i,k}\mathrm{E}|\gamma_{ik}|^2\sum_{i,k}\mathrm{E}|b_{ik}|^2\right)^{1/2}\\ &= \left(\mathrm{E}\operatorname{tr}\left(\mathbf{A}_1^{-2}(z)\mathbf{A}_1^{-2}(\bar{z})\right)\mathrm{E}\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar{z})\mathbf{x}_n\right)^{1/2}\\ &\le C\frac{\sqrt{N}}{v^2}\left(\frac{1}{v_y}+\frac{\Delta^H}{v}\right).\end{aligned}$$

Similarly, one can prove the other terms of (4.4) share this common bound.

In summary, when $v^2v_y \ge C_0N^{-1}$ it holds that

$$\left|\mathrm{E}\xi_1^2(z)\alpha_1(z)\right| \le \frac{C}{N^2v}\left(\frac{1}{v_y}+\frac{\Delta^H}{v}\right). \tag{4.5}$$

For the last term in the upper bound of $|\delta_{11}|$, we apply Lemma A.4 and the Cauchy–Schwarz inequality again. In particular, for any fixed $t>0$, we have that

$$\begin{aligned}\left|\mathrm{E}\beta_1(z)\xi_1^3(z)\alpha_1(z)\right| &\le C\mathrm{E}\left|\xi_1^3(z)\alpha_1(z)\right| + o\left(N^{-t}\right)\\ &\le C\left(\mathrm{E}\left|\xi_1(z)\right|^6\right)^{1/2}\left(\mathrm{E}\left|\alpha_1(z)\right|^2\right)^{1/2} + o\left(N^{-t}\right)\\ &\le \frac{C}{N^{5/2}v^2v_y^{3/2}}\left(\frac{1}{v_y}+\frac{\Delta^H}{v}\right)^{1/2}.\end{aligned} \tag{4.6}$$

The last inequality in (4.6) is due to Lemmas A.2 and A.7. Therefore, for any $v^2v_y \ge C_0N^{-1}$, (4.3), (4.5) and (4.6) lead us to

$$|\delta_{11}| \le \frac{C}{Nvv_y}\left(\frac{1}{v_y}+\frac{\Delta^H}{v}\right). \tag{4.7}$$

To establish the upper bound for $|\delta_{12}|$, we will make use of the following equality:

$$\mathbf{A}_1^{-1}(z)-\mathbf{A}^{-1}(z) = \beta_1(z)\mathbf{A}_1^{-1}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{A}_1^{-1}(z). \tag{4.8}$$

Note that (4.1) implies that

$$
\begin{aligned}
|\delta_{12}| &\le \frac{C}{v_y}\big|\mathrm{E}\beta_1(z)\mathbf{x}_n^*\big(\mathbf{A}_1^{-1}(z)-\mathbf{A}^{-1}(z)\big)\mathbf{x}_n\big| \\
&\le \frac{C}{v_y}\big|\mathrm{E}\beta_1^2(z)\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\big| \\
&\le \frac{C}{Nv_y}\mathrm{E}\big|\mathbf{X}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{X}_1\big| + o\big(N^{-t}\big) \qquad \text{(see Lemma A.4)} \\
&\le \frac{C}{Nv_y}\big|\mathrm{E}\,\mathrm{tr}\big(\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\big)\big| \qquad \text{(see Lemma B.5)} \\
&= \frac{C}{Nv_y}\big|\mathrm{E}\mathbf{x}_n^*\mathbf{A}_1^{-2}(z)\mathbf{x}_n\big| \\
&\le \frac{C}{Nvv_y}\left(\frac{1}{v_y}+\frac{\Delta^H}{v}\right).
\end{aligned} \tag{4.9}
$$

At last, we establish the upper bound for δ_{13}.

By (4.1), Lemma A.3, and the fact

$$\beta_1(z) = b_1(z) - b_1(z)\beta_1(z)\xi_1(z), \tag{4.10}$$

we obtain

$$
\begin{aligned}
|\delta_{13}| &\le \frac{C}{v_y}\big|\mathrm{E}\big\{\beta_1(z)\mathbf{x}_n^*\big(\mathbf{A}^{-1}(z)-\mathrm{E}\mathbf{A}^{-1}(z)\big)\mathbf{x}_n\big\}\big| \\
&= \frac{C}{v_y}\big|\mathrm{E}\big\{b(z)\beta_1(z)\xi_1(z)\mathbf{x}_n^*\big(\mathbf{A}^{-1}(z)-\mathrm{E}\mathbf{A}^{-1}(z)\big)\mathbf{x}_n\big\}\big| \\
&\le \frac{C}{v_y}\big|\mathrm{E}\big\{\xi_1(z)\mathbf{x}_n^*\big(\mathbf{A}^{-1}(z)-\mathrm{E}\mathbf{A}^{-1}(z)\big)\mathbf{x}_n\big\}\big| \\
&\quad + \frac{C}{v_y}\big|\mathrm{E}\big\{\beta_1(z)\xi_1^2(z)\mathbf{x}_n^*\big(\mathbf{A}^{-1}(z)-\mathrm{E}\mathbf{A}^{-1}(z)\big)\mathbf{x}_n\big\}\big|.
\end{aligned}
$$

From C_r inequality (see Loève [14]), we have $|\delta_{13}| \le \frac{C}{v_y}(|\mathrm{II}_1|+|\mathrm{II}_2|+|\mathrm{II}_3|)$, where

$$
\begin{aligned}
\mathrm{II}_1 &= \mathrm{E}\big\{\xi_1(z)\mathbf{x}_n^*\big(\mathbf{A}_1^{-1}(z)-\mathrm{E}\mathbf{A}_1^{-1}(z)\big)\mathbf{x}_n\big\}, \\
\mathrm{II}_2 &= \mathrm{E}\big\{\xi_1(z)\mathbf{x}_n^*\big(\mathbf{A}^{-1}(z)-\mathbf{A}_1^{-1}(z)\big)\mathbf{x}_n\big\}, \\
\mathrm{II}_3 &= \mathrm{E}\big\{\beta_1(z)\xi_1(z)\mathbf{x}_n^*\mathrm{E}\big(\mathbf{A}^{-1}(z)-\mathbf{A}_1^{-1}(z)\big)\mathbf{x}_n\big\}.
\end{aligned}
$$

It should be noted that $\mathrm{E}(\xi_1(z) \mid \mathbf{A}_1^{-1}(z)) = 0$, and $\mathbf{r}_1$ and $\mathbf{A}_1^{-1}(z)$ are independent. Then we have

$$\mathrm{II}_1 = \mathrm{E}\big[\mathrm{E}\big\{\xi_1(z)\mathbf{x}_n^*\big(\mathbf{A}_1^{-1}(z)-\mathrm{E}\mathbf{A}_1^{-1}(z)\big)\mathbf{x}_n \mid \mathbf{A}_1^{-1}(z)\big\}\big] = 0.$$

By the results in (4.8) and (4.10), we have

$$\begin{aligned}|\mathrm{II}_2| &= \big|\mathrm{E}\xi_1(z)\beta_1(z)\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\big| \\ &\le \big|b_1(z)\big|\big|\mathrm{E}\xi_1(z)\mathbf{r}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{r}_1\big| \\ &\quad + \big|b_1(z)\big|\big|\mathrm{E}\beta_1(z)\xi_1^2(z)\mathbf{r}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{r}_1\big| \\ &=: \mathrm{III}_1 + \mathrm{III}_2,\end{aligned}$$

where

$$\begin{aligned}\mathrm{III}_1 &= \big|b_1(z)\big|\bigg|\mathrm{E}\xi_1(z)\bigg(\mathbf{r}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{r}_1 - \frac{1}{N}\mathbf{x}_n^*\mathbf{A}_1^{-2}(z)\mathbf{x}_n\bigg)\bigg| \\ &\le \frac{C}{N^2}\bigg|\mathrm{E}\,\mathrm{tr}\big(\mathbf{x}_n^*\mathbf{A}_1^{-3}(z)\mathbf{x}_n\big) + \mathrm{E}\sum_i\big(\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\big)_{ii}a_{ii}\bigg| \\ &= \frac{C}{N^2v^2}\bigg(\frac{1}{v_y} + \frac{\Delta^H}{v}\bigg).\end{aligned}$$

The above inequality follows from Lemmas B.3 and A.3. By Lemma A.4 and the Cauchy–Schwarz inequality, it holds that

$$\begin{aligned}\mathrm{III}_2 &\le C\mathrm{E}\big|\xi_1^2(z)\mathbf{r}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{r}_1\big| \\ &\le C\big(\mathrm{E}\big|\xi_1(z)\big|^4\big)^{1/2}\big(\mathrm{E}\big|\mathbf{r}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{r}_1\big|^2\big)^{1/2} \\ &\le C\bigg(\frac{1}{N^2v^2v_y^2}\frac{1}{N^2}\mathrm{E}\big|\mathbf{x}_n^*A_1^{-2}(z)\mathbf{x}_n\big|^2\bigg)^{1/2} \qquad \text{(see Lemmas A.2 and B.5)} \\ &\le \frac{C}{N^2v^2v_y}\bigg(\frac{1}{v_y} + \frac{\Delta^H}{v}\bigg).\end{aligned}$$

Hence, we have shown that

$$|\mathrm{II}_2| \le \frac{C}{N^2v^2v_y}\bigg(\frac{1}{v_y} + \frac{\Delta^H}{v}\bigg).$$

Moreover, Lemmas A.2 and A.5, and the Cauchy–Schwarz inequality lead us to the following:

$$\begin{aligned}|\mathrm{II}_3| &\le C\big(\mathrm{E}\big|\xi_1(z)\big|^2\mathrm{E}\big|\mathbf{x}_n^*\big(\mathbf{A}^{-1}(z) - \mathbf{A}_1^{-1}(z)\big)\mathbf{x}_n\big|^2\big)^{1/2} \\ &\le \frac{C}{N^{1/2}v^{1/2}v_y^{1/2}}\frac{C}{Nv}\bigg(\frac{1}{v_y} + \frac{\Delta^H}{v}\bigg) \\ &= \frac{C}{N^{3/2}v^{3/2}v_y^{1/2}}\bigg(\frac{1}{v_y} + \frac{\Delta^H}{v}\bigg).\end{aligned}$$

Therefore, it follows that

$$|\delta_{13}| \le \frac{C}{v_y}\big(|\mathrm{II}_1| + |\mathrm{II}_2| + |\mathrm{II}_3|\big) \le \frac{C}{N^{3/2}v^{3/2}v_y}\left(\frac{1}{v_y} + \frac{\Delta^H}{v}\right). \tag{4.11}$$

As it has been shown in (4.7), (4.9) and (4.11), we conclude that

$$|\delta_1| \le \frac{C}{Nvv_y}\left(\frac{1}{v_y} + \frac{\Delta^H}{v}\right). \tag{4.12}$$

5. Proof of Theorem 1.6. From Lemma 2.2, and by replacing $\mathrm{E}H^{\mathbf{S}_n}(x)$ and $\mathrm{E}m_n^H(z)$ by $H^{\mathbf{S}_n}(x)$ and $m_n^H(z)$, respectively, we have

$$\begin{aligned}
\mathrm{E}\Delta_p^H &=: \mathrm{E}\|H^{\mathbf{S}_n}(x) - F_{y_n}(x)\| \\
&\le K_1 \int_{-A}^{A} \mathrm{E}\big|m_n^H(z) - \mathrm{E}m_n^H(z)\big|\,du + K_1 \int_{-A}^{A} \big|\mathrm{E}m_n^H(z) - m_y(z)\big|\,du \\
&\quad + K_2 v^{-1} \int_{|x|>B} \big|\mathrm{E}H^{\mathbf{S}_n}(x) - F_{y_n}(x)\big|\,dx \\
&\quad + K_3 v^{-1} \sup_x \int_{|t|<v} \big|F_{y_n}(x+t) - F_{y_n}(x)\big|\,dt \\
&\le K_1 \int_{-A}^{A} \mathrm{E}\big|m_n^H(z) - \mathrm{E}m_n^H(z)\big|\,du + \Delta^H.
\end{aligned}$$

As the convergence rate of Δ^H has already been established in Theorem 1.1, we only focus on the convergence rate of $\mathrm{E}|m_n^H(z) - \mathrm{E}m_n^H(z)|$.

By Lemma A.6 and the Cauchy–Schwarz inequality, it follows that

$$\begin{aligned}
\mathrm{E}\big|m_n^H(z) - \mathrm{E}m_n^H(z)\big| &\le \big(E\big|m_n^H(z) - \mathrm{E}m_n^H(z)\big|^2\big)^{1/2} \\
&\le \frac{C}{\sqrt{N}v}\left(\frac{1}{v_y} + \frac{\Delta^H}{v}\right).
\end{aligned}$$

Together with Lemma 3.1 that $\Delta^H \le Cv/v_y$, when $v^2 v_y \ge O(N^{-1})$, we have

$$\mathrm{E}\|H^{\mathbf{S}_n}(x) - F_{y_n}(x)\| \le \left(\frac{1}{\sqrt{N}v} + v\right)\frac{C}{v_y}.$$

By choosing $v = O(N^{-1/4})$, we obtain

$$\mathrm{E}\|H^{\mathbf{S}_n}(x) - F_{y_n}(x)\| \le \begin{cases} O\big(N^{-1/4}a^{-1/2}\big), & \text{when } a \ge N^{-1/4}, \\ O\big(N^{-1/8}\big), & \text{otherwise.} \end{cases}$$

The proof of Theorem 1.6 is complete.

6. Proof of Theorem 1.8. Notice that the proof of Theorem 1.8 is almost the same as that of Theorem 1.6.

By Lemma 2.2, choosing $v = O(N^{-1/4})$,

$$\|H^{\mathbf{S}_n} - F_{y_n}\| \le \int_{-A}^{A} |m_n^H(z) - \mathrm{E}m_n^H(z)|\, du + Cv/v_y.$$

By Lemma A.6, we have

$$\mathrm{E}|m_n^H(z) - \mathrm{E}m_n^H(z)|^{2l} \le CN^{-l}v^{-2l}\left(\frac{1}{v_y} + \frac{\Delta^H}{v}\right)^{2l}.$$

When $a < N^{-1/4}$, with $v = O(N^{-1/4})$,

$$N^{2l(1/8-\eta)}\mathrm{E}|m_n^H(z) - \mathrm{E}m_n^H(z)|^{2l} \le CN^{-2l\eta},$$

which implies that if we choose an l such that $2l\eta > 1$,

$$\int_{-A}^{A} |m_n^H(z) - \mathrm{E}m_n^H(z)|\, du = o_{\mathrm{a.s.}}(N^{-1/8+\eta}).$$

When $a \ge N^{-1/4}$, in this case, by choosing $v = O(N^{-1/4})$, we have

$$a^l N^{2l(1/4-\eta)}\mathrm{E}|m_n^H(z) - \mathrm{E}m_n^H(z)|^{2l} \le CN^{-2l\eta}.$$

Theorem 1.8 then follows by setting $l > \frac{1}{2\eta}$. This completes the proof of Theorem 1.8.

APPENDIX A

In this section, we establish some lemmas which are used in the proofs of the main theorems.

LEMMA A.1. *Under the conditions of Theorem* 1.6, *for all* $|z| < A$ *and* $v^2 v_y \ge C_0 N^{-1}$, *we have*

$$|\mathrm{E}m_n(z) - m_y(z)| \le \frac{C}{Nv^{3/2}v_y^2},$$

where C_0 *is a constant and* $v_y = 1 - \sqrt{y_n} + \sqrt{v}$.

PROOF. Since

$$\begin{aligned}
\mathrm{E}m_n(z) &= \frac{1}{n}\mathrm{E}\,\mathrm{tr}(\mathbf{S}_n - z\mathbf{I}_n)^{-1} \\
&= \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}\frac{1}{s_{kk} - z - N^{-2}\boldsymbol{\alpha}_k^*(\mathbf{S}_{nk} - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k} \\
&= \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}\frac{1}{\epsilon + 1 - y_n - z - y_n z\mathrm{E}m_n(z)} \qquad \text{(A.1)}\\
&= -\frac{1}{z + y_n - 1 + y_n z\mathrm{E}m_n(z)} + \delta_n,
\end{aligned}$$

where

$$
\begin{aligned}
s_{kk} &= \frac{1}{N}\sum_{j=1}^{N}|X_{kj}|^2, \\
\mathbf{S}_{nk} &= \frac{1}{N}\mathbf{X}_{(k)}\mathbf{X}_{(k)}^*, \\
\boldsymbol{\alpha}_k &= \mathbf{X}_{(k)}\bar{\mathbf{X}}_k, \\
\epsilon_k &= (s_{kk}-1) + y_n + y_n z\mathrm{E}m_n(z) - \frac{1}{N^2}\boldsymbol{\alpha}_k^*(\mathbf{S}_{nk} - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k, \\
\delta_n &= -\frac{1}{n}\sum_{k=1}^{n} b_n \mathrm{E}\beta_k\epsilon_k, \\
b_n &= b_n(z) = \frac{1}{z + y_n - 1 + y_n z\mathrm{E}m_n(z)}, \\
\beta_k &= \beta_k(z) = \frac{1}{z + y_n - 1 + y_n z\mathrm{E}m_n(z) - \epsilon_k},
\end{aligned}
$$

where $\mathbf{X}_{(k)}$ is the $(n-1)\times N$ matrix obtained from $\mathbf{X}$ with its kth row removed and $\mathbf{X}_k^*$ is the kth row of $\mathbf{X}$. It has proved that one of the roots of equation (A.1) is (see (3.1.7) in [3])

$$
\mathrm{E}m_n(z) = -\frac{1}{2y_n z}\left(z + y_n - 1 - y_n z\delta_n - \sqrt{(z + y_n - 1 + y_n z\delta_n)^2 - 4y_n z}\right).
$$

The Stieltjes transform of the Marčenko–Pastur distribution with index y is given by (see (2.3) in [3])

$$
m_y(z) = -\frac{y_n + z - 1 - \sqrt{(1 + y_n - z)^2 - 4y_n}}{2y_n z}.
$$

Thus,

$$
\begin{aligned}
&\left|\mathrm{E}m_n(z) - m_y(z)\right| \\
&\quad\le \frac{|\delta_n|}{2}\left[1 + \frac{|2(z + y_n - 1) - y_n z\delta_n|}{|\sqrt{(z + y_n - 1)^2 - 4y_n z} + \sqrt{(z + y_n - 1 + y_n z\delta_n)^2 - 4y_n z}|}\right].
\end{aligned}
$$

Let us define by convention

$$
\Re(\sqrt{z}) = \frac{\Im(z)}{\sqrt{2(|z| - \Re(z))}}, \qquad \Im(\sqrt{z}) = \frac{|\Im(z)|}{\sqrt{2(|z| + \Re(z))}}.
$$

If $|u - y_n - 1| \ge \frac{1}{5(A+1)}$, then the real parts of $\sqrt{(z + y_n - 1)^2 - 4y_n z}$ and $\sqrt{(z + y_n - 1 + y_n z\delta_n)^2 - 4y_n z}$ have the same sign. Since they both have positive imaginary parts, it follows that

$$\begin{aligned}&\left|\sqrt{(z+y_n-1)^2-4y_nz}+\sqrt{(z+y_n-1+y_nz\delta_n)^2-4y_nz}\right|\\&\qquad\geq\sqrt{\left|\Im\big((z+y_n-1)^2-4y_nz\big)\right|}=\sqrt{2v(u-y_n-1)}\geq\left(\frac{2v}{5(A+1)}\right)^{1/2}.\end{aligned}$$

Thus,

$$\left|\mathrm{E}m_n(z)-m_y(z)\right|\leq\frac{|\delta_n|}{2}\left(1+\frac{C}{\sqrt{v}}\right)\leq\frac{C|\delta_n|}{\sqrt{v}}. \tag{A.2}$$

If $|u-y_n-1|<\frac{1}{5(A+1)}$, we have $|\mathrm{E}m_n(z)-m_y(z)|\leq C|\delta_n|$.

In [7] [see the inequality above (8.3.16)], we have

$$|\delta_n|\leq\frac{C}{Nv^3}\left(\Delta+\frac{v}{v_y}\right)^2\leq\frac{C}{Nvv_y^2}.$$

Combined with (A.2), we get

$$\left|\mathrm{E}m_n(z)-m_y(z)\right|\leq\frac{C}{Nv^{3/2}v_y^2}.$$

The proof of the lemma is complete.

In addition, the following relevant result which is involved in the proof of Lemma 3.1 is presented here:

$$\begin{aligned}&\int_{-A}^{A}\left|\mathrm{E}m_n(z)-m_y(z)\right|du\\&\qquad=\int_{|u-y_n-1|\geq1/(5(A+1)),|u|\leq A}\left|\mathrm{E}m_n(z)-m_y(z)\right|du\\&\qquad\quad+\int_{|u-y_n-1|<1/(5(A+1)),|u|\leq A}\left|\mathrm{E}m_n(z)-m_y(z)\right|du\\&\qquad\leq Cv.\end{aligned} \tag{A.3}$$

□

LEMMA A.2. *Under the conditions of Theorem 1.6, for $v^2v_y\geq C_0N^{-1}$ and $1\leq l\leq 3$, there exists a constant C, such that*

$$\mathrm{E}\left|\xi_1(z)\right|^{2l}\leq\frac{C}{N^lv^lv_y^l}.$$

PROOF. By the C_r-inequality (see Loève [14]), it follows that

$$\begin{aligned}\mathrm{E}\left|\xi_1(z)\right|^{2l}&\leq C\left(\mathrm{E}\left|\frac{1}{N}\operatorname{tr}\mathbf{A}_1^{-1}(z)-\frac{1}{N}\mathrm{E}\operatorname{tr}\mathbf{A}_1^{-1}(z)\right|^{2l}+\mathrm{E}\left|\hat{\xi}_1(z)\right|^{2l}\right)\\&=:I_1+I_2.\end{aligned}$$

From Lemmas B.6, B.8, B.10 and the C_r-inequality with $v^2 v_y \ge C_0 N^{-1}$, we have

$$\begin{aligned}
I_1 &\le \mathrm{E}\left|\frac{1}{N}\operatorname{tr}\mathbf{A}_1^{-1}(z) - \frac{1}{N}\operatorname{tr}\mathbf{A}^{-1}(z)\right|^{2l} + \mathrm{E}\left|\frac{1}{N}\operatorname{tr}\mathbf{A}^{-1}(z) - \frac{1}{N}\mathrm{E}\operatorname{tr}\mathbf{A}^{-1}(z)\right|^{2l} \\
&\quad + \left|\frac{1}{N}\mathrm{E}\operatorname{tr}\mathbf{A}^{-1}(z) - \frac{1}{N}\mathrm{E}\operatorname{tr}\mathbf{A}_1^{-1}(z)\right|^{2l} \\
&\le C\left\{\left(\frac{1}{Nv}\right)^{2l} + \frac{1}{N^{2l}v^{4l}}\left(\Delta + \frac{v}{v_y}\right)^{l} + \left(\frac{1}{Nv}\right)^{2l}\right\} \\
&\le \frac{C}{N^{2l}v^{3l}v_y^l}.
\end{aligned}$$

Under finite 8th moment assumption, for $l \ge 2$ we have

$$\mathrm{E}|X_{11}|^{4l} = \mathrm{E}\{|X_{11}|^8 |X_{11}|^{4l-8}\mathrm{I}(|X_{11}| \le \eta_N N^{1/4})\} \le C N^{l-2}.$$

It can be shown that $\mathbf{B}_1 - z\mathbf{I}_n = \mathbf{A}_1$, $\|(\mathbf{B}_1 - z\mathbf{I}_n)^{-1}\| \le 1/v$, and

$$\operatorname{tr}\big((\mathbf{B}_1 - z\mathbf{I}_n)^{-1}(\mathbf{B}_1 - \bar{z}\mathbf{I}_n)^{-1}\big) = v^{-1}\Im\big(\operatorname{tr}(\mathbf{B}_1 - z\mathbf{I}_n)^{-1}\big).$$

Hence, by Lemma B.5,

$$\begin{aligned}
I_2 &= \frac{1}{N^{2l}}\mathrm{E}\big|\mathbf{X}_1^*\mathbf{A}_1^{-1}(z)\mathbf{X}_1 - \operatorname{tr}\mathbf{A}_1^{-1}(z)\big|^{2l} \\
&\le \frac{C}{N^{2l}}\mathrm{E}\{CN^{l-2}\operatorname{tr}\big(\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar{z})\big)^l + \big(C\operatorname{tr}\big(\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar{z})\big)\big)^l\} \\
&= \frac{C}{N^{2l}}\mathrm{E}\{CN^{l-2}v^{-2l+1}\Im\big(\operatorname{tr}(\mathbf{B}_1 - z\mathbf{I}_n)^{-1}\big) \\
&\qquad\qquad + C^l v^{-l}\big(\Im\big(\operatorname{tr}(\mathbf{B}_1 - z\mathbf{I}_n)^{-1}\big)\big)^l\} \\
&= \frac{C}{N^{2l}}\{CN^{l-1}v^{-2l+1}\mathrm{E}\big(\Im\big(m_{F^{\mathbf{B}_1}}(z)\big)\big) \\
&\qquad\qquad + C^l N^l v^{-l}\mathrm{E}\big(\Im\big(m_{F^{\mathbf{B}_1}}(z)\big)\big)^l\} \\
&\le \frac{C}{N^{l+1}v^{2l-1}v_y} + \frac{C}{N^l v^l v_y^l}.
\end{aligned}$$

The last inequality is due to Lemmas B.6, B.7, B.8 and

$$\begin{aligned}
\big|\mathrm{E}\big(\Im\big(m_{F^{\mathbf{B}_1}}(z)\big)\big)^l\big| &\le \mathrm{E}\big|m_{F^{\mathbf{B}_1}}(z) - m_n(z)\big|^l + \mathrm{E}\big|m_n(z) - \mathrm{E}m_n(z)\big|^l \\
&\quad + \big|\mathrm{E}m_n(z) - m_y(z)\big|^l + \big|m_y(z)\big|^l \\
&\le C v_y^{-l}.
\end{aligned}$$

Here let $\Delta = \|EF^{\mathbf{S}_n} - F_{y_n}\|$, by integration by parts and Lemma B.10, we have

$$\big|\mathrm{E}m_n(z) - m_y(z)\big| \le \frac{C\Delta}{v} \le \frac{C}{v_y}. \tag{A.4}$$

Therefore, for $1 \le l \le 3$ and $v^2 v_y \ge C_0 N^{-1}$, it follows that

$$\mathrm{E}\big|\xi_1(z)\big|^l \le I_1 + I_2 \le \frac{C}{N^l v^l v_y^l}.$$

This completes the proof. $\square$

LEMMA A.3. *Under the conditions of Theorem 1.6, for all $|z| < A$ and $v^2 v_y \ge C_0 N^{-1}$, we have*

$$\big|b_1(z)\big| \le C.$$

PROOF. From (2.3) in [3], we have

$$m_y(z) = \frac{1 - y_n - z + \sqrt{(1 - y_n - z)^2 - 4y_n z}}{2y_n z},$$

where the square root of a complex number is defined as the one with positive imaginary part. It then can be verified that

$$\begin{aligned} b_0(z) =: \frac{1}{1 + y_n m_y(z)} &= 1 + \frac{1}{2}\Big(z - y_n - 1 - \sqrt{(z - y_n - 1)^2 - 4y_n}\Big) \\ &= 1 + \sqrt{y_n}\, m_{\mathrm{semi}}\bigg(\frac{z - y_n - 1}{\sqrt{y_n}}\bigg), \end{aligned}$$

where m_{semi} denotes the Stieltjes transform of the semicircular law. As $|m_{\mathrm{semi}}| \le 1$, we conclude that

$$\big|b_0(z)\big| \le 1 + \sqrt{y_n}. \tag{A.5}$$

By the relationship between b_0 and b_1, we have

$$b_1(z) = \frac{b_0(z)}{1 + y_n b_0(z)((1/n)\mathrm{E}\,\mathrm{tr}\,\mathbf{A}_1^{-1}(z) - m_y(z))}.$$

When C_0 is chosen large enough, by Lemma A.1, for all large N we have

$$\begin{aligned} \bigg|\frac{1}{n}\mathrm{E}\,\mathrm{tr}\,\mathbf{A}_1^{-1}(z) - m_y(z)\bigg| &\le \frac{1}{n}\big|\mathrm{E}\,\mathrm{tr}\big(\mathbf{A}_1^{-1}(z) - \mathbf{A}^{-1}(z)\big)\big| + \big|\mathrm{E}m_n(z) - m_y(z)\big| \\ &\le \frac{1}{nv} + \frac{1}{3(1 + \sqrt{y_n})} \le \frac{2}{3(1 + \sqrt{y_n})} \end{aligned}$$

and consequently we obtain

$$\big|b_1(z)\big| \le 3(1 + \sqrt{y_n}) \le C.$$

Thus, the proof is complete. $\square$

LEMMA A.4. *If* $|b_1(z)| \le C$, *then for any fixed* $t > 0$,

$$\mathrm{P}\big(|\beta_1(z)| > 2C\big) = o\big(N^{-t}\big).$$

PROOF. Note that if $|b_1(z)\xi_1(z)| \le 1/2$, by Lemma A.3, we get

$$|\beta_1(z)| = \frac{|b_1(z)|}{|1 + b_1(z)\xi_1(z)|} \le \frac{|b_1(z)|}{1 - |b_1(z)\xi_1(z)|} \le 2|b_1(z)| \le 2C.$$

As a result,

$$\begin{aligned}\mathrm{P}\big(|\beta_1(z)| > 2C\big) &\le \mathrm{P}\bigg(|b_1(z)\xi_1(z)| > \frac{1}{2}\bigg) \\ &\le \mathrm{P}\bigg(|\xi_1(z)| > \frac{1}{2C}\bigg) \qquad \text{(see Lemma A.3)} \\ &\le (2C)^p \mathrm{E}|\xi_1(z)|^p.\end{aligned}$$

By the C_r-inequality, Lemmas B.8 and B.9, for some $\eta = \eta_N N^{-1/4}$ and $p \ge \log N$, we have

$$\begin{aligned}\mathrm{E}|\xi_1(z)|^p &= \mathrm{E}|\hat{\xi}_1(z)|^p + \mathrm{E}\bigg|\frac{1}{N}\operatorname{tr}\mathbf{A}_1^{-1}(z) - \frac{1}{N}\mathrm{E}\operatorname{tr}\mathbf{A}_1^{-1}(z)\bigg|^p \\ &\le C\big(N\eta_N^4 N^{-1}\big)^{-1}\big(v^{-1}\eta_N^2 N^{-1/2}\big)^p + \frac{C}{N^p v^{3p/2} v_y^{p/2}} \\ &\le C\eta_N^{2p-4} \le C\eta_N^p.\end{aligned}$$

For any fixed $t > 0$, when N is large enough so that $\log \eta_N^{-1} > t + 1$, it can be shown that

$$\begin{aligned}\mathrm{E}|\xi_1(z)|^p &\le Ce^{-p\log\eta_N^{-1}} \\ &\le Ce^{-p(t+1)} \\ &\le Ce^{-(t+1)\log N} \\ &= CN^{-t-1} = o\big(N^{-t}\big).\end{aligned}$$

We finish the proof. □

LEMMA A.5. *If* $v^2 v_y \ge C_0 N^{-1}$, C_0 *is a large constant. For* $l \ge 1$, *it holds that*

$$\mathrm{E}|m_{H^{\mathbf{B}_1}}(z)|^{2l} \le C\mathrm{E}|m_n^H(z)|^{2l}.$$

PROOF. Recall that

$$\mathbf{A}_j^{-1}(z) - \mathbf{A}^{-1}(z) = \beta_j(z)\mathbf{A}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{A}_j^{-1}(z).$$

By Lemmas A.4 and B.5, it holds that

$$
\begin{aligned}
&\mathrm{E}\big|m_{H^{\mathbf{B}_1}}(z) - m_n^H(z)\big|^{2l} \\
&\quad = \mathrm{E}\big|\mathbf{x}_n^*\big(\mathbf{A}_1^{-1}(z) - \mathbf{A}^{-1}(z)\big)\mathbf{x}_n\big|^{2l} \\
&\quad = \mathrm{E}\big|\mathbf{x}_n^*\beta_1(z)\mathbf{A}_1^{-1}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\big|^{2l} \\
&\quad = \mathrm{E}\big|\mathbf{x}_n^*\beta_1(z)\mathbf{A}_1^{-1}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\big|^{2l}\mathrm{I}\big(\big|\beta_1(z)\big| \le C\big) \\
&\qquad + \mathrm{E}\big|\mathbf{x}_n^*\beta_1(z)\mathbf{A}_1^{-1}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\big|^{2l}\mathrm{I}\big(\big|\beta_1(z)\big| > C\big) \\
&\quad \le \frac{C}{N^{2l}}\mathrm{E}\big|\mathbf{X}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{X}_1\big|^{2l} + o\big(N^{-t}\big) \\
&\quad \le \frac{C}{N^{2l}}\mathrm{E}\big|\mathbf{X}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{X}_1 - \mathbf{x}_n^*\mathbf{A}_1^{-2}(z)\mathbf{x}_n\big|^{2l} + \frac{C}{N^{2l}}\mathrm{E}\big|\mathbf{x}_n^*\mathbf{A}_1^{-2}(z)\mathbf{x}_n\big|^{2l} \\
&\quad \le \frac{C}{N^{2l}}\big[\mathrm{E}\big(\operatorname{tr}\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(\bar z)\big)^l \\
&\qquad\qquad + N^{l-2}\mathrm{E}\operatorname{tr}\big(\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(\bar z)\big)^l\big] \\
&\qquad + \frac{C}{N^{2l}}\mathrm{E}\big|\mathbf{x}_n^*\mathbf{A}_1^{-2}(z)\mathbf{x}_n\big|^{2l} \\
&\quad \le \frac{C}{N^{l+2}v^{2l}}\mathrm{E}\big|m_{H^{\mathbf{B}_1}}(z)\big|^{2l}.
\end{aligned}
$$

The last step follows the fact that

$$
\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{A}_1^{-1}(\bar z)\mathbf{x}_n = v^{-1}\Im\big(\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\big) = v^{-1}\Im\big(m_{H^{\mathbf{B}_1}}(z)\big).
$$

For $v^2 v_y \ge C_0 N^{-1}$, which implies that $v \ge C_0 N^{-1/2}$. Choose C_0 and N large enough, such that $\frac{C}{N^{l+2}v^{2l}} \le \frac{1}{2}$. Further, by C_r-inequality, we obtain

$$
\begin{aligned}
\mathrm{E}\big|m_{H^{\mathbf{B}_1}}(z)\big|^{2l} &\le C\mathrm{E}\big|m_{H^{\mathbf{B}_1}}(z) - m_n^H(z)\big|^{2l} + C\mathrm{E}\big|m_n^H(z)\big|^{2l} \\
&\le \tfrac{1}{2}\mathrm{E}\big|m_{H^{\mathbf{B}_1}}(z)\big|^{2l} + C\mathrm{E}\big|m_n^H(z)\big|^{2l}.
\end{aligned}
$$

That is, $\mathrm{E}|m_{H^{\mathbf{B}_1}}(z)|^{2l} \le C\mathrm{E}|m_n^H(z)|^{2l}$, for some constant C. This finishes the proof. $\square$

LEMMA A.6. *Under the conditions of Theorem 1.6, for $v^2 v_y \ge C_0 N^{-1}$, we have*

$$
\mathrm{E}\big|m_n^H(z) - \mathrm{E}m_n^H(z)\big|^{2l} \le \frac{C}{N^l v^{2l}}\left(\frac{1}{v_y} + \frac{\Delta^H}{v}\right)^{2l}.
$$

PROOF. Write $\mathrm{E}_j(\cdot)$ as the conditional expectation given $\{\mathbf{r}_1, \dots, \mathbf{r}_j\}$. It can then be shown that $m_n^H(z) - \mathrm{E}m_n^H(z) = \sum_{j=1}^N \gamma_j$, where

$$\begin{aligned}\gamma_j &=: \mathrm{E}_j(\mathbf{x}_n^* \mathbf{A}^{-1}(z)\mathbf{x}_n) - \mathrm{E}_{j-1}(\mathbf{x}_n^* \mathbf{A}^{-1}(z)\mathbf{x}_n)\\ &= (\mathrm{E}_j - \mathrm{E}_{j-1})\{\mathbf{x}_n^*(\mathbf{A}^{-1}(z) - \mathbf{A}_j^{-1}(z))\mathbf{x}_n\}\\ &= -(\mathrm{E}_j - \mathrm{E}_{j-1})\{\beta_j(z)\mathbf{x}_n^*\mathbf{A}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{A}_j^{-1}(z)\mathbf{x}_n\}.\end{aligned}$$

Therefore, by Lemmas B.4(b), we have

$$\mathrm{E}|m_n^H(z) - \mathrm{E}m_n^H(z)|^{2l} \le C\mathrm{E}\left(\sum_{j=1}^N \mathrm{E}_{j-1}|\gamma_j|^2\right)^l + C\sum_{j=1}^N \mathrm{E}|\gamma_j|^{2l}.$$

Using Lemmas A.4 and B.5, we have

$$\begin{aligned}\mathrm{E}_{j-1} &= \mathrm{E}_{j-1}|(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z)\mathbf{x}_n^*\mathbf{A}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{A}_j^{-1}(z)\mathbf{x}_n|^2\\ &\le \frac{C}{N^2}\mathrm{E}_{j-1}|\mathbf{X}_j^*\mathbf{A}_j^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_j^{-1}(z)\mathbf{X}_j - \mathbf{x}_n^*\mathbf{A}_j^{-2}(z)\mathbf{x}_n|^2\\ &\quad + \frac{C}{N^2}\mathrm{E}_{j-1}|\mathbf{x}_n^*\mathbf{A}_j^{-2}(z)\mathbf{x}_n|^2\\ &\le \frac{C}{N^2}\mathrm{E}_{j-1}\operatorname{tr}(\mathbf{A}_j^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_j^{-1}(z)\mathbf{A}_j^{-1}(\bar{z})\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_j^{-1}(\bar{z}))\\ &\quad + \frac{C}{N^2}\mathrm{E}_{j-1}|\mathbf{x}_n^*\mathbf{A}_j^{-2}(z)\mathbf{x}_n|^2.\end{aligned}$$

By the fact that $\mathbf{x}_n^*\mathbf{A}_j^{-1}(z)\mathbf{A}_j^{-1}(\bar{z})\mathbf{x}_n = v^{-1}\Im(\mathbf{x}_n^*\mathbf{A}_j^{-1}(z)\mathbf{x}_n)$ and $\|\mathbf{A}_j^{-1}(z)\| \le v^{-1}$, we have

$$\mathrm{E}_{j-1}|\gamma_j|^2 \le \frac{C}{N^2v^2}\mathrm{E}_{j-1}|m_{H^{\mathbf{B}_j}}(z)|^2.$$

On the other side,

$$\begin{aligned}\mathrm{E}|\gamma_j|^{2l} &= \frac{1}{N^{2l}}\mathrm{E}|\mathbf{X}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{X}_1|^{2l}\\ &\le \frac{C}{N^{2l}}\mathrm{E}|\mathbf{X}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{X}_1 - \mathbf{x}_n^*\mathbf{A}_1^{-2}(z)\mathbf{x}_n|^{2l}\\ &\quad + \frac{C}{N^{2l}}\mathrm{E}|\mathbf{x}_n^*\mathbf{A}_1^{-2}(z)\mathbf{x}_n|^{2l}\\ &\le \frac{C}{N^{2l}} \times \frac{N^{l-2}}{v^{2l}}\mathrm{E}(\Im(\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n))^{2l} + \frac{C}{N^{2l}}\mathrm{E}|\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n|^{2l}\\ &\le \frac{C}{N^{l+2}v^{2l}}\mathrm{E}|m_{H^{\mathbf{B}_1}}(z)|^{2l}.\end{aligned}$$

Thus, we obtain

$$\begin{aligned}\mathrm{E}\big|m_n^H(z)-\mathrm{E}m_n^H(z)\big|^{2l} &\le \frac{C}{N^l v^{2l}}\mathrm{E}\big|m_{H^{\mathbf{B}_1}}(z)\big|^{2l}+\frac{C}{N^{l+1}v^{2l}}\mathrm{E}\big|m_{H^{\mathbf{B}_1}}(z)\big|^{2l}\\ &\le \frac{C}{N^l v^{2l}}\mathrm{E}\big|m_n^H(z)\big|^{2l} \qquad \text{(by Lemma A.5).}\end{aligned}$$

Further

$$\mathrm{E}\big|m_n^H(z)\big|^{2l}\le \mathrm{E}\big|m_n^H(z)-\mathrm{E}m_n^H(z)\big|^{2l}+\big|\mathrm{E}m_n^H(z)-m_y(z)\big|^{2l}+\big|m_y(z)\big|^{2l}.$$

For $v\ge C_0N^{-1/2}$, choose C_0 large enough, such that $\frac{C}{N^l v^{2l}}\le\frac{1}{2}$. And using integration by parts, it is easy to find that

$$\big|\mathrm{E}m_n^H(z)-m_y(z)\big|\le\frac{C\Delta^H}{v}, \tag{A.6}$$

where $\Delta^H=\|EH^{\mathbf{S}_n}-F_{y_n}\|$.

Besides, from Lemma B.7, we know that $|m_y(z)|\le\frac{C}{v_y}$.

Therefore, we obtain

$$\mathrm{E}\big|m_n^H(z)-\mathrm{E}m_n^H(z)\big|^{2l}\le\frac{C}{N^l v^{2l}}\left(\frac{1}{v_y}+\frac{\Delta^H}{v}\right)^{2l}.$$

The proof is then complete. $\square$

LEMMA A.7.

$$\mathrm{E}\big|\alpha_1(z)\big|^2\le\frac{C}{N^2v}\left(\frac{1}{v_y}+\frac{\Delta^H}{v}\right).$$

PROOF. Lemma B.5 implies that

$$\begin{aligned}\mathrm{E}\big|\alpha_1(z)\big|^2&=\frac{1}{N^2}\mathrm{E}\big|\mathbf{X}_1^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\mathbf{x}_n^*\mathbf{X}_1-\mathbf{x}_n^*\mathbf{A}_1^{-1}(z)\mathbf{x}_n\big|^2\\ &\le\frac{C}{N^2}\mathrm{E}\big(\mathbf{x}_n^*\mathbf{A}_1^{-1}(\bar z)\mathbf{A}_1^{-1}(z)\mathbf{x}_n\big)\\ &\le\frac{C}{N^2v}\big|\mathrm{E}m_{H^{\mathbf{B}_1}}(z)\big|.\end{aligned}$$

Using Lemmas A.5 and A.6 and integration by parts, we have

$$\begin{aligned}\mathrm{E}\big|m_{H^{\mathbf{B}_1}}(z)\big|^2&\le\mathrm{E}\big|m_n^H(z)\big|^2\\ &\le\mathrm{E}\big|m_n^H(z)-\mathrm{E}m_n^H(z)\big|^2+\big|\mathrm{E}m_n^H(z)-m_y(z)\big|^2+\big|m_y(z)\big|^2\\ &\le C\left(\frac{1}{v_y}+\frac{\Delta^H}{v}\right)^2.\end{aligned}$$

This finishes the proof. $\square$

LEMMA A.8. *Under the conditions in Theorem 1.1, for any fixed $t > 0$,*

$$\int_B^\infty \big|\mathrm{E}H^{\mathbf{S}_n}(x) - F_{y_n}(x)\big|\,dx = o\big(N^{-t}\big).$$

PROOF. For any fixed $t > 0$, by Lemma B.12, it follows that

$$\mathrm{P}\big(\lambda_{\max}(\mathbf{S}_n) \ge B + x\big) \le CN^{-t-1}(B + x - \varepsilon)^{-2}$$

and

$$\begin{aligned}
\int_B^\infty \big|\mathrm{E}H^{\mathbf{S}_n}(x) - F_{y_n}(x)\big|\,dx &\le \int_B^\infty \big(1 - \mathrm{E}H^{\mathbf{S}_n}(x)\big)\,dx \\
&= \int_B^\infty \left(1 - \sum_{i=1}^N |y_i|^2 \mathrm{P}(\lambda_i \le x)\right) dx \\
&\le \int_B^\infty \left(\sum_{i=1}^N |y_i|^2 - \sum_{i=1}^N |y_i|^2 \mathrm{P}(\lambda_{\max} \le x)\right) dx \\
&\le \int_B^\infty N^{-t-1}(B + x - \varepsilon)^{-2}\,dx = o\big(N^{-t}\big).
\end{aligned}$$

The proof is complete. $\square$

APPENDIX B

In what follows, we will present some existing results which are of substantial importance in proving the main theorems.

LEMMA B.1 (Theorem 2.2 in [2]). *Let F be a distribution function and let G be a function of bounded variation satisfying $\int |F(x) - G(x)|\,dx < \infty$. Denote their Stieltjes transforms by $f(z)$ and $g(z)$, respectively. Then*

$$\begin{aligned}
\|F - G\| &= \sup_x \big|F(x) - G(x)\big| \\
&\le \frac{1}{\pi(1-\kappa)(2\gamma - 1)} \\
&\quad \times \left(\int_{-A}^{A} \big|f(z) - g(z)\big|\,du + \frac{2\pi}{v}\int_{|x|>B} \big|F(x) - G(x)\big|\,dx \right. \\
&\qquad \left. + \frac{1}{v}\sup_x \int_{|y| \le 2v\tau} \big|G(x+y) - G(x)\big|\,dy\right),
\end{aligned}$$

where $z = u + iv$ is a complex variable, γ, κ, τ, A and B are positive constants such that $A > B$,

$$\kappa = \frac{4B}{\pi(A-B)(2\gamma - 1)} < 1$$

and

$$\gamma = \frac{1}{\pi}\int_{|u|<\tau}\frac{1}{u^2+1}\,du > \frac{1}{2}.$$

LEMMA B.2 (Lemma 8.15 in [7]). *For any $v>0$, we have*

$$\sup_x \int_{|u|<v} \left|F_{y_n}(x+u) - F_{y_n}(x)\right| du < \frac{11\sqrt{2(1+y_n)}}{3\pi y_n} v^2/v_y,$$

where F_{y_n} is the c.d.f. of the Marčenko–Pastur distribution with index $y_n \le 1$, and $v_y = 1-\sqrt{y_n}+\sqrt{v}$.

LEMMA B.3 ((1.15) in [6]). *Let $\mathbf{A} = (a_{ij})_{n\times n}$ and $\mathbf{B} = (b_{ij})_{n\times n}$ be two nonrandom matrices. Let $\mathbf{X} = (X_1,\dots,X_n)^*$ be a random vector of independent complex entries. Assume that $\mathrm{E}X_i = 0$ and $\mathrm{E}|X_i|^2 = 1$. Then we have*

$$\begin{aligned}&\mathrm{E}\big(\mathbf{X}^*\mathbf{A}\mathbf{X} - \operatorname{tr}\mathbf{A}\big)\big(\mathbf{X}^*\mathbf{B}\mathbf{X} - \operatorname{tr}\mathbf{B}\big)\\ &\quad = \sum_{i=1}^n \big(\mathrm{E}|X_i|^4 - |\mathrm{E}X_i^2|^2 - 2\big)a_{ii}b_{ii} + \left|\mathrm{E}X_i^2\right|^2 \operatorname{tr}\mathbf{A}\mathbf{B}^T + \operatorname{tr}\mathbf{A}\mathbf{B}.\end{aligned}$$

LEMMA B.4 (Burkholder inequalities (Lemmas 2.1 and 2.2 in [5])). *Let $\{X_k\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$, and let E_k denote the conditional expectation with respect to $\mathcal{F}_k$. Then we have*:

(a) *for $p>1$,*

$$\mathrm{E}\left|\sum_{k=1}^n X_k\right|^p \le K_p \mathrm{E}\left(\sum_{k=1}^n |X_k|^2\right)^{p/2},$$

(b) *for $p\ge 2$,*

$$\mathrm{E}\left|\sum_{k=1}^n X_k\right|^p \le K_p\left(\mathrm{E}\left(\sum_{k=1}^n \mathrm{E}_{k-1}|X_k|^2\right)^{p/2} + \mathrm{E}\sum_{k=1}^n |X_k|^p\right),$$

where K_p is a constant which depends on p only.

LEMMA B.5 (Lemma 2.7 in [5]). *Let $\mathbf{A} = (a_{ij})$ be an $n\times n$ nonrandom matrix and $\mathbf{X} = (X_1,\dots,X_n)^*$ be random vector of independent complex entries. Assume that $\mathrm{E}X_i = 0$, $\mathrm{E}|X_i|^2 = 1$ and $\mathrm{E}|X_i|^l \le V_l$. Then for any $p \ge 2$,*

$$\mathrm{E}\left|\mathbf{X}^*\mathbf{A}\mathbf{X} - \operatorname{tr}\mathbf{A}\right|^p \le K_p\big(\big(V_4 \operatorname{tr}(\mathbf{A}\mathbf{A}^*)\big)^{p/2} + V_{2p}\operatorname{tr}(\mathbf{A}\mathbf{A}^*)^{p/2}\big),$$

where K_p is a constant depending on p only.

LEMMA B.6 (Lemma 2.6 in [24]). *Let $z \in \mathbb{C}^+$ with $v = \Im(z)$, $\mathbf{A}$ and $\mathbf{B}$ $n \times n$ with $\mathbf{B}$ Hermitian, $\tau \in \mathbb{R}$, and $\mathbf{q} \in \mathbb{C}^n$. Then*

$$\left|\operatorname{tr}\left((\mathbf{B} - z\mathbf{I}_n)^{-1} - \left(\mathbf{B} + \tau\mathbf{q}\mathbf{q}^* - z\mathbf{I}_n\right)^{-1}\right)\mathbf{A}\right| \le \frac{\|\mathbf{A}\|}{v},$$

where $\|A\|$ denotes spectral norm on matrices.

LEMMA B.7 ((8.4.9) in [7]). *For the Stieltjes transform of the Marčenko–Pastur distribution, we have*

$$\left|m_y(z)\right| \le \frac{\sqrt{2}}{\sqrt{y}v_y},$$

where $v_y = \sqrt{a} + \sqrt{v} = 1 - \sqrt{y_n} + \sqrt{v}$.

LEMMA B.8 (Lemma 8.20 in [7]). *If $|z| < A$, $v^2 v_y \ge C_0 N^{-1}$ and $l \ge 1$, then*

$$\mathrm{E}\left|m_n(z) - \mathrm{E}m_n(z)\right|^{2l} \le \frac{C}{N^{2l} v^{4l} y_n^{2l}}\left(\Delta + \frac{v}{v_y}\right)^l,$$

where A is a positive constant, $v_y = 1 - \sqrt{y_n} + \sqrt{v}$ and $\Delta := \|\mathrm{E}F^{\mathbf{S}_n} - F_{y_n}\|$.

LEMMA B.9 (Lemma 9.1 in [7]). *Suppose that X_i, $i = 1, \dots, n$, are independent, with $\mathrm{E}X_i = 0$, $\mathrm{E}|X_i|^2 = 1$, $\sup \mathrm{E}|X_i|^4 = \nu < \infty$ and $|X_i| \le \eta\sqrt{n}$ with $\eta > 0$. Assume that $\mathbf{A}$ is a complex matrix. Then for any given p such that $2 \le p \le b\log(n\nu^{-1}\eta^4)$ and $b > 1$, we have*

$$\mathrm{E}\left|\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \operatorname{tr}(\mathbf{A})\right|^p \le \nu n^p \left(n\eta^4\right)^{-1}\left(40b^2\|\mathbf{A}\|\eta^2\right)^p,$$

where $\boldsymbol{\alpha} = (X_1, \dots, X_n)^T$.

LEMMA B.10 (Theorem 8.10 in [7]). *Let $\mathbf{S}_n = \mathbf{X}\mathbf{X}^*/N$, where $\mathbf{X} = (X_{ij}(n))_{n\times N}$. Assume that the following conditions hold:*

(1) *For each n, $X_{ij}(n)$ are independent,*
(2) $\mathrm{E}X_{ij}(n) = 0$, $\mathrm{E}|X_{ij}(n)|^2 = 1$, *for all i, j,*
(3) $\sup_n \sup_{i,j} \mathrm{E}|X_{ij}(n)|^6 < \infty$.

Then we have

$$\Delta =: \|\mathrm{E}F^{\mathbf{S}_n} - F_{y_n}\| = \begin{cases} O\left(N^{-1/2}a^{-1}\right), & \text{if } a > N^{-1/3}, \\ O\left(N^{-1/6}\right), & \text{otherwise}, \end{cases}$$

where $y_n = n/N \le 1$ and a is defined in the Marčenko–Pastur distribution.

LEMMA B.11 (Theorem 5.11 in [7]). *Assume that the entries of* $\{X_{ij}\}$ *is a double array of i.i.d. complex random variables with mean zero, variance* σ^2 *and finite* 4*th moment. Let* $\mathbf{X} = (X_{ij})_{n \times N}$ *be the* $n \times N$ *matrix of the upper-left corner of the double array. If* $n/N \to y \in (0, 1)$, *then, with probability one, we have*

$$\lim_{n\to\infty} \lambda_{\min}(\mathbf{S}_n) = \sigma^2(1 - \sqrt{y})^2$$

and

$$\lim_{n\to\infty} \lambda_{\max}(\mathbf{S}_n) = \sigma^2(1 + \sqrt{y})^2.$$

LEMMA B.12 (Theorem 5.9 in [7]). *Suppose that the entries of the matrix* $\mathbf{X} = (X_{ij})_{n\times N}$ *are independent* (*not necessarily identically distributed*) *and satisfy*:

(1) $\mathrm{E}X_{ij} = 0$,
(2) $|X_{ij}| \le \sqrt{N}\delta_N$,
(3) $\max_{ij} |\mathrm{E}|X_{ij}|^2 - \sigma^2| \to 0$ *as* $N \to \infty$ *and*
(4) $\mathrm{E}|X_{ij}|^l \le b(\sqrt{N}\delta_N)^{l-3}$ *for all* $l \ge 3$, *where* $\delta_N \to 0$ *and* $b > 0$. *Let* $\mathbf{S}_n = \mathbf{X}\mathbf{X}^*/N$. *Then*, *for any* $x > \epsilon > 0$, $n/N \to y$, *and fixed integer* $\ell \ge 2$, *we have*

$$\mathrm{P}\big(\lambda_{\max}(\mathbf{S}_n) \ge \sigma^2(1 + \sqrt{y})^2 + x\big) \le CN^{-\ell}\big(\sigma^2(1 + \sqrt{y})^2 + x - \epsilon\big)^{-\ell}$$

for some constant $C > 0$.

APPENDIX C

Note that the data matrix $\mathbf{X} = (X_{ij})_{n\times N}$ consists of i.i.d. complex random variables with mean 0 and variance 1. In what follows, we will further assume that every $|X_{ij}|$ is bounded by $\eta_N N^{1/4}$ for some carefully selected η_N. The proofs presented in the following three steps jointly justify such a convenient assumption.

C.1. Truncation for Theorem 1.1.

Choose $\eta_N \downarrow 0$ and $\eta_N N^{1/4} \uparrow \infty$ as $N \to \infty$ such that

$$\lim_{N\to\infty} \eta_N^{-10}\mathrm{E}|X_{11}|^{10}\mathrm{I}\big(|X_{11}| > \eta_N N^{1/4}\big) = 0.$$

Let $\underline{\widehat{\mathbf{X}}}_n$ denote the truncated data matrix whose entry on the ith row and jth column is $X_{ij}\mathrm{I}(|X_{ij}| \le \eta_N N^{1/4})$, $i = 1, \dots, n$, $j = 1, \dots, N$. Define $\widehat{\mathbf{S}}_n = \underline{\widehat{\mathbf{X}}}_n\underline{\widehat{\mathbf{X}}}_n^*/N$. Then

$$\begin{aligned}
\mathrm{P}(\mathbf{S}_n \ne \widehat{\mathbf{S}}_n) &\le nN\mathrm{P}\big(|X_{ij}| > \eta_N N^{1/4}\big)\\
&\le nN^{-3/2}\eta_N^{-10}\mathrm{E}|X_{11}|^{10}\mathrm{I}\big(|X_{11}| > \eta_N N^{1/4}\big)\\
&= o\big(N^{-1/2}\big).
\end{aligned}$$

C.2. Truncation for Theorems 1.6 and 1.8. Choose $\eta_N \downarrow 0$ and $\eta_N N^{1/4} \uparrow \infty$ as $N \to \infty$ such that

$$\lim_{N\to\infty} \eta_N^{-8} \mathrm{E}|X_{11}|^8 \mathrm{I}\big(|X_{11}| > \eta_N N^{1/4}\big) = 0. \tag{C.1}$$

Let $\underline{\widehat{\mathbf{X}}}_n$ denote the truncated data matrix whose entry on the ith row and jth column is $X_{ij}\mathrm{I}(|X_{ij}| \le \eta_N N^{1/4})$, $i = 1, \dots, n$, $j = 1, \dots, N$. Define $\widehat{\mathbf{S}}_n = \underline{\widehat{\mathbf{X}}}_n \underline{\widehat{\mathbf{X}}}_n^*/N$. Then

$$\begin{aligned}
\mathrm{P}(\mathbf{S}_n \neq \widehat{\mathbf{S}}_n, \text{i.o.}) &= \lim_{k\to\infty} \mathrm{P}\left(\bigcup_{N=k}^{\infty} \bigcup_{i=1}^{n} \bigcup_{j=1}^{N} |X_{ij}| > \eta_N N^{1/4}\right) \\
&= \lim_{k\to\infty} \mathrm{P}\left(\bigcup_{t=k}^{\infty} \bigcup_{N\in[2^t,2^{t+1})} \bigcup_{i=1}^{n} \bigcup_{j=1}^{N} |X_{ij}| > \eta_N N^{1/4}\right) \\
&\le \lim_{k\to\infty} \sum_{t=k}^{\infty} \mathrm{P}\left(\bigcup_{i=1}^{(y_n+1)2^{t+1}} \bigcup_{j=1}^{2^{t+1}} |X_{ij}| > \eta_{2^t} 2^{t/4}\right) \\
&\le C \lim_{k\to\infty} \sum_{t=k}^{\infty} (2^{t+1})^2 \mathrm{P}\big(|X_{11}| > \eta_{2^t} 2^{t/4}\big) \\
&\le C \lim_{k\to\infty} \sum_{t=k}^{\infty} \sum_{l=t}^{\infty} 4^t \mathrm{P}\big(\eta_{2^l} 2^{l/4} < |X_{11}| \le \eta_{2^{l+1}} 2^{(l+1)/4}\big) \\
&= C \lim_{k\to\infty} \sum_{l=k}^{\infty} \sum_{t=k}^{l} 4^t \mathrm{P}\big(\eta_{2^l} 2^{l/4} < |X_{11}| \le \eta_{2^{l+1}} 2^{(l+1)/4}\big) \\
&\le \lim_{k\to\infty} \sum_{l=k}^{\infty} C\eta_{2^l}^{-8} \mathrm{E}|X_{11}|^8 \mathrm{I}\big(\eta_{2^l} 2^{l/4} < |X_{11}| \le \eta_{2^{l+1}} 2^{(l+1)/4}\big) \\
&= 0.
\end{aligned}$$

The last equality is due to (C.1).

C.3. Centralization. The centralization procedures for three theorems are identical, only 8th moment is required and thus we treat them uniformly. Let $\underline{\widetilde{\mathbf{X}}}_n$ denote the centralized version of $\underline{\widehat{\mathbf{X}}}_n$. More explicitly, on the ith row and jth column of $\underline{\widetilde{\mathbf{X}}}_n$, the entry is

$$X_{ij}\mathrm{I}\big(|X_{ij}| \le \eta_N N^{1/4}\big) - \mathrm{E}\big(X_{ij}\mathrm{I}\big(|X_{ij}| \le \eta_N N^{1/4}\big)\big).$$

Notice that according to Theorem 3.1 of [26], $\|(\mathbf{S}_n - z\mathbf{I}_n)^{-1}\|$ is bounded by $1/v$, where $\|\cdot\|$ denotes the spectral norm for a matrix. Define $\widetilde{\mathbf{S}}_n = \underline{\widetilde{\mathbf{X}}}_n \underline{\widetilde{\mathbf{X}}}_n^*/N$.

Suppose that $v \geq C_0 N^{-1/2}$, we obtain

$$
\begin{aligned}
&\left|m_{H^{\widehat{\mathbf{S}}_n}}(z) - m_{H^{\widetilde{\mathbf{S}}_n}}(z)\right| \\
&\quad = \left|\mathbf{x}_n^*(\widehat{\mathbf{S}}_n - z\mathbf{I}_n)^{-1}\mathbf{x}_n - \mathbf{x}_n^*(\widetilde{\mathbf{S}}_n - z\mathbf{I}_n)^{-1}\mathbf{x}_n\right| \\
&\quad \leq \|(\widehat{\mathbf{S}}_n - z\mathbf{I}_n)^{-1}\| \|\widehat{\mathbf{S}}_n - \widetilde{\mathbf{S}}_n\| \|(\widetilde{\mathbf{S}}_n - z\mathbf{I}_n)^{-1}\| \\
&\quad \leq \frac{1}{v^2}\|\widehat{\mathbf{S}}_n - \widetilde{\mathbf{S}}_n\| \\
&\quad \leq \frac{1}{Nv^2}\left(\|\widehat{\underline{\mathbf{X}}}_n\| \|\widehat{\underline{\mathbf{X}}}_n^* - \widetilde{\underline{\mathbf{X}}}_n^*\| + \|\widehat{\underline{\mathbf{X}}}_n - \widetilde{\underline{\mathbf{X}}}_n\| \|\widetilde{\underline{\mathbf{X}}}_n^*\|\right) \qquad \text{by Lemma B.12} \\
&\quad \leq \frac{C}{\sqrt{N}v^2}\|\widehat{\underline{\mathbf{X}}}_n - \widetilde{\underline{\mathbf{X}}}_n\| \qquad \text{a.s.} \\
&\quad = \frac{C}{\sqrt{N}v^2}\left|\mathrm{E}\left\{X_{11}\mathrm{I}\left(|X_{11}| \leq \eta_N N^{1/4}\right)\right\}\right| \|\mathbf{1}_{n\times 1}\| \|\mathbf{1}'_{N\times 1}\| \\
&\quad \leq C\sqrt{N}v^{-2}\eta_N^{-7}N^{-7/4}\mathrm{E}\left(|X_{11}|^8\mathrm{I}\left(|X_{11}| > \eta_N N^{1/4}\right)\right) \\
&\quad = o\left(N^{-1/4}\right).
\end{aligned}
$$

To establish both the weak and the strong convergence rates of the VESD to the Marčenko–Pastur distribution, this $o(N^{-1/4})$ suffices. Moreover, for the convergence rate presented in Theorem 1.1, we shall prove the following. Let $m_y(z)$ denotes the Stieltjes transform of the Marčenko–Pastur distribution, thus $m_y(z)$ is bounded by $\frac{\sqrt{2}}{\sqrt{y_n}(1-\sqrt{y_n}+\sqrt{v})} \leq \frac{C}{\sqrt{v}}$ in Lemma B.7. Then

$$
\left|\mathrm{E}m_n^H(z)\right| \leq \left|\mathrm{E}m_n^H(z) - m_y(z)\right| + \left|m_y(z)\right| \leq C\left|m_y(z)\right| \leq \frac{C}{\sqrt{v}}.
$$

Besides $\mathbf{x}_n^*(\widehat{\mathbf{S}}_n - z\mathbf{I}_n)^{-1}\mathbf{x}_n$ can be considered as a Stieltjes transform of some VESD function. So, we have

$$
\mathrm{E}\|\mathbf{x}_n^*(\widehat{\mathbf{S}}_n - z\mathbf{I}_n)^{-1}\|^2 = v^{-1}\mathrm{E}\mathbf{x}_n^*(\widehat{\mathbf{S}}_n - z\mathbf{I}_n)^{-1}\mathbf{x}_n \leq \frac{C}{\sqrt{v}}.
$$

Thus,

$$
\begin{aligned}
&\mathrm{E}\left|m_{H^{\widehat{\mathbf{S}}_n}}(z) - m_{H^{\widetilde{\mathbf{S}}_n}}(z)\right| \\
&\quad \leq \mathrm{E}\|\widehat{\mathbf{S}}_n - \widetilde{\mathbf{S}}_n\| \|\mathbf{x}_n^*(\widehat{\mathbf{S}}_n - z\mathbf{I}_n)^{-1}\| \|(\widetilde{\mathbf{S}}_n - z\mathbf{I}_n)^{-1}\mathbf{x}_n\| \\
&\quad \leq C\sqrt{N}\eta_N^{-7}N^{-7/4}\mathrm{E}\left(|X_{11}|^8\mathrm{I}\left(|X_{11}| \geq \eta_N N^{1/4}\right)\right) \\
&\qquad \times \left(\mathrm{E}\|\mathbf{x}_n^*(\widehat{\mathbf{S}}_n - z\mathbf{I})^{-1}\|^2\right)^{1/2}\left(\mathrm{E}\|(\widetilde{\mathbf{S}}_n - z\mathbf{I})^{-1}\mathbf{x}_n\|^2\right)^{1/2} \\
&\quad \leq C\sqrt{N}v^{-3/2}\eta_N^{-7}N^{-7/4}\mathrm{E}\left(|X_{11}|^8\mathrm{I}\left(|X_{11}| \geq \eta_N N^{1/4}\right)\right) \\
&\quad \leq o\left(N^{-1/2}\right).
\end{aligned}
$$

C.4. Rescaling. The rescaling procedures for the three theorems are exactly the same, and only 8th moment is required. Thus, we treat them uniformly. Write $\underline{\mathbf{Y}}_n = \underline{\widetilde{\mathbf{X}}}_n/\sigma_1$, where

$$\sigma_1^2 = \mathrm{E}\big|X_{11}\mathrm{I}(|X_{11}| \le \eta_N N^{1/4}) - \mathrm{E}(X_{11}\mathrm{I}(|X_{11}| \le \eta_N N^{1/4}))\big|^2.$$

Notice that σ_1 tends to 1 as N goes to ∞. Define $\mathbf{G}_n = \underline{\mathbf{Y}}_n\underline{\mathbf{Y}}_n^*/N$, which is the sample covariance matrix of $\underline{\mathbf{Y}}_n$. We shall show that $\mathbf{G}_n$ and $\mathbf{S}_n$ are asymptotically equivalent, that is, the VESD of $\mathbf{G}_n$ and $\mathbf{S}_n$ have the same limit if either one limit exists. For $v \ge C_0 N^{-1/2}$,

$$\begin{aligned}
\big|m_{H^{\mathbf{G}_n}}(z) - m_{H^{\widetilde{\mathbf{S}}_n}}(z)\big| &= \big|\mathbf{x}_n^*(\widetilde{\mathbf{S}}_n - z\mathbf{I}_n)^{-1}(\widetilde{\mathbf{S}}_n - \mathbf{G}_n)(\mathbf{G}_n - z\mathbf{I}_n)^{-1}\mathbf{x}_n\big| \\
&\le \frac{1}{v^2}\big\|(1-\sigma_1^{-1})\widetilde{\mathbf{S}}_n\big\| \qquad \text{(see Lemma B.12)} \\
&\le \frac{C}{v^2}(1-\sigma_1^2) \qquad \text{a.s.} \\
&\le Cv^{-2}\eta_N^{-6}N^{-3/2}\mathrm{E}(|X_{11}|^8\mathrm{I}(|X_{11}| > \eta_N N^{1/4})) \\
&\le o(N^{-1/2}) \qquad \text{a.s.}
\end{aligned}$$

Hence, we shall without loss of generality assume that every $|X_{ij}|$ is bounded by $\eta_N N^{1/4}$, and every X_{ij} has mean 0 and variance 1.

REFERENCES

[1] ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Inst. Statist. Math.* **34** 122–148. MR0145620

[2] BAI, Z. D. (1993). Convergence rate of expected spectral distributions of large random matrices. I. Wigner matrices. *Ann. Probab.* **21** 625–648. MR1217559

[3] BAI, Z. D. (1993). Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. *Ann. Probab.* **21** 649–672. MR1217560

[4] BAI, Z. D., MIAO, B. Q. and PAN, G. M. (2007). On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab.* **35** 1532–1572. MR2330979

[5] BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051

[6] BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792

[7] BAI, Z. D. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

[8] BAI, Z. D. and XIA, N. (2013). Functional CLT of eigenvectors for large sample covariance matrices. *Statist. Papers*. To appear.

[9] CAI, T., MA, Z. M. and WU, Y. H. (2013). Sparse PCA: Optimal rates and adaptive estimation. Available at arXiv:1211.1309.

[10] ERDŐS, L., SCHLEIN, B. and YAU, H.-T. (2009). Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.* **37** 815–852. MR2537522

[11] GÖTZE, F. and TIKHOMIROV, A. (2004). Rate of convergence in probability to the Marchenko–Pastur law. *Bernoulli* **10** 503–548. MR2061442

[12] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

[13] KNOWLES, A. and YIN, J. (2013). Eigenvector distribution of Wigner matrices. *Probab. Theory Related Fields* **155** 543–582. MR3034787

[14] LOÈVE, M. (1977). *Probability Theory. I*, 4th ed. Springer, New York. *Graduate Texts in Mathematics* **45**. MR0651017

[15] MA, Z. M. (2013). Sparse principal component analysis and iterative thresholding. *Ann. Statist.* **41** 772–801.

[16] MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sb.* **1** 457–483.

[17] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865

[18] PILLAI, N. S. and YIN, J. (2013). Universality of covariance matrices. Available at arXiv:1110.2501v6.

[19] SILVERSTEIN, J. W. (1981). Describing the behavior of eigenvectors of random matrices using sequences of measures on orthogonal groups. *SIAM J. Math. Anal.* **12** 274–281. MR0605435

[20] SILVERSTEIN, J. W. (1984). Some limit theorems on the eigenvectors of large-dimensional sample covariance matrices. *J. Multivariate Anal.* **15** 295–324. MR0768500

[21] SILVERSTEIN, J. W. (1989). On the eigenvectors of large-dimensional sample covariance matrices. *J. Multivariate Anal.* **30** 1–16. MR1003705

[22] SILVERSTEIN, J. W. (1990). Weak convergence of random functions defined by the eigenvectors of sample covariance matrices. *Ann. Probab.* **18** 1174–1194. MR1062064

[23] SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. MR1370408

[24] SILVERSTEIN, J. W. and BAI, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.* **54** 175–192. MR1345534

[25] TAO, T. and VU, V. (2011). Universal properties of eigenvectors. Available at arXiv:1103.2801v2.

[26] YIN, Y. Q., BAI, Z. D. and KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521. MR0950344

N. XIA
Z. D. BAI
KLAS AND SCHOOL OF MATHEMATICS
AND STATISTICS
NORTHEAST NORMAL UNIVERSITY
CHANGCHUN 130024
CHINA
AND DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
SINGAPORE 117546
SINGAPORE
E-MAIL: xiann664@gmail.com
baizd@nenu.edu.cn

Y. QIN
DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
UNIVERSITY OF WATERLOO
WATERLOO, ON N2L 3G1
CANADA
E-MAIL: yingli.qin@uwaterloo.ca