Available online at www.sciencedirect.com
SciVerse ScienceDirect
Stochastic Processes and their Applications 122 (2012) 68–92
stochastic processes and their applications
www.elsevier.com/locate/spa

Convergence rates to the Marchenko–Pastur type distribution

Zhidong Bai[a,b], Jiang Hu[a,*], Wang Zhou[b]

[a] KLASMOE and School of Mathematics & Statistics, Northeast Normal University, Changchun, 130024, PR China
[b] Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546, Singapore

Received 24 March 2011; received in revised form 24 June 2011; accepted 3 October 2011
Available online 13 October 2011

Abstract

$\mathbf{S}_n = \frac{1}{n}\mathbf{T}_n^{1/2}\mathbf{X}_n\mathbf{X}_n^*\mathbf{T}_n^{1/2}$, where $\mathbf{X}_n = (x_{ij})$ is a $p \times n$ matrix consisting of independent complex entries with mean zero and variance one, $\mathbf{T}_n$ is a $p \times p$ nonrandom positive definite Hermitian matrix with spectral norm uniformly bounded in p. In this paper, if $\sup_n \sup_{i,j} \mathbb{E} \mid x_{ij}^8 \mid < \infty$ and $y_n = p/n < 1$ uniformly as $n \to \infty$, we obtain that the rate of the expected empirical spectral distribution of $\mathbf{S}_n$ converging to its limit spectral distribution is $O(n^{-1/2})$. Moreover, under the same assumption, we prove that for any $\eta > 0$, the rates of the convergence of the empirical spectral distribution of $\mathbf{S}_n$ in probability and the almost sure convergence are $O(n^{-2/5})$ and $O(n^{-2/5+\eta})$ respectively.
© 2011 Elsevier B.V. All rights reserved.

MSC: primary 60F15; secondary 62H99

Keywords: Convergence rate; Sample covariance matrix; Spectral distribution

1. Introduction and results

Let $\mathbf{X}_n = (x_{ij})_{p\times n}$, where x_{ij}'s are independent random complex variables with $\mathbb{E}x_{ij} = 0$ and $\mathbb{E}|x_{ij}|^2 = 1$, and let $\mathbf{T}_n$ be a $p \times p$ non-random positive definite Hermitian matrix whose spectral norm is bounded by a constant independent of p. In this paper, we consider a class of sample covariance matrices,

$$\mathbf{S}_n = \frac{1}{n}\mathbf{T}_n^{1/2}\mathbf{X}_n\mathbf{X}_n^*\mathbf{T}_n^{1/2},$$

* Corresponding author. Tel.: +86 3804356730.
E-mail addresses: baizd@nenu.edu.cn (Z. Bai), huj156@nenu.edu.cn (J. Hu), stazw@nus.edu.sg (W. Zhou).

0304-4149/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.spa.2011.10.002

where $\mathbf{T}_n^{1/2}$ is any square root of $\mathbf{T}_n$ and $\mathbf{X}_n^*$ denotes the conjugate transpose of the matrix $\mathbf{X}_n$. Here $\mathbf{S}_n$ can be viewed as the sample covariance matrix based on n samples from a p dimensional population, whose population matrix is $\mathbf{T}_n$ (see [24,7,9,4]). Moreover, if $\mathbf{T}_n$ is taken as the inverse of another sample covariance matrix which is independent of $\mathbf{X}_n$, then $\mathbf{S}_n$ is an F-matrix (see [23,4]).

In recent years, since high dimensional data occur in many modern scientific fields, the random matrix theory (RMT) in both theoretical investigations and applications becomes more and more important. For example, RMT was widely applied in the context of vector signal processing. Another typical application is in finance. Details can be found in [10,21,5,14,22,13].

We start with some basic facts about RMT. For any $n \times n$ matrix $\mathbf{A}$ with only real eigenvalues, let $F^{\mathbf{A}}$ denote the empirical spectral distribution function (ESDF) of $\mathbf{A}$, that is

$$F^{\mathbf{A}}(x) = \frac{1}{n}\sum_{i=1}^{n} I(\lambda_i^{\mathbf{A}} \le x),$$

where $\lambda_i^{\mathbf{A}}$ denotes the i-th smallest eigenvalue of $\mathbf{A}$. The results in [27,24] state that $F^{\mathbf{S}_n}$ converges almost surely (a.s.) to a non-random distribution $F^{y,H}$ under the assumption that, as $n \to \infty$, the ratio $y_n = p/n \to y < \infty$ and $H_n := F^{\mathbf{T}_n} \xrightarrow{\mathcal{D}} H$, a proper distribution. The notation $F^{y,H}$ means that the limiting spectral distribution function (LSDF) of the sequence $\{\mathbf{S}_n\}$, which is also known as Marchenko–Pastur (M–P) type distribution, depends on the limiting dimension to sample size ratio y and LSDF H of the population matrix. It is shown in [7] that the Stieltjes transform $s(z)$ of $F^{y,H}$ is a solution to the equation

$$s = \int \frac{1}{t(1-y-yzs-z)} dH(t), \tag{1.1}$$

which is unique in the set $\{s \in \mathbb{C} : -(1-y)/z + ys \in \mathbb{C}^+\}$. Here for any function of bounded variation G on the real line, its Stieltjes transform is defined by

$$s_G(z) = \int \frac{1}{\lambda - z} dG(\lambda), \quad z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C}^+ : \Im z > 0\}.$$

Silverstein and Choi in [26] derive some analytic properties of $F^{y,H}$, such as the continuous dependence of $F^{y,H}$ on y and H, and the fact that the limit of $s(z)$ as $z \downarrow c \in \mathbb{R} \setminus 0$ exists and is continuous. Later Bai and Silverstein in [9] establish the central limit theorem for linear spectral statistics of $\mathbf{S}_n$. Recently, Jing et al. in [20] give another way of understanding the LSDF $F^{y,H}$, i.e., using kernel estimators to estimate $F^{y,H}$.

Clearly, in order to refine the above approximation, we need to obtain the convergence rate. In this paper, we focus on establishing the convergence rates of the ESDF of $\mathbf{S}_n$. In other words, we will study the rate of the following Kolmogorov distances tending to zero,

$$\Delta_n := \sup_x |\mathbb{E}F^{\mathbf{S}_n}(x) - F^{y_n,H_n}(x)| \quad \text{and} \quad \tilde{\Delta}_n := \sup_x |F^{\mathbf{S}_n}(x) - F^{y_n,H_n}(x)|$$

where F^{y_n,H_n} is the distribution function obtained from $F^{y,H}$ by replacing y and H with y_n and H_n respectively.

The rates of convergence play important roles in the applications of the spectral analysis of large random matrices which can be viewed in [12,20,11]. If $\mathbf{T}_n = \mathbf{I}$, the $p \times p$ identity matrix, given the condition that $y = \lim_{n\to\infty} y_n$ stays away from 0 and 1, Bai [2] proves that the convergence rate of Δ_n to zero is $O(n^{-1/4})$ under the assumption that $\sup_n \sup_{i,j} \mathbb{E}x_{ij}^4 I(|x_{ij}| \geq M) \to 0$ as $M \to \infty$. Ten years later, Bai et al. [6] and Götze and Tikhomirov [18] improve the rate to $O(n^{-1/2})$ at the cost of the eighth moments of the matrix entries. If y is close to 1, since the limit density function and its Stieltjes transform have singularities, the research of the rates of convergence becomes more difficult. In this case, Bai [2] proves that the order is $O(n^{-5/48})$. In [6], it is improved to $O(n^{-1/8})$. In [19], Götze and Tikhomirov obtain the bound $O(n^{-1/2})$ under the assumption $\sup_{i,j} \mathbb{E}x_{ij}^4 < \infty$ and $0 < \theta \leq y_n \leq 1$. When $\mathbf{T}_n$ is not the identity matrix, Jing et al. in [20] prove that $\Delta_n = O(n^{-2/5})$, assuming that x_{ij}'s are independent and identically distributed, $\mathbb{E}x_{11}^{16} < \infty$, and $y_n \to y \in (0, 1)$; and they apply this result to estimate the spectral density functions of sample covariance matrices. However, the exact rates and the optimal conditions of the convergence are still open. Jing et al. in [20] conjecture that the rate of $\tilde{\Delta}_n$ could be $O(n^{-1}\sqrt{\log n})$ in probability (i.p.) under the fourth moment condition. Our main results of this paper are as follows.

Theorem 1.1. *Assume that* $0 < \theta \leq y_n \leq \Theta < 1$ *for positive constants* θ *and* Θ, $\sup_n \sup_{i,j} \mathbb{E}|x_{ij}^8| < \infty$, $\lim_{n\to\infty} \lambda_1^{\mathbf{T}_n} = \lambda_0 > 0$, $\sup_n \lambda_p^{\mathbf{T}_n} < \infty$ *and* $F^{\mathbf{T}_n} \xrightarrow{\mathcal{D}} H$, *a proper distribution function. Then we have*

$$\Delta_n = O(n^{-1/2}). \tag{1.2}$$

Remark 1.2. The condition $\sup_n \lambda_p^{\mathbf{T}_n} < \infty$ is equivalent to $\sup_n \lambda_p^{\mathbf{T}_n} \leq 1$. One can see this via re-scaling $\mathbf{T}_n$ by $(\sup_n \lambda_p^{\mathbf{T}_n})^{-1}$. From now on, we will assume that $\sup_n \lambda_p^{\mathbf{T}_n} \leq 1$.

Theorem 1.3. *Under the same assumptions of Theorem 1.1, we have*

$$\tilde{\Delta}_n = O(n^{-2/5}) \quad \textit{i.p.}, \tag{1.3}$$

and for any $\eta > 0$,

$$\tilde{\Delta}_n = O(n^{-2/5+\eta}) \quad \textit{a.s.} \tag{1.4}$$

Remark 1.4. It is worth noting that, as in [2], it is impossible to establish any rate of $\sup_x |\mathbb{E}F^{\mathbf{S}_n}(x) - F^{y,H}(x)|$ because we know nothing about the rates of $y_n \to y$ and $F^{\mathbf{T}_n} \xrightarrow{\mathcal{D}} H$. This is why we choose F^{y_n,H_n} instead of $F^{y,H}$ in these theorems.

Remark 1.5. The condition $\sup_n \sup_{i,j} \mathbb{E}|x_{ij}^8| < \infty$ is just a technical assumption. This assumption is essentially required when we apply Lemma 6.5 and the truncation Lemma 3.1. However we believe that the fourth moment is sufficient.

The rest of this paper is organized as follows. The main tools of proving the theorems and some basic consequences are introduced in Section 2. In Section 3, we take the truncation and centralization step. Theorem 1.1 is proved in Section 3 and the proof of Theorem 1.3 is provided in Section 4. Some technical lemmas are given in Section 5. Throughout this paper, constants appearing in inequalities are represented by C which are nonrandom and may take different values from one appearance to another.

2. The main tools and easy consequences

Our main tools to prove the theorems are two Berry–Esseen type inequalities which are proved by Götze and Tikhomirov in [17] and by Bai in [1].

Proposition 2.1 (*Lemma 2.1 in [19]*). *Let F and G be distribution functions satisfying $\int |F(x) - G(x)|dx < \infty$. Denote their Stieltjes transforms by $s_F(z)$ and $s_G(z)$ respectively, where $z \in \mathbb{C}^+$. Assume that the distribution G has a support contained in the bounded interval $[a, b]$ and admits a density g such that $g(x) \leq c_g$ for some positive constant $c_g > 0$. Then there exists some constant $C > 0$ depending only on c_g such that, for any $0 < v < V$,*

$$\sup_x |F(x) - G(x)| \leq C\left(\int_{-\infty}^{\infty} |s_F(u + iV) - s_G(u + iV)|du \right.$$
$$\left. + v + \sup_{u\in[a,b]} \left|\Re\left\{\int_v^V (s_F(u + ix) - s_G(u + ix))dx\right\}\right|\right). \qquad (2.1)$$

Proposition 2.2 (*Theorem 2.2 in [1]*). *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 $s_F(z)$ and $s_G(z)$ respectively, where $z = u + iv \in \mathbb{C}^+$. Then*

$$\sup_x |F(x) - G(x)| \leq \frac{1}{\pi(1-\zeta)(2\rho - 1)} \left(\int_{-A}^{A} |s_F(z) - s_G(z)|du \right.$$
$$\left. + 2\pi v^{-1} \int_{|x|>B} |F(x) - G(x)|dx + v^{-1} \sup_x \int_{|u|\leq 2vc_*} |G(x + u) - G(x)|du\right) \qquad (2.2)$$

where the constants $A > B > 0$, ζ and c_* are restricted by $\rho = \frac{1}{\pi}\int_{|u|\leq c_*} \frac{1}{u^2+1}du > \frac{1}{2}$, and $\zeta = \frac{4B}{\pi(A-B)(2\rho-1)} \in (0, 1)$.

We use Proposition 2.1 to prove Theorem 1.1 and use Proposition 2.2 to prove Theorem 1.3. Moreover, the bound of the density function of $F^{y,H}$ is required when we apply the two propositions. For example, in Proposition 2.1, we require the density function of $F^{y,H}$ to be bounded by some constant. It is also useful when we estimate the third integral in (2.2). So here we present a lemma on it.

Lemma 2.3. *Let $s(z)$ be the Stieltjes transform of the LSDF of matrices $\mathbf{S}_n$, where $z = u + iv \in \mathbb{C}^+$. If $u \in [a, b]$ with $a > 0$, then there exists a positive constant C such that for all $v > 0$, we have*

$$\sup_{u\in[a,b]} |s(z)| \leq C \quad \text{and} \quad f^{y,H}(x) \leq \frac{1}{\pi\sqrt{\lambda_0 x y}},$$

where $f^{y,H}(x)$ is the density function of $F^{y,H}(x)$.

Before proving this lemma, we give some notation and facts which will be well used in the following parts. Let $\mathbf{I}$ be the identity matrix. Define $\mathbf{D} = \mathbf{D}(z) = \mathbf{S}_n - z\mathbf{I}$, $\mathbf{r}_j = n^{-1/2}\mathbf{T}_n^{1/2}\mathbf{X}_{(\cdot j)}$

where $\mathbf{X}_{(\cdot j)}$ is the j-th column of $\mathbf{X}_n$ and $\mathbf{D}_j = \mathbf{D}_j(z) = \mathbf{D}(z) - \mathbf{r}_j\mathbf{r}_j^*$. Moreover introduce

$$\begin{aligned}
&a_j = \mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{T}_n\mathbf{D}_j^{-1}\mathbf{r}_j, \qquad \alpha_j = a_j - n^{-1}\mathbb{E}\mathrm{tr}(\mathbf{D}_j^{-1}\mathbf{T}_n\mathbf{D}_j^{-1}\mathbf{T}_n),\\
&\gamma_j = \mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j - n^{-1}\mathbb{E}\mathrm{tr}(\mathbf{D}_j^{-1}\mathbf{T}_n), \qquad \hat{\gamma}_j = \mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j - n^{-1}\mathrm{tr}(\mathbf{D}_j^{-1}\mathbf{T}_n),\\
&\beta_j = (1+\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j)^{-1}, \qquad b_j = (1+n^{-1}\mathbb{E}\mathrm{tr}(\mathbf{D}_j^{-1}\mathbf{T}_n))^{-1},\\
&s_n = s_n(z) = s_{F^{\mathbf{S}_n}}(z), \qquad s = s(z) = s_{F^{y,H}}(z), \qquad s_0 = s_0(z) = s_{F^{y_n,H_n}}(z).
\end{aligned} \tag{2.3}$$

Let $\underline{\mathbf{S}}_n = \frac{1}{n}\mathbf{X}_n^*\mathbf{T}_n\mathbf{X}_n$, $\underline{s}_n = \underline{s}_n(z) = s_{F^{\underline{\mathbf{S}}_n}}(z)$, $\underline{F}^{y,H} = \lim_{n\to\infty} F^{\underline{\mathbf{S}}_n}$, $\underline{s} = \underline{s}(z) = s_{\underline{F}^{y,H}}(z)$ and $\underline{s}_0 = \underline{s}_0(z) = s_{\underline{F}^{y_n,H_n}}(z)$. Since the spectra of $\underline{\mathbf{S}}_n$ and $\mathbf{S}_n$ differ by $|n-p|$ zero eigenvalues, it follows that

$$F^{\underline{\mathbf{S}}_n} = (1-y_n)I([0,\infty)) + y_nF^{\mathbf{S}_n},$$

which implies

$$\underline{F}^{y,H} = (1-y)I([0,\infty)) + yF^{y,H},$$

$$z\underline{s}_n(z) = y_n - 1 + zy_ns_n(z), \tag{2.4}$$

and

$$z\underline{s}(z) = y - 1 + zys(z). \tag{2.5}$$

Substituting (2.5) into (1.1), we obtain

$$s(z) = -\frac{1}{z}\int\frac{1}{1+t\underline{s}}dH(t). \tag{2.6}$$

Moreover $\underline{s}(z)$ has an inverse

$$z(\underline{s}) = -\frac{1}{\underline{s}} + y\int\frac{t}{1+t\underline{s}}dH(t). \tag{2.7}$$

Note that if $\mathbf{T}_n$ is the identity matrix, there is an explicit solution to (1.1). Thus, in this case we can get the explicit expression of the destiny function $f^{y,H}$ by the inversion formula of the Stieltjes transform (Lemma 6.8). Then analysing the asymptomatics of the ESDF $F^{\mathbf{S}_n}$ becomes much simpler. But for general $\mathbf{T}_n$, there is no explicit solution to (1.1), so we know nothing about $F^{y,H}$ except the Eq. (1.1). Thus what we can do is to investigate the properties of Eq. (1.1), or equivalently the Eqs. (2.6) and (2.7). This makes the problems more complicated.

Proof of Lemma 2.3. First, from taking the imaginary part of (2.7), we have for all $z \in \mathbb{C}^+$,

$$y\int\frac{t^2}{|1+t\underline{s}(z)|^2}dH(t) = \frac{1}{|\underline{s}(z)|^2} - \frac{v}{\Im\underline{s}(z)} \geq 0. \tag{2.8}$$

Also, (2.5) and (2.6) imply that

$$z\underline{s}(z) - y + 1 = zys(z) = -y\int\frac{1}{1+t\underline{s}(z)}dH(t).$$

It follows from Hölder's inequality and (2.8) that

$$|zys(z)| \leq \sqrt{y}\left(y\int \frac{t^2}{|1+t\underline{s}(z)|^2}dH(t)\int t^{-2}dH(t)\right)^{1/2}$$
$$\leq \frac{\sqrt{y}}{|\underline{s}(z)\lambda_0|},$$

which implies

$$|s(z)\underline{s}(z)| \leq \frac{1}{\lambda_0\sqrt{y}|z|}.$$

This leads to, for all $u \in [a, b]$,

$$\sup_{u\in[a,b]} |\underline{s}(z)| \leq C \quad \text{and} \quad \sup_{u\in[a,b]} |s(z)| \leq C.$$

Let Ψ be the support of $F^{y,H}(x)$. For any $u \in \Psi \setminus \partial\Psi$, via Lemma 6.8, we have $f^{y,H}(u) = \pi^{-1}\Im s(u)$. Then taking the real part of (2.7),

$$u = -\frac{\Re\underline{s}(u)}{|\underline{s}(u)|^2} + y\int \frac{t(1+t\Re\underline{s}(u))dH(t)}{|1+t\underline{s}(u)|^2}.$$

Using (2.8) with $v \downarrow 0$, we obtain

$$u = y\int \frac{tdH(t)}{(1+t\Re\underline{s}(u))^2 + t^2\Im\underline{s}(u)^2} \leq y\int \frac{dH(t)}{t\Im\underline{s}(u)^2}.$$

Therefore,

$$\Im\underline{s}(u) \leq \left(\frac{\int_0^\infty yt^{-1}dH(t)}{x}\right)^{1/2}.$$

If $u \in \Psi^c \cup \partial\Psi$, we have $\Im\underline{s}(u) = 0$. Therefore, by the fact that $\underline{s}(z) = -\frac{1-y}{z} + ys(z)$, the proof is complete. □

3. Truncation, centralization and rescale

We first truncate the random variables x_{ij} at $n^{3/16}$ and then centralize and rescale the variables. In the following sections, we denote $\|\mathbf{A}\|$ as the spectral norm of any matrix $\mathbf{A}$, i.e. the square root of the maximum eigenvalue of $\mathbf{A}^*\mathbf{A}$. Let us define

$$\begin{array}{lll}
\hat{x}_{ij} = x_{ij}I(|x_{ij}| \leq n^{3/16}) & \tilde{x}_{ij} = \hat{x}_{ij} - \mathbb{E}\hat{x}_{ij} & \check{x}_{ij} = \tilde{x}_{ij}/(\mathbb{E}|\hat{x}_{ij} - \mathbb{E}\hat{x}_{ij}|^2)^{1/2} \\
\hat{\mathbf{X}}_n = (\hat{x}_{ij})_{p\times n} & \tilde{\mathbf{X}}_n = (\tilde{x}_{ij})_{p\times n} & \check{\mathbf{X}}_n = (\check{x}_{ij})_{p\times n} \\
\hat{\mathbf{S}}_n = \frac{1}{n}\mathbf{T}_n^{\frac{1}{2}}\hat{\mathbf{X}}_n\hat{\mathbf{X}}_n^*\mathbf{T}_n^{\frac{1}{2}} & \tilde{\mathbf{S}}_n = \frac{1}{n}\mathbf{T}_n^{\frac{1}{2}}\tilde{\mathbf{X}}_n\tilde{\mathbf{X}}_n^*\mathbf{T}_n^{\frac{1}{2}} & \check{\mathbf{S}}_n = \frac{1}{n}\mathbf{T}_n^{\frac{1}{2}}\check{\mathbf{X}}_n\check{\mathbf{X}}_n^*\mathbf{T}_n^{\frac{1}{2}}.
\end{array}$$

In this section, we will show that under the assumptions in Theorem 1.1, we have almost surely

$$\sup_x |F^{\mathbf{S}_n}(x) - F^{y_n,H_n}(x)| \leq C\max\left\{\sup_x |F^{\check{\mathbf{S}}_n}(x) - F^{y_n,H_n}(x)|, n^{-1/2}\right\}. \tag{3.1}$$

This means that we can add the condition $|x_{ij}| \leq n^{3/16}$ for all i, j in Theorems 1.1 and 1.3.

From [8], we know that the support of $F^{y,H}(x)$ is a subset of $[\lambda_0(1-\sqrt{y})^2, (1+\sqrt{y})^2]$. Thus applying Lemma 2.3, we get

$$\begin{aligned}\sup_x |F^{y_n,H_n}(x+h) - F^{y_n,H_n}(x)| &= \sup_x \left| \int_x^{x+h} f^{y_n,H_n}(t)dt \right| \\ &\le \sup_x \frac{h}{2\pi\sqrt{\lambda_1^{\mathbf{T}_n} y_n}\left(\sqrt{x+h}+\sqrt{x}\right)} \\ &\le \frac{\sqrt{h}}{\sqrt{y_n \lambda_1^{\mathbf{T}_n}}}.\end{aligned}$$

This shows that F^{y_n,H_n} satisfies the condition in Lemma 6.3. Thus to prove (3.1), one has to prove the following lemma.

Lemma 3.1. *Under the same assumptions of Theorem 1.1, we have*

I: $\sup_x |F^{\mathbf{S}_n}(x) - F^{\hat{\mathbf{S}}_n}(x)| = O(n^{-1/2})$ *a.s.*
II: $L(F^{\hat{\mathbf{S}}_n} - F^{\check{\mathbf{S}}_n}) = o(n^{-1/2})$ *a.s.*

where $L(\cdot,\cdot)$ denotes the Lévy distance.

Proof. (I) By Lemma 6.1, we have

$$\sup_x |F^{\mathbf{S}_n}(x) - F^{\hat{\mathbf{S}}_n}(x)| \le \frac{1}{p}\operatorname{rank}\left(\frac{1}{\sqrt{n}}\mathbf{T}_n^{1/2}(\mathbf{X}_n - \hat{\mathbf{X}}_n)\right) \le \frac{1}{p}\sum_{ij} I(|x_{ij}| > n^{3/16}).$$

Under the assumption of Theorem 1.1, we have

$$\mathbb{E}\left(\frac{1}{p}\sum_{ij} I(|x_{ij}| > n^{3/16})\right) \le \frac{pn \max_{ij} \mathbb{E}|x_{ij}|^8}{pn^{3/2}} = O(n^{-1/2}),$$

and

$$\mathbb{V}ar\left(\frac{1}{p}\sum_{ij} I(|x_{ij}| > n^{3/16})\right) = O(n^{-3/2}).$$

For any $\varepsilon > n^{1/2}p^{-1}\sum_{ij}\mathbb{P}(|x_{ij}| > n^{3/16})$ and by the Bernstein inequality, we have

$$\mathbb{P}\left(\frac{1}{p}\sum_{ij} I(|x_{ij}| > n^{3/16}) > n^{-1/2}\varepsilon\right) \le 2e^{-Cn^{1/2}},$$

which is summable. Therefore from the Borel–Cantelli lemma, we get (I).

(II) Let $\mathbf{M}_n = \left(1 - \sigma_{ij}^{-1}\right)_{p\times n}$, with $\sigma_{ij}^2 = \mathbb{V}ar(\hat{x}_{ij})$ and $\circ$ be the Hadamard product of matrices. Now let us estimate the Lévy distance. By Lemma 6.2, we obtain that

$$L(F^{\hat{\mathbf{S}}_n}, F^{\tilde{\mathbf{S}}_n}) \le 2\left\|\frac{1}{\sqrt{n}}\mathbf{T}_n^{1/2}\hat{\mathbf{X}}_n\right\| \left\|\frac{1}{\sqrt{n}}\mathbf{T}_n^{1/2}\mathbb{E}(\hat{\mathbf{X}}_n)\right\| + \left\|\frac{1}{\sqrt{n}}\mathbf{T}_n^{1/2}\mathbb{E}(\hat{\mathbf{X}}_n)\right\|^2$$

$$L(F^{\check{\mathbf{S}}_n}, F^{\tilde{\mathbf{S}}_n}) \le 2\left\|\frac{1}{\sqrt{n}}\mathbf{T}_n^{1/2}\tilde{\mathbf{X}}_n\right\| \left\|\frac{1}{\sqrt{n}}\mathbf{T}_n^{1/2}(\tilde{\mathbf{X}}_n \circ \mathbf{M}_n)\right\| + \left\|\frac{1}{\sqrt{n}}\mathbf{T}_n^{1/2}(\tilde{\mathbf{X}}_n \circ \mathbf{M}_n)\right\|^2.$$

From Theorem 1.1 in [8] and the condition $\|\mathbf{T}_n\| \le 1$, we have

$$\limsup \left\| \frac{1}{\sqrt{n}} \mathbf{T}_n^{1/2} \hat{\mathbf{X}} \right\| = \limsup \left\| \frac{1}{\sqrt{n}} \mathbf{T}_n^{1/2} \tilde{\mathbf{X}}_n \right\| \le 1 + \sqrt{y} \quad \text{a.s.}$$

In addition, by the assumption that $\mathbb{E}x_{ij} = 0$ and $\sup_n \sup_{i,j} \mathbb{E}|x_{ij}^8| < \infty$, we have

$$\left\| \frac{1}{\sqrt{n}} \mathbf{T}_n^{1/2} \mathbb{E}\hat{\mathbf{X}}_n \right\| \le \sqrt{n} \max_{i,j} \mathbb{E}|x_{i,j}| I(|x_{ij}| > n^{3/16}) = O(n^{-13/16}),$$

and

$$\begin{aligned} |\sigma_{ij} - 1| &\le |\sigma_{ij}^2 - 1| \\ &\le \mathbb{E}\left(|x_{ij}|^2 I(|x_{ij}| > n^{3/16})\right) + \left(\mathbb{E}x_{ij} I(|x_{ij}| \le n^{3/16})\right)^2 \\ &= O(n^{-9/8}). \end{aligned}$$

Thus, we have

$$\begin{aligned} \left\| \frac{1}{\sqrt{n}} \mathbf{T}_n^{1/2} (\tilde{\mathbf{X}}_n \circ \mathbf{M}_n) \right\|^2 &\le \frac{1}{n} \sum_{i,j} |\tilde{x}_{ij}|^2 (1 - \sigma_{ij}^{-1})^2 \\ &\le \frac{1}{n} \sum_{i,j} |\tilde{x}_{ij}|^2 \max_{i,j} \left(\frac{1 - \sigma_{ij}}{\sigma_{ij}} \right)^2 = o(n^{-1}) \quad \text{a.s.} \end{aligned}$$

From the above arguments, we can complete the proof. □

Remark 3.2. In proofs of theorems, we may assume that the entries of $\mathbf{X}_n$ are truncated at $n^{3/16}$, recentralized and rescaled. For brevity, we still use x_{ij} to denote the truncated and normalized variables $\check{x}_{ij}$ in the sequel.

4. Proof of Theorem 1.1

In [20], which is the first paper investigating the convergence rates of spectral distributions of $\mathbf{S}_n$ with $\mathbf{T}_n$ not being the identity matrix, Jing et al. prove that $\Delta = O(n^{-2/5})$. However, this order cannot be improved by the mathematical technique used in [20]. Therefore, in this section, we introduce another technique. The main tools used in the proof are properties of the Stieltjes transform and bounds on the moments of martingale difference sequences. Before proceeding, we introduce some notation.

Throughout this section, we use a and b to denote two constants such that $0 < a < \lambda_0(1 - \sqrt{y_n})^2$, $b > (1 + \sqrt{y_n})^2$. Let $z = u + iv$, $u \in [a, b]$, $1 \ge v \ge v_0 = \max\{\vartheta\Delta, C_0 n^{-1/2}\}$ with $0 < \vartheta < 1$ and C_0 be an appropriate constant. Both ϑ and C_0 will be specified later.

From Proposition 2.1, we know that the essential portion of the proof of Theorem 1.1 is to obtain the upper bound for $|\mathbb{E}s_n - s_0|$. The derivation of such a bound is one of the main technical works of this paper. In order to make the presentation easily followed, we start with some basic facts.

From the definition of the Stieltjes transform we have

$$s_n(z) = \frac{1}{p} \sum_{j=1}^{p} \frac{1}{\lambda_j^{\mathbf{S}_n} - z} = \frac{1}{p} \mathrm{tr}(\mathbf{S}_n - z\mathbf{I})^{-1}.$$

Write

$$\mathbf{S}_n - z\mathbf{I} + z\mathbf{I} = \sum_{i=j}^{n} \mathbf{r}_j \mathbf{r}_j^*.$$

Multiplying the inverse of $\mathbf{S}_n - z\mathbf{I}$ on both sides (refer to the notation in (2.3)), we have

$$\mathbf{I} + z\mathbf{D}^{-1} = \sum_{i=j}^{n} \beta_j \mathbf{r}_j \mathbf{r}_j^* \mathbf{D}_j^{-1}. \tag{4.1}$$

Taking the trace on both sides and dividing by n, we get

$$y_n + z y_n s_n = \frac{1}{n} \sum_{j=1}^{n} (1 - \beta_j),$$

which together with (2.4) implies

$$\underline{s}_n = -\frac{1}{zn} \sum_{j=1}^{n} \beta_j. \tag{4.2}$$

If we can prove that $\mathbb{E}\beta_j = g(z, \mathbb{E}\underline{s}_n(z)) + o(1)$ for some function g, then (4.2) gives (intuitively)

$$z\mathbb{E}\underline{s}_n = -g(z, \mathbb{E}\underline{s}_n(z)) + o(1).$$

So the LSDF $f^{y,H}$ exists and its Stieltjes transform is the solution to the equation

$$z\underline{s} = -g(z, \underline{s}).$$

This is the classical method to deal with the problems about LSDF of random matrices. Now we show how to find the function g rigorously in our case.

Let $\omega_n = \omega_n(z) = \frac{1}{p}\mathrm{tr}(-z\mathbb{E}\underline{s}_n \mathbf{T}_n - z\mathbf{I})^{-1} - \mathbb{E}s_n$. Later we will show that ω_n tends to 0 fast enough for our purpose. Using (2.4), we can check

$$\begin{aligned}
\omega_n &= -\frac{1}{y_n}\left(\frac{y_n}{z}\int \frac{dH_n(t)}{t\mathbb{E}\underline{s}_n + 1} + \mathbb{E}\underline{s}_n + \frac{1 - y_n}{z}\right) \\
&= -\frac{\mathbb{E}\underline{s}_n}{z y_n}\left(\frac{y_n}{\mathbb{E}\underline{s}_n}\int \frac{dH_n(t)}{t\mathbb{E}\underline{s}_n + 1} + z + \frac{1 - y_n}{\mathbb{E}\underline{s}_n}\right) \\
&= \frac{\mathbb{E}\underline{s}_n}{z y_n}\left(-z - \frac{1}{\mathbb{E}\underline{s}_n} + y_n \int \frac{t\,dH_n(t)}{t\mathbb{E}\underline{s}_n + 1}\right).
\end{aligned} \tag{4.3}$$

Denote $R_n = -z - \frac{1}{\mathbb{E}\underline{s}_n} + y_n \int \frac{t\,dH_n(t)}{t\mathbb{E}\underline{s}_n + 1}$. Then we have $R_n = z y_n \omega_n / \mathbb{E}\underline{s}_n$ and

$$\mathbb{E}\underline{s}_n = \frac{1}{-z + y_n \int \frac{t\,dH_n(t)}{t\mathbb{E}\underline{s}_n + 1} - R_n}.$$

Combining (2.5) and (2.6), we obtain

$$\underline{s}_0 = \frac{1}{-z + y_n \int \frac{t\,dH_n(t)}{t\underline{s}_0 + 1}}.$$

Thus we get

$$\mathbb{E}\underline{s}_n - \underline{s}_0 = \frac{(\mathbb{E}\underline{s}_n - \underline{s}_0) y_n \int \frac{t^2 dH_n(t)}{(t\underline{s}_0+1)(t\mathbb{E}\underline{s}_n+1)}}{\left(-z + y_n \int \frac{t dH_n(t)}{t\mathbb{E}\underline{s}_n+1} - R_n\right)\left(-z + y_n \int \frac{t dH_n(t)}{t\underline{s}_0+1}\right)} + \underline{s}_n \underline{s}_0 R_n,$$

which implies

$$\mathbb{E}\underline{s}_n - \underline{s}_0 = \underline{s}_0 y_n z \omega_n \left(1 - \frac{y_n \int \frac{t^2 dH_n(t)}{(1+t\mathbb{E}\underline{s}_n)(1+t\underline{s}_0)}}{\left(-z + y_n \int \frac{t dH_n(t)}{(1+t\mathbb{E}\underline{s}_n)} - R_n\right)\left(-z + y_n \int \frac{t dH_n(t)}{1+t\underline{s}_0}\right)}\right)^{-1}. \tag{4.4}$$

Next, we will bound the right side of (4.4). Noting (2.4) and (2.5), we are actually working on $|\mathbb{E}s_n - s_0|$. Here we need the following lemma which will be proved in Section 4.3.

Lemma 4.1. *Under the same assumptions of Theorem 1.1, we have for any $u \in [a, b]$ and $v \geq C_0 n^{-1/2}$,*

$$|\omega_n| = O(n^{-1}v^{-1}).$$

From integration by parts, we have

$$\begin{aligned} |\mathbb{E}s_n(z) - s_0(z)| &= \left| \int_0^\infty \frac{d(\mathbb{E}F^{\mathbf{S}_n}(x) - F^{y_n,H_n}(x))}{x - z} \right| \\ &= \left| \int_0^\infty \frac{\mathbb{E}F^{\mathbf{S}_n}(x) - F^{y_n,H_n}(x)}{(x-z)^2} dx \right| \leq \frac{\pi \Delta}{v} \leq \frac{\pi}{\vartheta}. \end{aligned}$$

By Lemma 2.3, we have

$$\sup_{u\in[a,b]} |s_0(z)| \leq C \quad \text{and} \quad \sup_{u\in[a,b]} |\underline{s}_0(z)| \leq C.$$

Then together with (2.4), if $u \geq a$, we have

$$|\mathbb{E}s_n(z)| \leq C, \qquad |\mathbb{E}\underline{s}_n(z)| \leq C. \tag{4.5}$$

In view of (2.4) and (4.3), we have

$$-zy_n\omega_n = y_n \int \frac{1}{t\mathbb{E}\underline{s}_n(z) + 1} dH_n(t) + z\mathbb{E}\underline{s}_n(z) - y_n + 1,$$

which via Lemma 4.1 implies that

$$\inf_{n,z} |\mathbb{E}\underline{s}_n| > C > 0.$$

Thus we have for $v > C_0 n^{-1/2}$ with appropriate C_0,

$$|R_n| = |y_n z \omega_n (\mathbb{E}\underline{s}_n(z))^{-1}| \leq \frac{C}{nv} < v. \tag{4.6}$$

Then by $|R_n| \le v$ and the Cauchy–Schwarz inequality, we get

$$\left|\frac{y_n \int \frac{t^2 dH_n(t)}{(1+t\mathbb{E}\underline{s}_n)(1+t\underline{s}_0)}}{\left(-z + y_n \int \frac{t dH_n(t)}{(1+t\mathbb{E}\underline{s}_n)} - R_n\right)\left(-z + y_n \int \frac{t dH_n(t)}{1+t\underline{s}_0}\right)}\right|$$
$$\le \left(\frac{y_n \int \frac{t^2 dH_n(t)}{|1+t\mathbb{E}\underline{s}_n|^2}}{\left|-z + y_n \int \frac{t dH_n(t)}{(1+t\mathbb{E}\underline{s}_n)} - R_n\right|^2}\right)^{1/2} \left(\frac{y_n \int \frac{t^2 dH_n(t)}{|1+t\underline{s}_0|^2}}{\left|-z + y_n \int \frac{t dH_n(t)}{1+t\underline{s}_0}\right|^2}\right)^{1/2}$$
$$= \left(\frac{y_n \Im\mathbb{E}\underline{s}_n \int \frac{t^2 dH_n(t)}{|1+t\mathbb{E}\underline{s}_n|^2}}{v + \Im R_n + \Im\mathbb{E}\underline{s}_n y_n \int \frac{t^2 dH_n(t)}{|1+t\mathbb{E}\underline{s}_n|^2}}\right)^{1/2} \left(\frac{y_n \Im\underline{s}_0 \int \frac{t^2 dH_n(t)}{|1+t\underline{s}_0|^2}}{v + \Im\underline{s}_0 y \int \frac{t^2 dH(t)}{|1+t\underline{s}_0|^2}}\right)^{1/2}$$
$$\le \left(\frac{y_n \Im\underline{s}_0 \int \frac{t^2 dH_n(t)}{|1+t\underline{s}_0|^2}}{v + \Im\underline{s}_0 y \int \frac{t^2 dH(t)}{|1+t\underline{s}_0|^2}}\right)^{1/2}. \tag{4.7}$$

Applying $\sqrt{1-a} \le 1 - \frac{1}{2}a$ for $a \le 1$, we have

$$\left(\frac{\Im\underline{s}_0 y_n \int \frac{t^2 dH(t)}{|1+t\underline{s}_0|^2}}{v + \Im\underline{s}_0 y_n \int \frac{t^2 dH(t)}{|1+t\underline{s}_0|^2}}\right)^{\frac{1}{2}} \le 1 - \frac{1}{2}\frac{v}{v + \Im\underline{s}_0 y_n \int \frac{t^2 dH(t)}{|1+t\underline{s}_0|^2}}, \tag{4.8}$$

which together with Lemma 2.3 and (4.4) implies

$$|\mathbb{E}\underline{s}_n - \underline{s}_0| \le \frac{C}{nv^2}\left|\Im\underline{s}_0 y_n \int \frac{t^2 dH(t)}{|1+t\underline{s}_0|^2}\right|.$$

Noting that (2.8) also holds when $\underline{s}$ is replaced by $\underline{s}_0$, we obtain that

$$y_n \Im\underline{s}_0 \int \frac{t^2 dH(t)}{|1+t\underline{s}_0|^2} = \frac{\Im\underline{s}_0}{|\underline{s}_0|^2} - v < C. \tag{4.9}$$

It follows from (2.4) and (2.5) that,

$$|\mathbb{E}s_n - s_0| = \frac{1}{y_n}|\mathbb{E}\underline{s}_n - \underline{s}_0| \le \frac{C}{nv^2}, \tag{4.10}$$

which implies that the second integral in (2.1) has the order $O(n^{-1}v^{-1})$.

Moreover, for any constant V, if $v = V$, then from (4.37) and (4.6), we know that

$$|\omega_n| = O\left(\frac{1}{n}\right) \quad \text{and} \quad |R_n| = O\left(\frac{1}{n}\right).$$

Therefore, it follows that

$$|\mathbb{E}\underline{s}_n - \underline{s}_0| \le \frac{C}{n}|\mathbb{E}\underline{s}_n \underline{s}_0| = \frac{C}{n}\mathbb{E}|\underline{s}_n^2| + \frac{C}{n}|\mathbb{E}\underline{s}_n||\mathbb{E}\underline{s}_n - \underline{s}_0|.$$

This implies, for large n,

$$|\mathbb{E}s_n - s_0| = \frac{1}{y_n}|\mathbb{E}\underline{s}_n - \underline{s}_0| \leq \frac{C}{ny_n}|\mathbb{E}\underline{s}_n^2|.$$

A direct calculation shows that

$$\int_{-\infty}^{\infty} |\mathbb{E}s_n(u+iV) - s_0(u+iV)|du$$
$$\leq \frac{C}{n}\mathbb{E}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{(x-u)^2+V^2} dud\mathbb{E}F^{\underline{\mathbf{S}}_n}(x) = O\left(\frac{1}{n}\right). \tag{4.11}$$

Therefore, we conclude from Proposition 2.1, (4.10) and (4.11) that for $0 < \theta \leq y_n \leq \Theta < 1, u \in [a,b], 1 \geq v \geq v_0 = \max\{\vartheta\Delta, C_0 n^{-1/2}\}$,

$$\Delta \leq C_1 v + \frac{C_2}{nv} + \frac{C_3}{n}. \tag{4.12}$$

Here we note that neither C_1 nor C_3 depends on ϑ. If $v_0 = C_0 n^{-1/2}$, then we have $\Delta \leq \vartheta^{-1} C_0 n^{-1/2}$. If $v \geq v_0 = \vartheta\Delta$, then choose $\vartheta = (2C_1)^{-1}$, from (4.12), we also have

$$\Delta = O(n^{-1/2}).$$

Therefore, it remains to prove Lemma 4.1.

The rest of this section is organized as follows. In Sections 4.1 and 4.2, we will introduce two important conclusions, Lemma 4.2 and (4.32), which will be used in the rest of the paper. The proof of Lemma 4.1 will be provided in Section 4.3, which is an improvement over the same part in [20]. Since the conditions in Theorem 1.1 are much weaker than those in [20], we have to make sure that many conclusions in [20] are also correct under our situations.

4.1. A bound for $\mathbb{E}|\mathbb{E}s_n - s_n|^{2l}$

In this section, we will use the martingale decomposition method to get a bound for $\mathbb{E}\,|s_n - \mathbb{E}s_n|^{2l}$. This method, devised by Girko in [15,16], is widely used in random matrix theory.

Lemma 4.2. *If $|b_i| \leq C$, $\sup_n \sup_{i,j} \mathbb{E}|x_{ij}|^8 < \infty$, $u \in [a,b]$ and $v \geq C_0 n^{-1/2}$, then there exist positive constants C_l depending on l, such that for all n and any $l \geq 1$.*

$$\mathbb{E}\left|\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} - \mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1}\right|^{2l} \leq \frac{C_l}{n^{2l}v^{3l}}. \tag{4.13}$$

Proof. Let $\mathbb{E}_0(\cdot)$ denote the expectation and $\mathbb{E}_k(\cdot)$ denote the conditional expectation with respect to the σ-field generated by $\mathbf{r}_1, \ldots, \mathbf{r}_k$. We have

$$\begin{aligned}
\mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} - \frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} &= -\frac{1}{p}\sum_{k=1}^{n}(\mathbb{E}_k\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} - \mathbb{E}_{k-1}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1}) \\
&= \frac{1}{p}\sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\left(\frac{\mathbf{r}_k^*\mathbf{D}_k^{-1}\mathbf{T}_n\mathbf{D}_k^{-1}\mathbf{r}_k}{1+\mathbf{r}_k^*\mathbf{D}_k^{-1}\mathbf{r}_k}\right) \\
&= \frac{1}{p}\sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\left(\alpha_k b_k - a_k\beta_k b_k\gamma_k\right).
\end{aligned}$$

Note that $\{(\mathbb{E}_k - \mathbb{E}_{k-1})\alpha_k b_k\}$ and $\{(\mathbb{E}_k - \mathbb{E}_{k-1})a_k\beta_k b_k\gamma_k\}$ form two sequences of bounded martingale differences respectively. Applying Lemma 6.9, it follows that for $l \geq 1$,

$$\mathbb{E}\left|\frac{1}{p}\sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\, b_k\alpha_k\right|^{2l} \leq C_l p^{-2l}\left(\mathbb{E}\left(\sum_{k=1}^{n}\mathbb{E}_{k-1}|b_k\alpha_k|^2\right)^l + \sum_{k=1}^{n}\mathbb{E}|b_k\alpha_k|^{2l}\right).$$

It follows from Lemma 6.7 that

$$|\beta_j a_j| = |\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} - \mathrm{tr}\mathbf{T}_n\mathbf{D}_j^{-1}| \leq v^{-1}. \tag{4.14}$$

From (4.14), one can check directly that

$$\begin{aligned}\mathrm{tr}\mathbf{T}_n\mathbf{D}_k^{-1}(\mathbf{T}_n\mathbf{D}_k^{-1})^* &= \frac{1}{v}\Im\left(\mathrm{tr}\mathbf{T}_n^2(\mathbf{D}_k - \mathbf{D}) + \mathrm{tr}\mathbf{T}_n^2\mathbf{D}\right)\\ &\leq \frac{1}{v}\left(\frac{1}{v} + \Im\mathrm{tr}\mathbf{T}_n\mathbf{D}\right).\end{aligned} \tag{4.15}$$

It follows from Lemma 6.5 that

$$\begin{aligned}\mathbb{E}_{k-1}|b_k\alpha_k|^2 &\leq \mathbb{E}_{k-1}\frac{C}{n^2}\mathrm{tr}(\mathbf{T}_n\mathbf{D}_k^{-1})^2(\mathbf{T}_n\mathbf{D}_k^{-1})^{2*}\\ &\leq \mathbb{E}_{k-1}\frac{C}{n^2v^2}\mathrm{tr}\mathbf{T}_n\mathbf{D}_k^{-1}(\mathbf{T}_n\mathbf{D}_k^{-1})^* \leq \mathbb{E}_{k-1}\frac{C}{n^2v^3}\left(\frac{1}{v} + \Im\mathrm{tr}\mathbf{T}_n\mathbf{D}\right).\end{aligned}$$

Furthermore, by Lemma 6.5 and the fact that

$$\nu_{2l} = \sup_n\sup_{i,j}\mathbb{E}|x_{ij}|^{2l} = \begin{cases}O(1), & l \leq 4,\\ O(n^{3l/8-3/2}), & l > 4,\end{cases}$$

for $l \geq 1$, we have

$$\begin{aligned}\mathbb{E}|\alpha_k|^{2l} &\leq \frac{C_l}{n^{2l}}\mathbb{E}\left(\left(\nu_4\mathrm{tr}(\mathbf{T}_n\mathbf{D}_k^{-1})^2(\mathbf{T}_n\mathbf{D}_k^{-1})^{2*}\right)^l + \nu_{4l}\mathrm{tr}\left((\mathbf{T}_n\mathbf{D}_k^{-1})^2(\mathbf{T}_n\mathbf{D}_k^{-1})^{2*}\right)^l\right)\\ &\leq \frac{C}{n^{2l}}\left(\mathbb{E}\left(v^{-2}\mathrm{tr}\mathbf{T}_n\mathbf{D}_k^{-1}(\mathbf{T}_n\mathbf{D}_k^{-1})^*\right)^l + \nu_{4l}v^{4l-2}\mathbb{E}\mathrm{tr}\mathbf{D}_k^{-1}(\mathbf{D}_k^{-1})^*\right)\\ &\leq \frac{C}{n^{2l}v^{3l}}\left(\frac{1}{v^l} + (\Im\mathbb{E}\mathrm{tr}\mathbf{T}_n\mathbf{D})^l\right) + \frac{\nu_{4l}}{n^{2l-1}v^{4l-1}},\end{aligned}$$

and,

$$\mathbb{E}|\alpha_k|^2 \leq \frac{C}{n^2v^3}\left(\frac{1}{v} + \Im\mathbb{E}\mathrm{tr}\mathbf{T}_n\mathbf{D}\right).$$

Therefore we have

$$\mathbb{E}\left|\frac{1}{p}\sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\, b_k\alpha_k\right|^{2l} \leq \frac{C_l}{n^{2l}v^{3l}}\left(1 + \mathbb{E}\left(\Im\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}\right)^l\right). \tag{4.16}$$

Similarly, using (4.14)

$$\mathbb{E}\left|\frac{1}{p}\sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\, a_k\beta_k b_k\gamma_k\right|^{2l} = C_l p^{-2l}v^{-2l}\left(\mathbb{E}\left(\sum_{k=1}^{n}\mathbb{E}_{k-1}|\gamma_k|^2\right)^l + \sum_{k=1}^{n}\mathbb{E}|\gamma_k|^{2l}\right).$$

For $l \geq 1$,

$$\mathbb{E}|\hat{\gamma}_k|^{2l} \leq \frac{C}{n^{2l}v^l}\left(\frac{1}{v^l} + (\Im\mathbb{E}\mathrm{tr}\mathbf{T}_n\mathbf{D})^l\right) + \frac{\nu_{4l}}{n^{2l-1}v^{2l-1}}, \tag{4.17}$$

and

$$\mathbb{E}|\hat{\gamma}_k|^2 \leq \frac{C}{n^2 v}\left(\frac{1}{v} + \Im\mathbb{E}\mathrm{tr}\mathbf{T}_n\mathbf{D}\right). \tag{4.18}$$

For any $l \geq 1$, by Lemma 6.7

$$\mathbb{E}|\gamma_j - \hat{\gamma}_j|^l \leq \frac{C_l}{n^l v^l} + C_l\mathbb{E}\left|\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} - \mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1}\right|^l.$$

So,

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{p}\sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\, a_k\beta_k b_k\gamma_k\right|^{2l} \\
&\leq \frac{C_l}{n^{2l}v^{3l}}\left(1 + \mathbb{E}\left(\Im\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}\right)^l\right) + \frac{C_l}{n^l v^{2l}}\mathbb{E}\left|\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} - \mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1}\right|^{2l}.
\end{aligned} \tag{4.19}$$

Combining (4.16), (4.19) and selecting C_0 such that $\frac{C_l}{n^l v^{2l}} \leq \frac{1}{2}$, we obtain that

$$\mathbb{E}\left|\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} - \mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1}\right|^{2l} \leq \frac{4C_l}{n^{2l}v^{3l}}\left(1 + \mathbb{E}\left(\Im\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}\right)^l\right). \tag{4.20}$$

We will use the inductive method to complete the proof. When $l = 1$, we have

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} - \mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1}\right|^2 \\
&\leq \frac{C}{n^2 v^3}\left(1 + \mathbb{E}\left(\Im\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}\right)\right) \leq \frac{C}{n^2v^3}(1 + |\mathbb{E}s_n(z)|) \leq \frac{C}{n^2v^3}.
\end{aligned}$$

This shows that the lemma holds for $l = 1$. When $l \in (1, 2]$, from (4.20) we have

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} - \mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1}\right|^{2l} \\
&\leq \frac{C_l}{n^{2l}v^{3l}}\left(1 + \left(\mathbb{E}\left|\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} - \mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1}\right|^2\right)^{l/2} + |\mathbb{E}s_n|^l\right) \leq \frac{C}{n^{2l}v^{3l}}.
\end{aligned}$$

Assume that the lemma holds for $l \in (2^t, 2^{t+1}]$. Consider the case where $l \in (2^{t+1}, 2^{t+2}]$

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} - \mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1}\right|^{2l} \\
&\leq \frac{C_l}{n^{2l}v^{3l}}\left(1 + \mathbb{E}\left|\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} - \mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1}\right|^l + |\mathbb{E}s_n|^l\right) \leq \frac{C}{n^{2l}v^{3l}},
\end{aligned}$$

which completes the proof of the lemma. □

Lemma 4.3. *Under the assumptions of Lemma 4.2, for any $l \geq 1$, we have*

$$\mathbb{E}\,|s_n - \mathbb{E}s_n|^{2l} \leq \frac{C_l}{n^{2l}v^{3l}}. \tag{4.21}$$

Proof. Lemma 4.3 is obtained by repeating the argument of Lemma 4.2 and using

$$\begin{aligned}\mathbb{E}|\gamma_j - \hat{\gamma}_j|^{2l} &\leq \frac{C_l}{n^{2l}v^{2l}} + C_l\mathbb{E}\left|\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1} - \mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{T}_n\mathbf{D}^{-1}\right|^{2l}\\ &\leq \frac{C_l}{n^{2l}v^{2l}} + \frac{C_l}{n^{2l}v^{3l}}. \quad \square\end{aligned}$$

Remark 4.4. From Lemmas 4.2 and 4.3, we know that for any $l \leq 4$,

$$\mathbb{E}|\gamma_j|^{2l} = O(n^{-l}v^{-l}), \tag{4.22}$$

and for any $v > C_0 n^{-1/2}$,

$$\mathbb{E}|s_n(z)| \leq \mathbb{E}|s_n(z) - \mathbb{E}s_n(z)| + |\mathbb{E}s_n(z)| \leq C. \tag{4.23}$$

4.2. The proof that $|b_j|$ is bounded

Let $b_0 = (1 + n^{-1}\mathbb{E}\mathrm{tr}\mathbf{D}_n^{-1}\mathbf{T}_n)^{-1}$. By the fact that $\Im(n^{-1}\mathrm{tr}\mathbf{D}_j^{-1}) > 0$ and $\Im(\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j) > 0$, we have

$$|\beta_j| \leq |z|v^{-1} \quad \text{and} \quad |b_j| \leq |z|v^{-1}. \tag{4.24}$$

Using Lemma 6.7 and (4.24), we have

$$|b_0 - b_j| \leq \frac{1}{nv}|b_0 b_j| \leq \frac{|z|}{nv^2}|b_0|, \tag{4.25}$$

which implies,

$$|b_j| \leq \left(\frac{|z|}{nv^2} + 1\right)|b_0|.$$

Therefore if $v > C_0 n^{-1/2}$ with appropriate C_0, then we have

$$|b_j| \leq C|b_0|.$$

Recalling that $-z\underline{s}_n = \frac{1}{n}\sum_{j=1}^n \beta_j$ and $z\underline{s}_n(z) = y_n - 1 + zy_n s_n(z)$, we get

$$b_0 = 1 - y_n - zy_n\mathbb{E}s_n(z) + \kappa_n \tag{4.26}$$

where $\kappa_n = \frac{1}{n}\sum_{j=1}^n (b_0 - \mathbb{E}\beta_j)$.

Lemma 4.5. *If $\Im(z + \kappa_n) \geq 0$, then there exists a positive constant C such that*

$$|b_0| \leq C.$$

Proof. Consider the case $\Im(\mathbb{E}s_n(z)) \geq v > 0$ first. It follows from (4.26) and the assumption that

$$\Im(y_n - 1 + zy_n\mathbb{E}s_n(z) + z) \geq -\Im(b_0)$$
$$= -|b_0|\Im(1 + n^{-1}\mathbb{E}\overline{\mathrm{tr}\mathbf{D}^{-1}(z)\mathbf{T}_n}) \geq |b_0|^2\lambda_1^{\mathbf{T}_n} y_n\Im(\mathbb{E}s_n).$$

Note that

$$\Im(y_n - 1 + zy_n\mathbb{E}s_n(z) + z) = v + vy_n\Re(\mathbb{E}s_n(z)) + uy_n\Im(\mathbb{E}s_n(z)).$$

Thus we have

$$|b_0|^2 \leq \frac{v + vy_n\Re(\mathbb{E}s_n(z)) + uy_n\Im(\mathbb{E}s_n(z))}{\lambda_1^{\mathbf{T}_n}\Im(\mathbb{E}s_n)}$$
$$\leq \frac{(1 + y_n|\mathbb{E}s_n(z)| + uy_n)\Im(\mathbb{E}s_n(z))}{\lambda_1^{\mathbf{T}_n} y_n\Im(\mathbb{E}s_n)} \leq C. \tag{4.27}$$

Consider the case $\Im(\mathbb{E}s_n(z)) \leq v$ next. Note that for $u \in [a, b]$,

$$\Im(\mathbb{E}s_n(z)) \geq \frac{v}{C + v^2}. \tag{4.28}$$

From (4.27) we have

$$|b_0|^2 \leq \frac{(C + v^2)(1 + y_n|\mathbb{E}s_n(z)| + uy_n)v}{y_n v} \leq C,$$

which completes the proof. □

Next, we aim at proving that for any $v \geq C_0 n^{-1/2}$, we have $\Im(z+\kappa_n) > 0$. First, if $\Im(z+\kappa_n) = 0$, then we have $|\kappa_n| \geq \Im(\kappa_n) = v$ and $|b_0| < C$. Remembering the notation (2.3), we have

$$\beta_j = b_j - b_j^2\gamma_j + \beta_j b_j^2\gamma_j^2. \tag{4.29}$$

Then from (4.22) and (4.25), we get

$$|\kappa_n| = \frac{1}{n}\left|\sum_{j=1}^{n}\mathbb{E}\left((b_j - b_0) + b_j^3\gamma_j^2 - b_j^4\gamma_j^3 + b_j^4\beta_j\gamma_j^4\right)\right| \leq \frac{C}{nv}. \tag{4.30}$$

Thus, choosing an appropriate $C_0 > (2C)^{1/4}$, if $v > C_0 n^{-1/2}$, then

$$|\kappa_n| \leq \frac{v}{2},$$

which contradicts with $|\kappa_n| \geq v$. Therefore we have

$$\Im(z + \kappa_n) \neq 0. \tag{4.31}$$

When taking $v = 1$, we have $|b_j| \leq |z|$. Then from (4.30),

$$|\kappa_n| \leq \frac{C}{n},$$

which implies that for $v = 1$ and large n,

$$\Im(z + \kappa_n) > 0.$$

Combining (4.31) and the continuity of the function, we get $\Im(z+\kappa_n) > 0$. Therefore we get for any $j = 0, 1, \ldots, n$,

$$|b_j| < C. \tag{4.32}$$

4.3. Proof of Lemma 4.1

In this part, we will proceed the proof of Lemma 4.1. If we estimate it by the method in [20], we can see that ω_n is controlled by $\mathbb{E}n^{-1}\mathrm{tr}\mathbf{D}^{-2}(\mathbf{D}^{-2})^*$. As far as we know, in the literature, the best order of $\mathbb{E}n^{-1}\mathrm{tr}\mathbf{D}^{-2}(\mathbf{D}^{-2})^*$ is $O(v^{-3})$ under the assumptions of Theorem 1.1. This forces us to use different techniques as follows.

Denote $\mathbf{K}_n = (\mathbb{E}\underline{s}_n\mathbf{T}_n + \mathbf{I})^{-1}$. By the fact that $\mathbf{T}_n$ is a positive definite Hermitian matrix, we have

$$\mathbf{T}_n = \mathbf{U}_n^* \begin{pmatrix} \lambda_1^{\mathbf{T}_n} & & \\ & \ddots & \\ & & \lambda_p^{\mathbf{T}_n} \end{pmatrix} \mathbf{U}_n,$$

where $\mathbf{U}_n^*\mathbf{U}_n = \mathbf{I}$. This implies

$$\mathbf{K}_n = \mathbf{U}_n^* \begin{pmatrix} (\mathbb{E}\underline{s}_n\lambda_1^{\mathbf{T}_n} + 1)^{-1} & & \\ & \ddots & \\ & & (\mathbb{E}\underline{s}_n\lambda_p^{\mathbf{T}_n} + 1)^{-1} \end{pmatrix} \mathbf{U}_n.$$

Let

$$\mathbf{K}_n^{1/2} = \mathbf{U}_n^* \begin{pmatrix} (\mathbb{E}\underline{s}_n\lambda_1^{\mathbf{T}_n} + 1)^{-1/2} & & \\ & \ddots & \\ & & (\mathbb{E}\underline{s}_n\lambda_p^{\mathbf{T}_n} + 1)^{-1/2} \end{pmatrix} \mathbf{U}_n.$$

Denote $\mathbf{Q} = \mathbf{K}_n^{1/2}(\mathbf{K}_n^{1/2})^*$, then we have

$$\mathbf{Q}^2 = \mathbf{K}_n\mathbf{K}_n^*.$$

Using the same method of calculating (4.1), we get that

$$(\mathbb{E}\underline{s}_n\mathbf{T}_n + \mathbf{I})^{-1} + z\mathbf{D}^{-1} = \sum_{j=1}^{n} \beta_j \left(\mathbf{K}_n\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1} - \frac{1}{n}\mathbf{K}_n\mathbf{T}_n\mathbf{D}^{-1} \right).$$

Multiplying $n^{-1}\mathbf{Q}^m$ for $m = 0, 1$ on both sides of the above equality and taking the trace, we get

$$\begin{aligned}
&\frac{1}{n}\mathrm{tr}\mathbf{Q}^m(\mathbb{E}\underline{s}_n\mathbf{T}_n + \mathbf{I})^{-1} + \frac{z}{n}\mathbb{E}\mathrm{tr}\mathbf{Q}^m\mathbf{D}^{-1} \\
&\quad = \frac{1}{n}\mathbb{E}\sum_{j=1}^{n}\beta_j\left(\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{Q}^m\mathbf{K}_n\mathbf{r}_j - \frac{1}{n}\mathrm{tr}\mathbf{Q}^m\mathbf{K}_n\mathbf{T}_n\mathbf{D}^{-1}\right) \\
&\quad = \frac{1}{n}\sum_{j=1}^{n} d_j = \frac{1}{n}\sum_{j=1}^{n}(d_{j1} - d_{j2} + d_{j3} - d_{j4} + d_{j5})
\end{aligned}$$

where

$$d_{j1} = b_j \mathbb{E}\left(\frac{1}{n}\mathrm{tr}\mathbf{Q}^m\mathbf{K}_n\mathbf{T}_n\mathbf{D}_j^{-1} - \frac{1}{n}\mathrm{tr}\mathbf{Q}^m\mathbf{K}_n\mathbf{T}_n\mathbf{D}^{-1}\right)$$

$$d_{j2} = b_j^2 \mathbb{E}\gamma_j\left(\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{Q}^m\mathbf{K}_n\mathbf{r}_j - \frac{1}{n}\mathrm{tr}\mathbf{Q}^m\mathbf{K}_n\mathbf{T}_n\mathbf{D}^{-1}\right)$$

$$d_{j3} = b_j^3 \mathbb{E}\gamma_j^2\left(\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{Q}^m\mathbf{K}_n\mathbf{r}_j - \frac{1}{n}\mathrm{tr}\mathbf{Q}^m\mathbf{K}_n\mathbf{T}_n\mathbf{D}^{-1}\right)$$

$$d_{j4} = b_j^4 \mathbb{E}\gamma_j^3\left(\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{Q}^m\mathbf{K}_n\mathbf{r}_j - \frac{1}{n}\mathrm{tr}\mathbf{Q}^m\mathbf{K}_n\mathbf{T}_n\mathbf{D}^{-1}\right)$$

$$d_{j5} = b_j^4 \mathbb{E}\beta_j\gamma_j^4\left(\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{Q}^m\mathbf{K}_n\mathbf{r}_j - \frac{1}{n}\mathrm{tr}\mathbf{Q}^m\mathbf{K}_n\mathbf{T}_n\mathbf{D}^{-1}\right).$$

Next, we will investigate the bound of d_j. From (4.14) and (4.29),

$$\begin{aligned} d_{j1} &= \frac{b_j^2}{n^2}\mathbb{E}\left(\mathrm{tr}\mathbf{Q}^m\mathbf{K}_n\mathbf{T}_n\mathbf{D}_j^{-1}\mathbf{T}_n\mathbf{D}_j^{-1}\right) \\ &\quad + \frac{1}{n}\mathbb{E}(-b_j^3\gamma_j + b_j^4\gamma_j^2 - b_j^5\gamma_j^3 + \beta_j b_j^5\gamma_j^4)\left(\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{Q}^m\mathbf{K}_n\mathbf{T}_n\mathbf{D}_j^{-1}\mathbf{r}_j\right) \\ &= d_{j1}^{(1)} + d_{j1}^{(2)}. \end{aligned}$$

Note that $\mathbf{T}_n\mathbf{K}_n = \mathbf{K}_n\mathbf{T}_n$ and $\mathbf{Q}^m\mathbf{T}_n = \mathbf{T}_n\mathbf{Q}^m$, by lemma (4.32), we can verify that

$$|d_{j1}^{(1)}| \le \frac{C}{n^2}\mathbb{E}\mathrm{tr}\mathbf{Q}^{m+1}\mathbf{D}^{-1}(\mathbf{D}^{-1})^*.$$

Let $\psi_j = \mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{Q}^m\mathbf{K}_n\mathbf{T}_n\mathbf{D}_j^{-1}\mathbf{r}_j - n^{-1}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{Q}^m\mathbf{K}_n\mathbf{T}_n\mathbf{D}_j^{-1}\mathbf{T}_n$, from (4.15), Hölder's inequality and Lemma 6.5, we get

$$\begin{aligned} \mathbb{E}|\gamma_j^l\left(\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{Q}^m\mathbf{K}_n\mathbf{T}_n\mathbf{D}_j^{-1}\mathbf{r}_j\right)| &\le \mathbb{E}|\gamma_j^l\psi_j| + n^{-1}\mathbb{E}|\gamma_j^l\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{Q}^m\mathbf{K}_n\mathbf{T}_n\mathbf{D}_j^{-1}\mathbf{T}_n| \\ &\le \left(\mathbb{E}|\gamma_j|^{2l}\right)^{1/2}\frac{C\|\mathbf{Q}^m\|}{n^{1/2}v^{3/2}}\left(\frac{1}{n}\mathrm{tr}\mathbf{Q}^2\right)^{1/2}. \end{aligned}$$

By (4.22), (4.32) and (4.24), we have

$$|d_{j1}^{(2)}| \le \frac{C\|\mathbf{Q}^m\|}{n^2v^3}\left(\frac{1}{n}\mathrm{tr}\mathbf{Q}^2\right)^{1/2}.$$

Therefore, we conclude that for $v > C_0n^{-1/2}$,

$$|d_{j1}| \le \frac{C}{n^2}\mathrm{tr}\mathbf{Q}^{m+1}\mathbf{D}^{-1}(\mathbf{D}^{-1})^* + \frac{C\|\mathbf{Q}^m\|}{nv}\left(\frac{1}{n}\mathrm{tr}\mathbf{Q}^2\right)^{1/2}.$$

Using $\gamma_j = \hat{\gamma}_j + n^{-1}(\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{T}_n - \mathbb{E}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{T}_n)$, we have

$$\begin{aligned} d_{j2} &= b_j^2 \mathbb{E}\hat{\gamma}_j \left(\mathbf{r}_j^* \mathbf{D}_j^{-1} \mathbf{Q}^m \mathbf{K}_n \mathbf{r}_j - \frac{1}{n} \mathrm{tr}\mathbf{Q}^m \mathbf{K}_n \mathbf{T}_n \mathbf{D}_j^{-1} \right) \\ &\quad + \frac{b_j^2}{n} \mathbb{E}\gamma_j \left(\beta_j \mathbf{r}_j^* \mathbf{D}_j^{-1} \mathbf{Q}^m \mathbf{K}_n \mathbf{T}_n \mathbf{D}_j^{-1} \mathbf{r}_j \right) \\ &= d_{j2}^{(1)} + d_{j2}^{(2)}. \end{aligned}$$

We obtain from Lemma 6.4 and (4.32),

$$|d_{j2}^{(1)}| \le \frac{C}{n^2} \mathrm{tr}\mathbf{Q}^{m+1} \mathbf{D}_j^{-1} (\mathbf{D}_j^{-1})^*.$$

In view of (4.29) and the estimate of $d_{j1}^{(2)}$, we have

$$|d_{j2}^{(2)}| \le \frac{C\|\mathbf{Q}^m\|}{n^2 v^3} \left(\frac{1}{n} \mathrm{tr}\mathbf{Q}^2 \right)^{1/2},$$

which implies that for $v > C_0 n^{-1/2}$,

$$d_{j2} \le \frac{C}{n^2} \mathrm{tr}\mathbf{Q}^{m+1} \mathbf{D}_j^{-1} (\mathbf{D}_j^{-1})^* + \frac{C\|\mathbf{Q}^m\|}{nv} \left(\frac{1}{n} \mathrm{tr}\mathbf{Q}^2 \right)^{1/2}.$$

Similarly, using (4.22), (4.24) and Hölder's inequality, for $v > C_0 n^{-1/2}$, we also have

$$|d_{jk}| \le \frac{C\|\mathbf{Q}^m\|}{nv} \left(\frac{1}{n} \mathrm{tr}\mathbf{Q}^2 \right)^{1/2},$$

where $k = 3, 4, 5$. Therefore, summarizing the above arguments yields that for $v > C_0 n^{-1/2}$,

$$\left| \frac{1}{n} \mathrm{tr}\mathbf{Q}^m \mathbf{K}_n + \frac{z}{n} \mathrm{tr}\mathbf{Q}^m \mathbf{D}^{-1} \right| \le \frac{C}{n^2} \mathrm{tr}\mathbf{Q}^{m+1} \mathbf{D}_j^{-1} (\mathbf{D}_j^{-1})^* + \frac{C\|\mathbf{Q}^m\|}{nv} \left(\frac{1}{n} \mathrm{tr}\mathbf{Q}^2 \right)^{1/2}.$$

If $m = 1$, from Lemma 6.6, it is straightforward to get

$$\left| \frac{1}{n} \mathrm{tr}\mathbf{Q}\mathbf{K}_n \right| \le \frac{1}{n} \mathrm{tr}\mathbf{Q}^2$$

and

$$\frac{C}{n^2} \mathrm{tr}\mathbf{Q}^2 \mathbf{D}_j^{-1} (\mathbf{D}_j^{-1})^* \le \frac{C}{n^2 v^2} \mathrm{tr}\mathbf{Q}^2.$$

Furthermore, using Lemma 6.6, we have for $v < 1$,

$$\|\mathbf{Q}\| = \|\mathbf{K}_n\| \le \frac{4}{v}.$$

Therefore, for $1 > v > C_0 n^{-1/2}$ and $u \in [a, b]$, we conclude that

$$\left| \frac{1}{n} \mathrm{tr}\mathbf{Q}\mathbf{D}^{-1} \right| \le \frac{C}{n} \mathrm{tr}\mathbf{Q}^2 + \left(\frac{C}{n} \mathrm{tr}\mathbf{Q}^2 \right)^{1/2}. \tag{4.33}$$

On the other hand, if $m = 0$, then we have

$$\left|\frac{1}{n}\mathrm{tr}\mathbf{K}_n + \frac{z}{n}\mathrm{tr}\mathbf{D}^{-1}\right| \leq \frac{C}{n^2}\mathrm{tr}\mathbf{Q}\mathbf{D}_j^{-1}(\mathbf{D}_j^{-1})^* + \frac{C}{nv}\left(\frac{1}{n}\mathrm{tr}\mathbf{Q}^2\right)^{1/2}.$$

In addition, by Lemma 6.7 we obtain

$$\frac{1}{n^2}|\mathbb{E}(\mathrm{tr}\mathbf{Q}\mathbf{D}_j^{-1}(\mathbf{D}_j^{-1})^*)| = \frac{1}{n^2}|\mathbb{E}(\mathrm{tr}\mathbf{Q}\mathbf{D}^{-1}(\mathbf{D}^{-1})^*)| + O\left(\frac{1}{n^2v^3}\right).$$

Since $\mathbf{D}$ has the decomposition

$$\mathbf{D} = \mathbf{V}_n^* \begin{pmatrix} (\lambda_1^{\mathbf{S}_n} - z)^{-1} & & \\ & \ddots & \\ & & (\lambda_p^{\mathbf{S}_n} - z)^{-1} \end{pmatrix} \mathbf{V}_n,$$

where $\mathbf{V}_n^*\mathbf{V}_n = \mathbf{I}$. This implies

$$\begin{aligned} \mathbb{E}\mathrm{tr}\mathbf{Q}\mathbf{D}^{-1}(\mathbf{D}^{-1})^* &= \mathbb{E}\mathrm{tr}\mathbf{Q}\mathbf{V}_n^* \begin{pmatrix} |\lambda_1^{\mathbf{S}_n} - z|^{-2} & & \\ & \ddots & \\ & & |\lambda_p^{\mathbf{S}_n} - z|^{-2} \end{pmatrix} \mathbf{V}_n \\ &= \mathbb{E}v^{-1}\mathrm{tr}\mathbf{Q}\mathbf{V}_n^* \begin{pmatrix} \Im(\lambda_1^{\mathbf{S}_n} - z)^{-1} & & \\ & \ddots & \\ & & \Im(\lambda_p^{\mathbf{S}_n} - z)^{-1} \end{pmatrix} \mathbf{V}_n \\ &= v^{-1}\mathbb{E}\Im\mathrm{tr}\mathbf{Q}\mathbf{D}^{-1} \leq v^{-1}|\mathbb{E}\mathrm{tr}\mathbf{Q}\mathbf{D}^{-1}|. \end{aligned}$$

Then combining (4.33) and the definition of ω_n, we conclude that

$$|zy_n\omega_n| \leq \frac{C}{n^2v}\mathrm{tr}\mathbf{Q}^2 + \frac{C}{nv}\left(\frac{1}{n}\mathrm{tr}\mathbf{Q}^2\right)^{1/2}, \tag{4.34}$$

which implies that for $v > C_0n^{-1/2}$ with $C_0 > \sqrt{2C}$,

$$|zy_n\omega_n| \leq \frac{v}{2n}\mathrm{tr}\mathbf{Q}^2 + \frac{v}{2}\left(\frac{1}{n}\mathrm{tr}\mathbf{Q}^2\right)^{1/2}. \tag{4.35}$$

Moreover, recalling

$$-z\omega_n = \int \frac{1}{t\mathbb{E}\underline{s}_n(z) + 1}dH_n(t) + z\mathbb{E}s_n(z). \tag{4.36}$$

Taking the imaginary parts on both side of (4.36), we have

$$\int \frac{t\Im\mathbb{E}\underline{s}_n}{|1 + t\mathbb{E}\underline{s}_n(z)|^2}dH_n(t) = -\Im(z\omega_n) - \Im(z\mathbb{E}\underline{s}_n(z)).$$

If $\Im \mathbb{E}\underline{s}_n \geq v > 0$, then from (4.35) we have

$$\frac{|\Im(z\omega_n)|}{\Im \mathbb{E}\underline{s}_n(z)} = \frac{|v\Re\omega_n + u\Im\omega_n|}{\Im \mathbb{E}\underline{s}_n(z)} \leq \left(\frac{C}{n^2 v} + \frac{1}{2n}\right)\mathrm{tr}\mathbf{Q}^2 + \left(\frac{C}{nv} + \frac{1}{2}\right)\left(\frac{1}{n}\mathrm{tr}\mathbf{Q}^2\right)^{1/2}.$$

And using (2.4), we get

$$\frac{|\Im z \mathbb{E}s_n(z)|}{\Im \mathbb{E}\underline{s}_n(z)} = \frac{|v\Re \mathbb{E}\underline{s}_n(z) + u\Im \mathbb{E}\underline{s}_n(z)|}{y_n \Im \mathbb{E}\underline{s}_n(z)} < C.$$

On the other hand, if $\Im \mathbb{E}\underline{s}_n \leq v$, then by (4.28), we also have the above two inequalities. Therefore we have

$$\int \frac{t}{|1 + t\mathbb{E}\underline{s}_n|^2} dH_n(t) \leq \left(\frac{C}{n^2 v} + \frac{1}{2n}\right)\mathrm{tr}\mathbf{Q}^2 + \left(\frac{C}{nv} + \frac{1}{2}\right)\left(\frac{1}{n}\mathrm{tr}\mathbf{Q}^2\right)^{1/2} + C.$$

Furthermore, we have

$$\frac{1}{n}\mathrm{tr}\mathbf{Q}^2 = \int \frac{1}{|1 + t\mathbb{E}\underline{s}_n|^2} dH_n(t) \leq \frac{1}{\lambda_1^{\mathbf{T}_n}} \int \frac{t}{|1 + t\mathbb{E}\underline{s}_n|^2} dH_n(t).$$

This yields for $v > C_0 n^{-1/2}$ and $u \in [a, b]$,

$$\frac{1}{n}\mathrm{tr}\mathbf{Q}^2 \leq C.$$

Then together with (4.34), we conclude that

$$|\omega_n| \leq \frac{C}{nv}, \tag{4.37}$$

which completes the proof of Lemma 4.1.

5. The proof of Theorem 1.3

From Theorem 1.1, we have

$$\tilde{\Delta}_n \leq \sup_x |F^{\mathbf{S}_n}(x) - \mathbb{E}F^{\mathbf{S}_n}(x)| + O(n^{-1/2}).$$

And also, by Proposition 2.2, we have

$$\sup_x |F^{\mathbf{S}_n}(x) - \mathbb{E}F^{\mathbf{S}_n}(x)| \leq C\left(\int_{-A}^{A} |s_n(z) - \mathbb{E}s_n(z)| du + 2\pi v^{-1} \int_{|x|>B} |F^{\mathbf{S}_n}(x) - \mathbb{E}F^{\mathbf{S}_n}(x)| dx + v^{-1} \sup_x \int_{|u| \leq 2vc_*} |F^{y_n, H_n}(x+u) - F^{y_n, H_n}(x)| du + O(n^{-1/2})\right).$$

Thus, by Lemma 4.3, to prove Theorem 1.3 we need to get the bound of $|F^{\mathbf{S}_n}(x) - \mathbb{E}F^{\mathbf{S}_n}(x)|$ and $|F^{y_n, H_n}(x+u) - F^{y_n, H_n}(x)|$ which are obtained as follows.

Lemma 5.1. *Under the assumptions of Theorem 1.1, we have*

$$\int_{|x|>5} |F^{\mathbf{S}_n}(x) - F^{y_n,H_n}(x)|dx = o(n^{-2}) \quad a.s.$$

Proof. Note that if $x > 5$ then $F^{y_n,H_n}(x) = 1$. Therefore we have

$$\int_{|x|>5} \mathbb{E}|F^{\mathbf{S}_n}(x) - F^{y_n,H_n}(x)|dx = \int_5^\infty (1 - \mathbb{E}F^{\mathbf{S}_n}(x))dx$$

$$= \int_5^\infty \frac{1}{p}\sum_{k=1}^{p} P(\lambda_k^{\mathbf{S}_n} > x)dx \le \int_5^\infty P(\lambda_p^{\mathbf{S}_n} > x)dx. \tag{5.1}$$

It follows from (3.15) of [6] that,

$$\int_{|x|>5} |F^{S_n}(x) - F^{y_n,H_n}(x)|dx = O(n^{-2}) \quad \text{a.s.} \quad \square \tag{5.2}$$

Lemma 5.2. *Under the conditions of Theorem 1.1, for* $\forall\ v > 0$ *we have*

$$\sup_x \int_{|u|\le v} |F^{y,H}(x+u) - F^{y,H}(x)|du \le \frac{4v^2}{\pi\sqrt{\lambda_0}\sqrt{y}(1-\sqrt{y}+\sqrt{v})}.$$

Proof. From [8], we can get that the support of $F^{y,H}(x)$ is a subset of $[\lambda_0(1-\sqrt{y})^2, (1+\sqrt{y})^2]$. Let $\Phi(x) = \int_0^v (F^{y,H}(x+u) - F^{y,H}(x))du$, we have

$$\begin{aligned}
&\sup_x \int_{|u|\le v} |F^{y,H}(x+u) - F^{y,H}(x)|du \\
&\le \sup_{x\ge\lambda_0(1-\sqrt{y})^2} \int_{|u|\le v} |F^{y,H}(x+u) - F^{y,H}(x)|du \\
&\le 2 \sup_{x\ge\lambda_0(1-\sqrt{y})^2} \Phi(x).
\end{aligned} \tag{5.3}$$

Then by Lemma 2.3, we have

$$\begin{aligned}
\Phi(x) &= \int_0^v \int_x^{x+u} f(t)dtdu \le \int_0^v \int_x^{x+u} \pi^{-1}(\lambda_0 yt)^{-1/2}dtdu \\
&= \int_x^{x+v} \pi^{-1}(\lambda_0 yt)^{-1/2}(x+v-t)dt \\
&\le \frac{v}{\pi\sqrt{\lambda_0 y_n}} \int_x^{x+v} \frac{dt}{\sqrt{t}} \\
&= \frac{2v}{\pi\sqrt{\lambda_0 y}}\left(\sqrt{x+v} - \sqrt{x}\right) \\
&= \frac{2v^2}{\pi\sqrt{\lambda_0 y}\left(\sqrt{x+v}+\sqrt{x}\right)},
\end{aligned}$$

which together with (5.3) implies the lemma. $\square$

So in order to prove (1.3), we only need to show that for $v = n^{-2/5}$,

$$\tilde{\Delta}_n = O(v). \tag{5.4}$$

From Lemma 4.3

$$\int_{-A}^{A} \mathbb{E}|s_n(z) - \mathbb{E}s_n(z)|^2 du \le \frac{C}{n^2 v^3} = O(v^2),$$

which together with Lemmas 5.1 and 5.2 implies (5.4).

Moreover, (1.4) follows from the inequality that for $v = n^{-2/5+\eta}$ and $l > (5\eta)^{-1}$, (by Lemma 4.3)

$$\int_{-A}^{A} \mathbb{E}|s_n(z) - \mathbb{E}s_n(z)|^{2l} du \le \frac{C}{n^{2l} v^{3l}} = O(v^{2l} n^{-5l\eta}).$$

Therefore, the proof of Theorem 1.3 is complete.

6. Some basic lemmas

In this section, we give some basic lemmas which are used in the paper.

Lemma 6.1 (*Lemma 2.6 in [3]*). *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $p \times n$ *complex matrices. Then*

$$\sup_x |F^{\mathbf{AA}^*}(x) - F^{\mathbf{BB}^*}(x)| \le \frac{1}{p}\mathrm{rank}(\mathbf{A} - \mathbf{B}).$$

Lemma 6.2 (*Lemma A.47 in [11]*). *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $p \times n$ *complex matrices. Then*

$$L\left(F^{\mathbf{AA}^*}, F^{\mathbf{BB}^*}\right) \le 2\|\mathbf{A}\|\,\|\mathbf{A} - \mathbf{B}\| + \|\mathbf{A} - \mathbf{B}\|^2,$$

where $L(\cdot, \cdot)$ *denotes the Lévy distance.*

Lemma 6.3 (*Lemma B.19 in [11]*). *Let* F_1, F_2 *be distribution functions and let* G *satisfy* $\sup_x |G(x+t) - G(x)| \le g(t)$, *for all* $t \ge 0$, *where* g *is an increasing and continuous function such that* $g(0) = 0$. *Then*

$$\sup_x |F_1(x) - G(x)| \le 3 \max\left\{\sup_x |F_2(x) - G(x)|, L(F_1, F_2), g(L(F_1, F_2))\right\}.$$

Lemma 6.4 (*(1.15) of [9]*). *Let* $\mathbf{A} = (a_{ij})_{p\times p}$ *and* $\mathbf{B} = (b_{ij})_{p\times p}$ *be nonrandom matrices and* $X = (x_1, \dots, x_n)^*$ *be a random vector of independent entries. Assume that* $\mathbb{E}x_i = 0$ *and* $\mathbb{E}|x_i|^2 = 1$. *Then we have,*

$$\begin{aligned}&\mathbb{E}(X^*\mathbf{A}X - \mathrm{tr}\mathbf{A})(X^*\mathbf{B}X - \mathrm{tr}\mathbf{B}) \\ &\quad = \sum_{i=1}^{p} (\mathbb{E}|x_i|^4 - |\mathbb{E}x_i^2|^2 - 2)a_{ii}b_{ii} + \mathrm{tr}\mathbf{A}_x\mathbf{B}_x^T + \mathrm{tr}\mathbf{AB},\end{aligned} \tag{6.1}$$

where $\mathbf{A}_x = \left(\mathbb{E}x_i^2 a_{ij}\right)_{p\times p}$ *and* $\mathbf{B} = \left(\mathbb{E}x_i^2 b_{ij}\right)_{p\times p}$.

Lemma 6.5 (*Lemma 2.7 of [7]*). *Let* $\mathbf{A}$ *be an* $n \times n$ *nonrandom matrix and* $X = (x_1, \ldots, x_n)^*$ *be a random vector of independent entries. Assume that* $\mathbb{E}x_i = 0, \mathbb{E}|x_i|^2 = 1$*, and* $E|x_j|^l \leq \nu_l$*. Then, for any* $p \geq 1$*,*

$$\mathbb{E}|X^*\mathbf{A}X - \mathrm{tr}\mathbf{A}|^p \leq C_p \left(\left(\nu_4 \mathrm{tr}(\mathbf{A}\mathbf{A}^*)\right)^{p/2} + \nu_{2p} \mathrm{tr}(\mathbf{A}\mathbf{A}^*)^{p/2} \right),$$

where C_p *is a constant depending on* p *only.*

Lemma 6.6 (*Lemma 2.3 of [24]*). *For* $z = u + iv \in \mathbb{C}^+$*, let* $s(z)$ *be the Stieltjes transform of any distribution function,* $\mathbf{A}$ *be* $n \times n$ *Hermitian nonnegative definite matrix. Then*

$$\|(s(z)\mathbf{A} + \mathbf{I})^{-1}\| \leq \max\{4\|\mathbf{A}\|/v, 2\}.$$

Lemma 6.7 (*Lemma 2.6 of [25]*). *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*

$$|\mathrm{tr}((\mathbf{B} - z\mathbf{I})^{-1} - (\mathbf{B} + \tau\mathbf{q}\mathbf{q}^* - z\mathbf{I})^{-1})\mathbf{A}| \leq \frac{\|\mathbf{A}\|}{v}.$$

Lemma 6.8 (*(1.2) of [26]*). *Let* G *be a function of bounded variation on the real line. Then for any continuity points* $a < b$ *of* G*, we have*

$$G\{[a, b]\} = \lim_{\epsilon \downarrow 0} \frac{1}{\pi} \int_a^b \Im s_G(x + i\epsilon) dx,$$

where s_G *is the Stieltjes transform of* G*.*

Lemma 6.9 (*Burkholder Inequality*). *Let* $\{X_k\}$ *be a complex martingale difference sequence with respect to the increasing* σ*-field* $\mathcal{F}_k$*, and let* $\mathbb{E}_k$ *denote conditional expectation with respect to* $\mathcal{F}_k$*. Then we have*

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

$$\mathbb{E}\left|\sum_{k=1}^n X_k\right|^p \leq K_p \mathbb{E}\left(\sum_{k=1}^n |X_k|^2\right)^{p/2}, \tag{6.2}$$

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

$$\mathbb{E}\left|\sum_{k=1}^n X_k\right|^p \leq K_p^* \left(\mathbb{E}\left(\sum_{k=1}^n \mathbb{E}_{k-1}|X_k|^2\right)^{p/2} + \mathbb{E}\sum_{k=1}^n |X_k|^p\right). \tag{6.3}$$

Acknowledgements

Z.D. Bai was partially supported by the CNSF 10871036. J. Hu was partially supported by the Fundamental Research Funds for the Central Universities. W. Zhou was partially supported by grant R-155-000-095-112 at the National University of Singapore.

References

[1] Z.D. Bai, Convergence rate of expected spectral distributions of large random matrices. Part I. Wigner matrices, The Annals of Probability 21 (2) (1993) 625–648.

[2] Z.D. Bai, Convergence rate of expected spectral distributions of large random matrices. Part II. Sample covariance matrices, The Annals of Probability 21 (2) (1993) 649–672.
[3] Z.D. Bai, Methodologies in spectral analysis of large dimensional random matrices, a review, Statistica Sinica 9 (1999) 611–677.
[4] Z.D. Bai, D.D. Jiang, J.F. Yao, S.R. Zheng, Corrections to LRT on large-dimensional covariance matrix by RMT, The Annals of Statistics 37 (6B) (2009) 3822–3840.
[5] Z.D. Bai, H.X. Liu, W.K. Wong, Enhancement of the applicability of markowitz's portfolio optimization by utilizing random matrix theory, Mathematical Finance 19 (4) (2009) 639–667.
[6] Z.D. Bai, B.Q. Miao, J.F. Yao, Convergence rates of spectral distributions of large sample covariance matrices, SIAM Journal on Matrix Analysis and Applications 25 (1) (2003) 105–127.
[7] Z.D. Bai, J.W. Silverstein, No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices, The Annals of Probability 26 (1) (1998) 316–345.
[8] Z.D. Bai, J.W. Silverstein, Exact separation of eigenvalues of large dimensional sample covariance matrices, Annals of Probability 27 (3) (1999) 1536–1555.
[9] Z.D. Bai, J.W. Silverstein, CLT for linear spectral statistics of large-dimensional sample covariance matrices, Annals of Probability 32 (1) (2004) 553–605.
[10] Z.D. Bai, J.W. Silverstein, On the signal-to-interference ratio of CDMA systems in wireless communications, The Annals of Applied Probability 17 (1) (2007) 81–101.
[11] Z.D. Bai, J.W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, 2nd ed., Springer Verlag, 2010.
[12] Z.D. Bai, X.Y. Wang, W. Zhou, Functional CLT for sample covariance matrices, Bernoulli 16 (4) (2010) 1086–1113.
[13] R. Couillet, M. Debbah, J.W. Silverstein, A deterministic equivalent for the capacity analysis of correlated multi-user MIMO channels, Preprint in IEEE Transactions on Information Theory.
[14] R. Couillet, J.W. Silverstein, M. Debbah, Eigen-inference for multi-source power estimation, in: IEEE International Symposium on Information Theory, 2010, pp. 1673–1677.
[15] V.L. Girko, Spectral theory of random matrices, Russian Mathematical Surveys 40 (1) (1985) 77–120.
[16] V.L. Girko, Asymptotics of the distribution of the spectrum of random matrices, Russian Mathematical Surveys 44 (4) (1989) 3–36.
[17] F. Götze, A. Tikhomirov, Rate of convergence to the semi-circular law, Probability Theory and Related Fields 127 (2) (2003) 228–276.
[18] F. Götze, A. Tikhomirov, Rate of convergence in probability to the Marchenko–Pastur law, Bernoulli 10 (3) (2004) 503–548.
[19] F. Götze, A. Tikhomirov, The rate of convergence of spectra of sample covariance matrices, Theory of Probability and its Applications 54 (1) (2010) 129–140.
[20] B.Y. Jing, G.M. Pan, Q.M. Shao, W. Zhou, Nonparametric estimate of spectral density functions of sample covariance matrices: a first step, The Annals of Statistics 38 (6) (2010) 3724–3750.
[21] H. Li, H.V. Poor, Large system spectral analysis of covariance matrix estimation, IEEE Transactions on Information Theory 55 (3) (2009) 1395–1422.
[22] R.R. Nadakuditi, J.W. Silverstein, Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples, IEEE Journal on Selected Topics in Signal Processing 4 (3) (2010) 468–480.
[23] J.W. Silverstein, The limiting eigenvalue distribution of a multivariate F matrix, SIAM Journal on Mathematical Analysis 16 (3) (1985) 641.
[24] J.W. Silverstein, Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices, Journal of Multivariate Analysis 55 (2) (1995) 331–339.
[25] J.W. Silverstein, Z.D. Bai, On the empirical distribution of eigenvalues of a class of largedimensional random matrices, Journal of Multivariate Analysis 54 (2) (1995) 175–192.
[26] J.W. Silverstein, S.I. Choi, Analysis of the limiting spectral distribution of large dimensional random matrices, Journal of Multivariate Analysis 54 (2) (1995) 295–309.
[27] Y.Q. Yin, Limiting spectral distribution for a class of random matrices, Journal of Multivariate Analysis 20 (1) (1986) 50–68.