Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2014, Vol. 50, No. 2, 655–677
DOI: 10.1214/12-AIHP506

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Comparison between two types of large sample covariance matrices

Guangming Pan

Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371.
E-mail: gmpan@ntu.edu.sg

Received 16 March 2011; revised 18 June 2012; accepted 21 June 2012

Abstract. Let $\{X_{ij}\}$, $i, j = \cdots$, be a double array of independent and identically distributed (i.i.d.) real random variables with $EX_{11} = \mu$, $E|X_{11} - \mu|^2 = 1$ and $E|X_{11}|^4 < \infty$. Consider sample covariance matrices (with/without empirical centering) $\mathcal{S} = \frac{1}{n}\sum_{j=1}^n (\mathbf{s}_j - \bar{\mathbf{s}})(\mathbf{s}_j - \bar{\mathbf{s}})^T$ and $\mathbf{S} = \frac{1}{n}\sum_{j=1}^n \mathbf{s}_j\mathbf{s}_j^T$, where $\bar{\mathbf{s}} = \frac{1}{n}\sum_{j=1}^n \mathbf{s}_j$ and $\mathbf{s}_j = \mathbf{T}_n^{1/2}(X_{1j}, \ldots, X_{pj})^T$ with $(\mathbf{T}_n^{1/2})^2 = \mathbf{T}_n$, non-random symmetric non-negative definite matrix. It is proved that central limit theorems of eigenvalue statistics of $\mathcal{S}$ and $\mathbf{S}$ are different as $n \to \infty$ with p/n approaching a positive constant. Moreover, it is also proved that such a different behavior is not observed in the average behavior of eigenvectors.

Résumé. Soit $\{X_{ij}\}$, $i, j = 1, 2, \ldots$, un tableau à double entrées, les X_{ij} étant des variables aléatoires réelles indépendantes et identiquement distribuées (i.i.d.) et où $\mathbf{E}X_{11} = \mu$, $\mathbf{E}|X_{11} - \mu|^2 = 1$ et $\mathbf{E}|X_{11}|^4 < \infty$. Considérons les matrices de covariances empiriques suivantes (avec/sans centrage empirique): $\mathcal{S} = \frac{1}{n}\sum_{j=1}^n (\mathbf{s}_j - \bar{\mathbf{s}})(\mathbf{s}_j - \bar{\mathbf{s}})^T$ et $\mathbf{S} = \frac{1}{n}\sum_{j=1}^n \mathbf{s}_j\mathbf{s}_j^T$, avec $\bar{\mathbf{s}} = \frac{1}{n}\sum_{j=1}^n \mathbf{s}_j$ et $\mathbf{s}_j = \mathbf{T}_n^{1/2}(X_{1j}, \ldots, X_{pj})^T$, où $(\mathbf{T}_n^{1/2})^2 = \mathbf{T}_n$ est une matrice déterministe définie positive. Nous démontrons que, sous le régime asymptotique $n \to \infty$ et p/n converge vers une constante positive, le théorème central limite pour la statistique $\mathcal{S}$ est différent de celui concernant la statistique $\mathbf{S}$. En outre, nous montrons que cette différence de comportement n'est pas observée pour le comportement moyen des vecteurs propres.

MSC: 15A52; 60F15; 62E20; 60F17

Keywords: Central limit theorems; Eigenvectors and eigenvalues; Sample covariance matrix; Stieltjes transform; Strong convergence

1. Introduction

Let $\{X_{ij}\}$, $i, j = \cdots$, be a double array of independent and identically distributed (i.i.d.) real random variables with $EX_{11} = \mu$, $E|X_{11} - \mu|^2 = 1$ and $E|X_{11}|^4 < \infty$. Write $\mathbf{x}_j = (X_{1j}, \ldots, X_{pj})^T$, $\mathbf{s}_j = \mathbf{T}_n^{1/2}\mathbf{x}_j$ and $\mathbf{X}_n = (\mathbf{s}_1, \ldots, \mathbf{s}_n)$ where $(\mathbf{T}_n^{1/2})^2 = \mathbf{T}_n$, non-random symmetric non-negative definite matrix. It is well known that the sample covariance matrix is defined by

$$\mathcal{S} = \frac{1}{n}\sum_{j=1}^n (\mathbf{s}_j - \bar{\mathbf{s}})(\mathbf{s}_j - \bar{\mathbf{s}})^T,$$

where $\bar{\mathbf{s}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{s}_j = \mathbf{T}_n^{1/2}\bar{\mathbf{x}}$ with $\bar{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j$. It lies at the roots of many methods in multivariate analysis (see [1]). However, the commonly used sample covariance matrix in random matrix theory is

$$\mathbf{S} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{s}_j\mathbf{s}_j^T.$$

Since the matrix $\mathbf{S}$ has been well studied in the literature (see [5]), a natural question is that whether the asymptotic results for eigenvalues and (or) eigenvectors of $\mathbf{S}$ apply for the matrix $\mathcal{S}$ as well. This paper attempts to address it.

First, observe that

$$\mathcal{S} = \mathbf{S} - \bar{\mathbf{s}}\bar{\mathbf{s}}^T \tag{1.1}$$

and thus by the rank inequality (see Theorem A.44 in [5]) there is no difference when the limiting empirical spectral distribution (ESD) of eigenvalues is our concern only. Here the ESD of a matrix $\mathbf{B}$ is defined by

$$F^{\mathbf{B}}(x) = \frac{1}{p}\sum_{i=1}^{p} I(\mu_i \le x), \tag{1.2}$$

where $\mu_1, \ldots, \mu_p$ are the eigenvalues of $\mathbf{B}$. When $F^{\mathbf{T}_n}$ converges weakly to a non-random distribution H and $p/n \to c > 0$, Marcenko and Pastur [14], Yin [23] and Silverstein [19] proved that, with probability one, $F^{(1/n)\mathbf{X}_n^T\mathbf{X}_n}(x)$ converges in distribution to a nonrandom distribution function $\underline{F}_{c,H}(x)$ whose Stieltjes transform $\underline{m}(z) = m_{\underline{F}_{c,H}}(z)$ is, for each $z \in \mathcal{C}^+ = \{z \in \mathcal{C}\colon \Im z > 0\}$, the unique solution to the equation

$$\underline{m} = -\left(z - c\int \frac{t\,\mathrm{d}H(t)}{1 + t\underline{m}}\right)^{-1}. \tag{1.3}$$

Here the Stieltjes transform $m_F(z)$ for any probability distribution function $F(x)$ is defined by

$$m_F(z) = \int \frac{1}{x - z}\,\mathrm{d}F(x), \quad z \in \mathcal{C}^+. \tag{1.4}$$

Noting the relationship between the spectra of $\mathbf{S}$ and $n^{-1}\mathbf{X}_n^T\mathbf{X}_n$, we have

$$\underline{m}(z) = -\frac{1-c}{z} + cm(z), \tag{1.5}$$

where $m(z)$ is the Stieltjes transform of $F_{c,H}(x)$, the limit of $F^{\mathbf{S}}$.

Particularly, when the population matrix $\mathbf{T}_n$ is the identity matrix the limiting ESD of $\mathcal{S}$ is Marcenko and Pastur's law $F_c(x)$ (see [14] and [12]), which has a density function

$$p_c(x) = \begin{cases} (2\pi cx)^{-1}\sqrt{(b-x)(x-a)}, & a \le x \le b, \\ 0, & \text{otherwise} \end{cases}$$

and has point mass $1 - c^{-1}$ at the origin if $c > 1$, where $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$.

Secondly, when $\mathbf{T}_n = \mathbf{I}$, it was reported in [11] that the maximum eigenvalues of $\mathbf{S}$ and $\mathcal{S}$ have the same limit, while, [22] announced that the minimum eigenvalues of $\mathbf{S}$ and $\mathcal{S}$ also share the identical limit. That is,

$$\lambda_{\max}(\mathcal{S}) \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2 \tag{1.6}$$

and when $c < 1$

$$\lambda_{\min}(\mathcal{S}) \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2, \tag{1.7}$$

where $\lambda_{\max}(\mathcal{S})$ and $\lambda_{\min}(\mathcal{S})$ denote, respectively, the maximum and minimum eigenvalues of $\mathcal{S}$. Moreover, it was proved in [10] that asymptotic distribution of the maximum eigenvalues of $\mathbf{S}$ after centering and scaling is the Tracy–Widom law of order 1 when $\mathbf{S}$ has a Wishart distribution $W_p(n, \mathbf{I})$. In this case, note that $n\mathcal{S}$ has the same distribution as $\sum_{j=1}^{n-1} \mathbf{x}_j \mathbf{x}_j^T$ and hence it is not difficult to prove that the maximum eigenvalues of $\mathbf{S}$ and $\mathcal{S}$ share the same central limit theorem.

Thirdly, remarkable central limit theorems for the eigevalues of $\mathbf{S}$ were established in [3]. Moreover, the constraint that $EX_{11}^4 = 3$, imposed in [3], was removed in [16]. Since the ESDs of $\mathbf{S}$ and $\mathcal{S}$ have the same limit it seems that they might have the same central limit theorem for eigenvalues. Surprisingly, a detailed investigation shows they have different central limit theorems. The main reason is that $\bar{\mathbf{s}}$ involved in $\mathcal{S}$ contributes to the central limit theorem as well. Formally, let

$$G_n(x) = p\bigl(F^{\mathcal{S}}(x) - F_{c_n, H_n}(x)\bigr),$$

where $c_n = \frac{p}{n}$ and $F_{c_n,H_n}(x)$ denotes the function by substituting c_n for c and H_n for H in $F_{c,H}(x)$.

Theorem 1. *Suppose that*

(1) *For each n $X_{ij} = X_{ij}^n, i, j = 1, 2, \ldots,$ are i.i.d. real random variables with $EX_{11} = \mu$, $E|X_{11} - \mu|^2 = 1$ and $E|X_{11}|^4 < \infty$.*
(2) $\lim_{n\to\infty} \frac{p}{n} = c \in (0, \infty)$.
(3) *$g_1, \ldots, g_k$ are functions on $\mathbb{R}$ analytic on an open region $\mathcal{D}$ of the complex plane, which contains the real interval*

$$\Bigl[\liminf_n \lambda_{\min}(\mathbf{T}_n) I_{(0,1)}(c)(1 - \sqrt{c})^2, \limsup_n \lambda_{\max}(\mathbf{T}_n)(1 + \sqrt{c})^2\Bigr].$$

(4) *Let $\mathbf{T}_n$ be a $p \times p$ non-random symmetric non-negative definite matrix with spectral norm bounded above by a positive constant such that $H_n = F^{\mathbf{T}_n}$ converges weakly to a non-random distribution H.*
(5) *Let $\mathbf{e}_i$ be the vector of size $p \times 1$ with the ith element 1 and others 0. The matrix $\mathbf{T}_n$ also satisfies*

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{e}_i^T \mathbf{T}_n^{1/2} \bigl(\underline{m}(z_1)\mathbf{T}_n + \mathbf{I}\bigr)^{-1} \mathbf{T}_n^{1/2} \mathbf{e}_i \mathbf{e}_i^T \mathbf{T}_n^{1/2} \bigl(\underline{m}(z_2)\mathbf{T}_n + \mathbf{I}\bigr)^{-1} \mathbf{T}_n^{1/2} \mathbf{e}_i \to h_1(z_1, z_2)$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{e}_i^T \mathbf{T}_n^{1/2} \bigl(\underline{m}(z)\mathbf{T}_n + \mathbf{I}\bigr)^{-1} \mathbf{T}_n^{1/2} \mathbf{e}_i \mathbf{e}_i^T \mathbf{T}_n^{1/2} \bigl(\underline{m}(z)\mathbf{T}_n + \mathbf{I}\bigr)^{-2} \mathbf{T}_n^{1/2} \mathbf{e}_i \to h_2(z).$$

Then $(\int g_1(x)\,\mathrm{d}G_n(x), \ldots, \int g_k(x)\,\mathrm{d}G_n(x))$ converges weakly to a Gaussian vector $(X_{g_1}, \ldots, X_{g_k})$, with mean

$$\begin{aligned} EX_g = {} & -\frac{c}{2\pi \mathrm{i}} \int g(z) \frac{\int \frac{\underline{m}^3(z) t^2 \,\mathrm{d}H(t)}{(1 + t\underline{m}(z))^3}}{(1 - c\int \frac{\underline{m}^2(z) t^2 \,\mathrm{d}H(t)}{(1 + t\underline{m}(z))^2})^2} \,\mathrm{d}z - \frac{c(E(X_{11} - \mu)^4 - 3)}{2\pi \mathrm{i}} \int g(z) \frac{\underline{m}^3(z) h_2(z)}{1 - c\int \frac{\underline{m}^2(z) t^2 \,\mathrm{d}H(t)}{(1 + t\underline{m}(z))^2}} \,\mathrm{d}z \\ & + \frac{c}{2\pi \mathrm{i}} \int g(z) \frac{\underline{m}(z) \int \frac{t \,\mathrm{d}H(t)}{(1 + t\underline{m}(z))^2}}{z(1 - c\int \frac{(\underline{m}(z))^2 t^2 \,\mathrm{d}H(t)}{(1 + t\underline{m}(z))^2})} \,\mathrm{d}z \end{aligned} \tag{1.8}$$

and covariance function

$$\begin{aligned} \operatorname{Cov}(X_{g_1}, X_{g_2}) = {} & -\frac{1}{2\pi^2} \int\!\!\int \frac{g_1(z_1) g_2(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} \frac{\mathrm{d}}{\mathrm{d}z_1} \underline{m}(z_1) \frac{\mathrm{d}}{\mathrm{d}z_2} \underline{m}(z_2) \,\mathrm{d}z_1 \,\mathrm{d}z_2 \\ & - \frac{c(E(X_{11} - \mu)^4 - 3)}{4\pi^2} \int\!\!\int g_1(z_1) g_2(z_2) \frac{\mathrm{d}^2}{\mathrm{d}z_1 \,\mathrm{d}z_2} \bigl[\underline{m}(z_1)\underline{m}(z_2) h_1(z_1, z_2)\bigr] \,\mathrm{d}z_1 \,\mathrm{d}z_2. \end{aligned} \tag{1.9}$$

The contours in (1.8) *and* (1.9) *are both contained in the analytic region for the functions* $g_1, \dots, g_k$ *and both enclose the support of* $F_{c_n,H_n}(x)$ *for large* n. *Moreover, the contours in* (1.9) *are disjoint.*

Remark 1. *From Theorem* 1.1 *of* [3], *Theorem* 1.4 *of* [16] *and Theorem* 1 *we see that functionals of eigenvalues of* $\mathbf{S}$ *and* $\mathcal{S}$ *have the same asymptotic variances but they have different asymptotic means. Indeed, the additional term in the asymptotic mean is the last term of* (1.8), *which can be further written as*

$$\frac{c}{2\pi \mathrm{i}} \int g(z) \frac{\underline{m}(z) \int \frac{t\,\mathrm{d}H(t)}{(1+t\underline{m}(z))^2}}{z(1 - c \int \frac{(\underline{m}(z))^2 t^2\,\mathrm{d}H(t)}{(1+t\underline{m}(z))^2})}\,\mathrm{d}z = -\frac{1}{\pi} \int_a^b g'(x) \arg\bigl[x\underline{m}(x)\bigr]\,\mathrm{d}x, \tag{1.10}$$

where $\underline{m}(x) = \lim_{z\to x} \underline{m}(z)$ *for* $0 \neq x \in \mathbb{R}$. *Here one should note that that Theorem* 1.1 *of* [3] *did not handle the case when the common mean of the underlying random variables is not equal to zero.*

Remark 2. *When developing central limit theorems, it is necessary to study the limits of the products of* $(E(X_{11} - \mu)^4 - 3)$ *and the diagonal elements of* $(\mathbf{S} - z\mathbf{I})^{-1}$. *For general* $\mathbf{T}_n$, *the limits of such diagonal elements are not necessary the same. That is why Assumption* 5 *has to be imposed. Roughly speaking, Assumption* 5 *is some kind of rotation invariance. For example, it is satisfied if* $\mathbf{T}_n$ *is the inverse of another sample covariance matrix with the population matrix being the identity. However, when* $E(X_{11} - \mu)^4 = 3$, *Assumption* 5 *can be removed as well because such diagonal elements disappear in this case thanks to the special fourth moment and one may see* [3]. *Also, when* $\mathbf{T}_n$ *is a diagonal matrix Assumption* 5 *in Theorem* 1 *is automatically satisfied, as pointed out in* [16], *and in this case*

$$h_1(z_1, z_2) = \int \frac{t^2\,\mathrm{d}H(t)}{((\underline{m}(z_1)t + 1))((\underline{m}(z_2)t + 1))}, \qquad h_2(z) = \int \frac{t^2\,\mathrm{d}H(t)}{(\underline{m}(z)t + 1)^3}.$$

Particulary, when $\mathbf{T}_n = \mathbf{I}$ *we have*

$$\begin{aligned} EX_g = {} & \frac{1}{2\pi} \int_a^b g'(x) \arg\biggl[1 - \frac{c\underline{m}^2(x)}{(1 + \underline{m}(x))^2}\biggr]\,\mathrm{d}x - \frac{1}{\pi} \int_a^b g'(x) \arg\bigl[x\underline{m}(x)\bigr]\,\mathrm{d}x \\ & - \frac{c(E(X_{11} - \mu)^4 - 3)}{\pi} \int_a^b g(x) \Im\biggl[\frac{\underline{m}^3(x)}{(\underline{m}(x) + 1)[(\underline{m}(x) + 1)^2 - c\underline{m}^2(x)]}\biggr]\,\mathrm{d}x, \end{aligned} \tag{1.11}$$

where

$$\underline{m}(x) = \frac{-(x + 1 - c) + \sqrt{(x - a)(b - x)}\,\mathrm{i}}{2x},$$

(*see* [3]). *For some simplified formulas for covariance function, refer to* (1.24) *in* [3] *and* (1.24) *in* [16].

Finally, let us take a look at the eigenvectors of $\mathbf{S}$ and $\mathcal{S}$. The matrix of orthonormal eigenvectors of $\mathbf{S}$ has a Haar measure on the orthogonal matrices when X_{11} is normally distributed and $\mathbf{T}_n = \mathbf{I}$. It is conjectured that for large n and for general X_{11} the matrix of orthonormal eigenvectors of $\mathbf{S}$ is close to being Haar distributed. Silverstein (1981) created an approach to address it. Further work along this direction can be found in Silverstein [18], [20], Bai, Miao and Pan [4] and Ledoit and Peche [13]. Here we would also like to point out that there are similar interests in eigenvectors of large Wigner matrix and one may see [6,7,9] and [21]. To consider the matrix $\mathcal{S}$, let $U_n \Lambda_n U_n^T$ denote the spectral decomposition of $\mathcal{S}$, where $\Lambda_n = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, $U_n = (u_{ij})$ is an orthogonal matrix consisting of the orthonormal eigenvectors of $\mathcal{S}$. Assume that $\mathbf{x}_n \in \mathbb{R}^p$, $\|\mathbf{x}_n\| = 1$, is an arbitrary non-random unit vector and $\mathbf{y} = (y_1, y_2, \dots, y_p)^T = U_n^T \mathbf{x}_n$. Following their approach, we define an empirical distribution function based on eigenvectors and eigenvalues as

$$F_1^{\mathcal{S}}(x) = \sum_{i=1}^{p} |y_i|^2 I(\lambda_i \le x), \tag{1.12}$$

where $\lambda_i's$ are eigenvalues of $\mathcal{S}$. Based on the above empirical spectral distribution function, we may further investigate central limit theorems of functions of eigenvalues and eigenvectors.

As will be seen below, the asymptotic properties of the eigenvectors matrix of $\mathcal{S}$ are the same as those of $\mathbf{S}$ in terms of this property (note: $EX_{11}=0$ required for $\mathbf{S}$). However, the advantage of $\mathcal{S}$ over $\mathbf{S}$ is that the common mean of underlying random variables is not necessarily zero to keep these types of properties. Unfortunately, this is not the case for $\mathbf{S}$. For example, consider $\mathbf{e}_1^T\mathbf{S}\mathbf{e}_1$ when X_{ij} are i.i.d. with $EX_{11}=\mu$ and $\operatorname{Var}(X_{11})=1$, and $\mathbf{T}_n=\mathbf{I}$. Then it is a simple matter to show that

$$\mathbf{e}_1^T\mathbf{S}\mathbf{e}_1 \xrightarrow{\text{a.s.}} 1+\mu^2$$

which is dependent on μ, different from Corollary 1 in Bai, Miao and Pan [4] when $\mu\neq 0$. Indeed, when dealing with central limit theorems of functions of eigenvalues and eigenvectors, [18] also proved that $EX_{11}=0$ is a necessary condition to keep this type of property for the matrix $\mathbf{S}$ with $\mathbf{T}_n=\mathbf{I}$.

Formally, let

$$G_{n1}(x)=\sqrt{n}\bigl(F_1^{\mathcal{S}}(x)-F_{c_n,H_n}(x)\bigr).$$

Theorem 2. *In addition to Conditions* (1), (2) *and* (4) *in Theorem* 1 *suppose that* $\mathbf{x}_n\in\mathbb{R}_1^p=\{\mathbf{x}\in\mathbb{R}^p, \|\mathbf{x}\|=1\}$ *and* $\mathbf{x}_n^T(\mathbf{T}_n-z\mathbf{I})^{-1}\mathbf{x}_n\to m_H(z)$ *where* $m_H(z)$ *denotes the Stieltjes transform of* $H(x)$. *Then, it holds that*

$$F_1^{\mathcal{S}}(x)\to F_{c,H}(x)\quad a.s.$$

By Theorem 2, we obtain

Corollary 1. *Let* $(\mathcal{S}^m)_{ii}, m=1,2,\dots,$ *denote the* i*th diagonal elements of matrices* $\mathcal{S}^m$. *Under the conditions of Theorem* 2 *for* $\mathbf{x}_n=\mathbf{e}_i$, *then for any fixed* m,

$$\lim_{n\to\infty}\left|\bigl(\mathcal{S}^m\bigr)_{ii}-\int x^m\,\mathrm{d}F_{c,H}(x)\right|\to 0,\quad a.s.$$

Theorem 3. *In addition to the assumptions in Theorem* 1 *and Theorem* 2 *suppose that as* $n\to\infty$

$$\sup_z\sqrt{n}\left|\mathbf{x}_n^T\bigl(\underline{m}_{F_{c,H}}(z)\mathbf{T}_n+\mathbf{I}\bigr)^{-1}\mathbf{x}_n-\int\frac{\mathrm{d}H_n(t)}{\underline{m}_{F_{c,H}}(z)t+1}\right|\to 0. \tag{1.13}$$

Then the following conclusions hold.

(a) *The random vector*

$$\left(\int g_1(x)\,\mathrm{d}G_{n1}(x),\dots,\int g_k(x)\,\mathrm{d}G_{n1}(x)\right)$$

forms a tight sequence.

(b) *If* $E(X_{11}-\mu)^4=3$, *the above random vector converges weakly to a Gaussian vector* $X_{g_1},\dots,X_{g_k}$, *with zero means and covariance function*

$$\operatorname{Cov}(X_{g_1},X_{g_2})=-\frac{1}{2\pi^2}\int\int g_1(z_1)g_2(z_2)\frac{(z_2\underline{m}(z_2)-z_1\underline{m}(z_1))^2}{c^2z_1z_2(z_2-z_1)(\underline{m}(z_2)-\underline{m}(z_1))}\,\mathrm{d}z_1\,\mathrm{d}z_2, \tag{1.14}$$

where the contours in (1.14) *are the same as in those in* (1.9).

(c) *Instead of* $E(X_{11}-\mu)^4=3$, *assuming that*

$$\max_i\bigl|\mathbf{e}_i^*\mathbf{T}_n^{1/2}\bigl(zm(z)\mathbf{T}_n+z\mathbf{I}\bigr)^{-1}\mathbf{x}_n\bigr|\to 0, \tag{1.15}$$

the assertions in (b) *still holds.*

Remark 3. *Assumptions* (1.13) *and* (1.15) *are imposed for the same reason as given in Remark* 2. *Besides*, *when* $\mathbf{T}_n$ *reduces to the identity matrix*, (1.15) *becomes*

$$\max_i |x_{ni}| \to 0. \tag{1.16}$$

As can be seen from (1.14) *the asymptotic covariance does not depend on the fourth moment of* X_{11} *and hence we have to impose the condition like* (1.16) *to ensure that the term involving the product of the fourth moment of* X_{11} *and the diagonal elements of* $(\mathbf{S} - z\mathbf{I})^{-1}$ *converges to zero in probability. One may see* [16] *for more details.*

Remark 4. *As pointed out in* [4], *when* $\mathbf{T}_n = I$ *condition* (1.13) *holds and formula* (1.14) *becomes*

$$\operatorname{Cov}(X_{g_1}, X_{g_2}) = \frac{2}{c}\left(\int g_1(x)g_2(x)\,\mathrm{d}F_c(x) - \int g_1(x_1)\,\mathrm{d}F_c(x_1)\int g_2(x_2)\,\mathrm{d}F_c(x_2)\right). \tag{1.17}$$

Theorem 1 essentially boils down to the study of the Stieltjes transform of $F^{\mathcal{S}}(x)$ while Theorems 2 and 3 boils down to the study of the Stieltjes transform of $F_1^{\mathcal{S}}(x)$. In view of this we conjecture the same phenomenon holds for other objects involving $F^{\mathcal{S}}(x)$ or $F_1^{\mathcal{S}}(x)$. For example we believe central limit theorems of the normalized empirical spectral distribution functions of $\mathcal{S}$ and $\mathbf{S}$ are different ([17] considers central limit theorems of the smoothed empirical spectral distribution functions).

We conclude this section by stating the structure of this work. Section 2 gives the proof of Theorem 1. Sections 3 and 4 pick up the proofs of Theorems 2 and 3, respectively. Throughout the paper M and C denote constants which may change from line to line, and all matrices and vectors are denoted by boldface letters.

2. Proof of Theorem 1 and (1.11)

The proof of Theorem 1 is essentially based on the Stieltjes transform following [3]. First, by analyticity of functions it is enough to consider the Stieltjes transforms of the empirical functions of sample covariance matrices. Moreover, note that the Stieltjes transform of the empirical function of $\mathcal{S}$ can be decomposed as the sum of the Stieltjes transform of the empirical function of $\mathbf{S}$ and some random variable involving sample mean and $\mathbf{S}$. Thus, our main aim is to prove that the extra random variable converges in probability in the C space to some nonrandom variable. Once this is finished, Theorem 1 follows from Slutsky's theorem, the continuous mapping theorem and the results in [3] and [16].

Before proceeding we observe the following useful fact. By the structure of $\mathcal{S}$, without loss of generality, we can assume $EX_{11} = 0$, $EX_{11}^2 = 1$ in the course of establishing Theorems 1–3 (but the fourth moment will be $E(X_{11} - \mu)^4$).

2.1. *Truncation of underlying random variables and random processes*

We begin the proof with the replacement of the underlying random variables X_{ij} with truncated and centralized variables. To this end, write

$$\mathcal{S} = \frac{1}{n}(\mathbf{X}_n - \mathbf{B})\big(\mathbf{X}_n^T - \mathbf{B}^T\big),$$

where $\mathbf{B} = (\bar{\mathbf{s}}, \bar{\mathbf{s}}, \ldots, \bar{\mathbf{s}})_{p\times n}$. Since the argument for (1.8) (and two lines below (1.8)) in [3] can be carried over directly to the present case we can choose a positive sequence ε_n such that

$$\varepsilon_n \to 0, \qquad \varepsilon_n^{-4} E|X_{11}|^4 I\big(|X_{11}| \ge \varepsilon_n\sqrt{n}\big) \to 0, \qquad \varepsilon_n n^{1/4} \to \infty. \tag{2.18}$$

Let $\hat{\mathcal{S}} = (1/n)(\hat{\mathbf{X}}_n - \hat{\mathbf{B}})(\hat{\mathbf{X}}_n^T - \hat{\mathbf{B}}^T)$ where $\hat{\mathbf{X}}_n$ and $\hat{\mathbf{B}}$ are respectively obtained from $\mathbf{X}_n$ and $\mathbf{B}$ with the entries X_{ij} replaced by $X_{ij}I(|X_{ij}| < \varepsilon_n\sqrt{n})$. We then obtain

$$\begin{aligned} P(\hat{\mathcal{S}} \ne \mathcal{S}) &\le P\Bigg(\bigcup_{i\le p, j\le n}\big(|X_{ij}| \ge \varepsilon_n\sqrt{n}\big)\Bigg) \le pnP\big(|X_{11}| \ge \varepsilon_n\sqrt{n}\big) \\ &\le M\varepsilon_n^{-4} E|X_{11}|^4 I\big(|X_{11}| \ge \varepsilon_n\sqrt{n}\big) \to 0, \end{aligned} \tag{2.19}$$

where and in what follows M denotes a constant which may change from line to line.

Define $\tilde{\mathcal{S}} = (1/n)(\tilde{\mathbf{X}}_n - \tilde{\mathbf{B}})(\tilde{\mathbf{X}}_n^T - \tilde{\mathbf{B}}^T)$ where $\tilde{\mathbf{X}}_n$ and $\tilde{\mathbf{B}}$ are respectively obtained from $\mathbf{X}_n$ and $\mathbf{B}$ with the entries X_{ij} replaced by $(\hat{X}_{ij} - E\hat{X}_{ij})/\sigma_n$ with $\hat{X}_{ij} = X_{ij} I(|X_{ij}| < \varepsilon_n \sqrt{n})$ and $\sigma_n^2 = E|\hat{X}_{ij} - E\hat{X}_{ij}|^2$. Denote by $\tilde{G}_n(x)$ and $\hat{G}_n(x)$ the analogues of $G_n(x)$ with the matrix $\mathcal{S}$ replaced by $\hat{\mathcal{S}}$ and $\tilde{\mathcal{S}}$ respectively. As in [3] (one may also refer to the proof of Corollary A.42 of [5]) we have for any $g(x) \in \{g_1(x), \dots, g_k(x)\}$,

$$\begin{aligned}
&\left(\int g_j(x)\,\mathrm{d}\tilde{G}_n(x) - \int g_j(x)\,\mathrm{d}\hat{G}_n(x)\right) \\
&\quad\le M \sum_{k=1}^{n} \bigl|\lambda_k^{\tilde{\mathcal{S}}} - \lambda_k^{\hat{\mathcal{S}}}\bigr| \\
&\quad\le \frac{M}{\sqrt{n}} \bigl[\operatorname{tr}\bigl((\hat{\mathbf{X}}_n - \hat{\mathbf{B}}) - (\tilde{\mathbf{X}}_n - \tilde{\mathbf{B}})\bigr)\bigl((\hat{\mathbf{X}}_n - \tilde{\mathbf{X}}_n) - (\hat{\mathbf{B}} - \tilde{\mathbf{B}})\bigr)^T\bigr]^{1/2} \bigl[\operatorname{tr}(\tilde{\mathcal{S}} + \hat{\mathcal{S}})\bigr]^{1/2} \\
&\quad= M\sigma_n \left(1 - \frac{1}{\sigma_n}\right) [\operatorname{tr}\tilde{\mathcal{S}}]^{1/2} \bigl[\operatorname{tr}(\tilde{\mathcal{S}} + \hat{\mathcal{S}})\bigr]^{1/2} \\
&\quad\le M\bigl(\sigma_n^2 - 1\bigr)\bigl[n\lambda_{\max}(\tilde{\mathcal{S}})\bigr]^{1/2} \bigl[n\lambda_{\max}(\tilde{\mathcal{S}}) + n\lambda_{\max}(\hat{\mathcal{S}})\bigr]^{1/2},
\end{aligned}$$

where the third step uses the fact that by the structure of $(\hat{\mathbf{X}}_n - \hat{\mathbf{B}})$,

$$\frac{1}{n}\bigl((\hat{\mathbf{X}}_n - \hat{\mathbf{B}}) - (\tilde{\mathbf{X}}_n - \tilde{\mathbf{B}})\bigr)\bigl((\hat{\mathbf{X}}_n - \tilde{\mathbf{X}}_n) - (\hat{\mathbf{B}} - \tilde{\mathbf{B}})\bigr)^T = \sigma_n^2 \left(1 - \frac{1}{\sigma_n}\right)^2 \tilde{\mathcal{S}}.$$

From [23] and $\|\mathbf{T}_n\| \le M$ we obtain $\limsup_n \lambda_{\max}(\tilde{\mathcal{S}})$ is bounded by $M(1 + \sqrt{c})^2$ with probability one. Similarly, $\limsup_n \lambda_{\max}(\hat{\mathcal{S}})$ is bounded by $M(1 + \sqrt{c})^2$ with probability one by the structure of $\hat{\mathcal{S}}$ and estimate (2.20) below. Moreover it is not difficult to prove that

$$\sigma_n^2 - 1 = \mathrm{o}\bigl(\varepsilon_n^2 n^{-1}\bigr). \tag{2.20}$$

Summarizing the above we obtain

$$\int g(x)\,\mathrm{d}\hat{G}_n(x) - \int g(x)\,\mathrm{d}\tilde{G}_n(x) \xrightarrow{\text{i.p.}} 0.$$

This, together with (2.19), implies that

$$\int g(x)\,\mathrm{d}G_n(x) - \int g(x)\,\mathrm{d}\tilde{G}_n(x) \xrightarrow{\text{i.p.}} 0.$$

Therefore, in what follows, we may assume that

$$|X_{ij}| \le \sqrt{n}\varepsilon_n, \qquad EX_{ij} = 0, \qquad E|X_{ij}|^2 = 1 \tag{2.21}$$

and for simplicity we shall suppress all subscripts or superscripts on the variables.

As in [3], by Cauchy's formula, (1.6) and (1.7), with probability one for all n large,

$$\int g(x)\,\mathrm{d}G_n(x) = -\frac{1}{2\pi\mathrm{i}} \int g(z)\bigl(\operatorname{tr}(\mathcal{S} - z\mathbf{I})^{-1} - nm_{c_n}(z)\bigr)\,\mathrm{d}z, \tag{2.22}$$

where $m_{c_n}(z)$ is obtained from $m(z)$ with c replaced by c_n and the complex integral is over $\mathcal{C}$. The contour $\mathcal{C}$ is specified below. Let $v_0 > 0$ be arbitrary and set $\mathcal{C}_u = \{u + \mathrm{i}v_0, u \in [u_l, u_r]\}$, where $u_r > \limsup_n \lambda_{\max}(\mathbf{T}_n)(1 + \sqrt{c})^2$ and $0 < u_l < \liminf_n \lambda_{\min}(\mathbf{T}_n) I_{(0,1)}(c)(1 - \sqrt{c})^2$ or u_l is any negative number if $\liminf_n \lambda_{\min}(\mathbf{T}_n) I_{(0,1)}(c)(1 - \sqrt{c})^2 = 0$. Then define

$$\mathcal{C}^+ = \bigl\{u_l + \mathrm{i}v\colon v \in [0, v_0]\bigr\} \cup \mathcal{C}_u \cup \bigl\{u_r + \mathrm{i}v\colon v \in [0, v_0]\bigr\}$$

and let $\mathcal{C}^-$ be the symmetric part of $\mathcal{C}^+$ about the real axis. Then set $\mathcal{C} = \mathcal{C}^+ \cup \mathcal{C}^-$. Moreover, since

$$\operatorname{tr}(\mathcal{S} - z\mathbf{I})^{-1} = \operatorname{tr}\mathbf{A}^{-1}(z) + \frac{\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}}{1 - \bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}},$$

we have

$$\bigl(\operatorname{tr}(\mathcal{S} - z\mathbf{I})^{-1} - nm_{c_n}(z)\bigr) = \bigl[\operatorname{tr}\mathbf{A}^{-1}(z) - nm_{c_n}(z)\bigr] + \frac{\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}}{1 - \bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}}, \tag{2.23}$$

where $\mathbf{A}^{-1}(z) = (\mathbf{S} - zI)^{-1}$. The first term on the right-hand side above was investigated in [3].

We next consider the second term on the right-hand side in (2.23). To this end, introduce a truncation version of $\bar{s}^T\mathbf{A}^{-2}(z)\bar{s}$. Define $\mathcal{C}_r = \{u_r + \mathrm{i}v\colon\ v \in [n^{-1}\rho_n, v_0]\}$,

$$\mathcal{C}_l = \begin{cases} \{u_l + \mathrm{i}v\colon\ v \in [n^{-1}\rho_n, v_0\} & \text{if } u_l > 0, \\ \{u_l + \mathrm{i}v\colon\ v \in [0, v_0]\} & \text{if } u_l < 0, \end{cases}$$

where

$$\rho_n \downarrow 0, \qquad \rho_n \ge n^{-\alpha}, \quad \text{for some } \alpha \in (0,1). \tag{2.24}$$

Let $\mathcal{C}_n^+ = \mathcal{C}_l \cup \mathcal{C}_u \cup \mathcal{C}_r$ and $\mathcal{C}_n^-$ denote the symmetric part of $\mathcal{C}_n^+$ with respect to the real axis. We then define the truncated process $\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-2}(z)}\bar{\mathbf{s}}$ of the process $\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}$ for $z = u + \mathrm{i}v$ by

$$\bar{s}^T\widehat{\mathbf{A}^{-2}(z)}\bar{s} = \begin{cases} \bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{s} & \text{for } z \in \mathcal{C}_n = \mathcal{C}_n^+ \cup \mathcal{C}_n^-, \\ \frac{nv+\rho_n}{2\rho_n}\bar{s}^T\mathbf{A}^{-2}(z_{r_1})\bar{s} + \frac{\rho_n - nv}{2\rho_n}\bar{s}^T\mathbf{A}^{-2}(z_{r_2})\bar{s} & \text{for } u = u_r,\ v \in \bigl[-n^{-1}\rho_n, n^{-1}\rho_n\bigr], \\ \frac{nv+\rho_n}{2\rho_n}\bar{s}^T\mathbf{A}^{-2}(z_{l_1})\bar{s} + \frac{\rho_n - nv}{2\rho_n}\bar{s}^T\mathbf{A}^{-2}(z_{l_2})\bar{s} & \text{for } u = u_l > 0,\ v \in \bigl[-n^{-1}\rho_n, n^{-1}\rho_n\bigr], \end{cases}$$

where $z_{r_1} = u_r - \mathrm{i}n^{-1}\rho_n$, $z_{r_2} = u_r + \mathrm{i}n^{-1}\rho_n$, $z_{l_1} = u_l - \mathrm{i}n^{-1}\rho_n$ and $z_{l_2} = u_l + \mathrm{i}n^{-1}\rho_n$. Similarly one may define the truncation version $\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-1}(z)}\bar{\mathbf{s}}$ of $\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}$.

From Theorem 2 in [15] we have

$$\|\bar{\mathbf{s}}^T\bar{\mathbf{s}}\| \le \|\mathbf{T}_n\| \|\bar{\mathbf{x}}^T\bar{\mathbf{x}}\| \le M\|\bar{\mathbf{x}}^T\bar{\mathbf{x}}\| \xrightarrow{\text{i.p.}} Mc. \tag{2.25}$$

Note that

$$\bar{\mathbf{s}}^T(\mathcal{S} - zI)^{-1}\bar{\mathbf{s}} = \bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}} + \frac{(\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}})^2}{1 - \bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}},$$

where we use the identity

$$\mathbf{r}^T\bigl(\mathbf{D} + a\mathbf{r}\mathbf{r}^T\bigr)^{-1} = \frac{r^T\mathbf{D}^{-1}}{1 + a\mathbf{r}^T\mathbf{D}^{-1}\mathbf{r}}, \tag{2.26}$$

where $\mathbf{D}$ and $\mathbf{D} + a\mathbf{r}\mathbf{r}^T$ are both invertible, $\mathbf{r} \in \mathbb{R}^p$ and $a \in \mathbb{R}$. This implies that

$$\frac{1}{1 - \bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}} = 1 + \bar{\mathbf{s}}^T(\mathcal{S} - zI)^{-1}\bar{\mathbf{s}}. \tag{2.27}$$

We then conclude that

$$\left|\frac{1}{1 - \bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}}\right| \le 1 + \left(\frac{M}{v_0} + \frac{M}{|\lambda_{\max}(\mathcal{S}) - u_r|} + \frac{M}{|\lambda_{\min}(\mathcal{S}) - u_l|}\right)\bar{\mathbf{s}}^T\bar{\mathbf{s}}. \tag{2.28}$$

This, together with (2.25), (1.6) and (1.7), ensures that

$$\begin{aligned}
&\left|\int g(z)\left(\frac{\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-2}(z)}\bar{\mathbf{s}}}{1-\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-1}(z)}\bar{\mathbf{s}}}-\frac{\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}}{1-\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}}\right)\mathrm{d}z\right|\\
&\quad\le\left|\int g(z)\left(\frac{(\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-2}(z)}\bar{\mathbf{s}}-\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}})}{1-\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-1}(z)}\bar{\mathbf{s}}}\right)\mathrm{d}z\right|\\
&\qquad+\left|\int g(z)\left(\frac{\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}(\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-1}(z)}\bar{\mathbf{s}}-\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}})}{(1-\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}})(1-\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-1}(z)}\bar{\mathbf{s}})}\right)\mathrm{d}z\right|\\
&\quad\le\frac{[(\bar{\mathbf{s}}^T\bar{\mathbf{s}})^2+(\bar{\mathbf{s}}^T\bar{\mathbf{s}})^4]M\rho_n}{n}\left(\frac{1}{|\lambda_{\max}(\mathcal{S})-u_r|}+\frac{1}{|\lambda_{\min}(\mathcal{S})-u_l|}\right)\left(\frac{1}{|\lambda_{\max}(\mathbf{S})-u_r|}+\frac{1}{|\lambda_{\min}(\mathbf{S})-u_l|}\right),
\end{aligned}\tag{2.29}$$

converging to zero in probability.

The aim of subsequent Sections 2.2–2.4 is to prove convergence of $\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-j}(z)}\bar{\mathbf{s}}$, $j=1,2$, in probability to nonrandom variables in the space of continuous functions for $z\in\mathcal{C}$. It is well known that convergence in probability is equivalent to convergence in distribution when the limiting variable is nonrandom (for example see page 25 of [8]). Thus, instead, we shall prove convergence of $\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-j}(z)}\bar{\mathbf{s}}$, $j=1,2$, in distribution to nonrandom variables in the space of continuous functions for $z\in\mathcal{C}$.

2.2. *Convergence of finite-dimensional distributions of* $\bar{\mathbf{s}}^T\mathbf{A}^{-j}(z)\bar{\mathbf{s}}-E(\bar{\mathbf{s}}^T\mathbf{A}^{-j}(z)\bar{\mathbf{s}})$, $j=1,2$

Consider $z\in\mathcal{C}_u$ and $\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}-E(\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}})$ first. Define the σ-field $\mathcal{F}_j=\sigma(\mathbf{s}_1,\dots,\mathbf{s}_j)$, and let $E_j(\cdot)=E(\cdot|\mathcal{F}_j)$ and $E_0(\cdot)$ be the unconditional expectation. Introduce $\bar{\mathbf{s}}_j=\bar{\mathbf{s}}-n^{-1}\mathbf{s}_j$,

$$\mathbf{A}_j^{-1}(z)=\left(\mathbf{S}-n^{-1}\mathbf{s}_j\mathbf{s}_j^T-z\mathbf{I}\right)^{-1},\qquad \gamma_j(z)=\frac{1}{n}\mathbf{s}_j^T\mathbf{A}_j^{-1}(z)\mathbf{s}_j-E\frac{1}{n}\operatorname{tr}\mathbf{A}_j^{-1}(z)\mathbf{T}_n,$$

$$\beta_j(z)=\frac{1}{1+(1/n)\mathbf{s}_j^T\mathbf{A}_j^{-1}(z)\mathbf{s}_j},\qquad b_1(z)=\frac{1}{1+(1/n)Etr\mathbf{A}_1^{-1}(z)\mathbf{T}_n}.$$

Write

$$\begin{aligned}
\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}-E\left(\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}\right)&=\sum_{j=1}^n E_j\left(\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}\right)-E_{j-1}\left(\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}\right)\\
&=\sum_{j=1}^n(E_j-E_{j-1})\left[\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}-\bar{\mathbf{s}}_j^T\mathbf{A}_j^{-2}(z)\bar{\mathbf{s}}_j\right]\\
&=\sum_{j=1}^n(E_j-E_{j-1})[d_{n1}+d_{n2}+d_{n3}],
\end{aligned}$$

where

$$d_{n1}=(\bar{\mathbf{s}}-\bar{\mathbf{s}}_j)^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}},\qquad d_{n2}=\bar{\mathbf{s}}_j^T\left(\mathbf{A}^{-2}(z)-\mathbf{A}_j^{-2}(z)\right)\bar{\mathbf{s}},$$

and

$$d_{n3}=\bar{\mathbf{s}}_j^T\mathbf{A}_j^{-2}(z)(\bar{\mathbf{s}}-\bar{\mathbf{s}}_j).$$

Consider d_{n3} first. It is proved in (3.12) of [15] that

$$E\left|\mathbf{x}_1^T\mathbf{B}\bar{\mathbf{x}}_1\right|^k=\mathrm{O}\left(n^{(k/2)-2}\varepsilon_n^{k-4}\right)\quad\text{for } k\geq 4, \tag{2.30}$$

where $\mathbf{x}_1$ and $\bar{\mathbf{x}}_j=\bar{\mathbf{x}}-n^{-1}\mathbf{x}_j$ are independent of $\mathbf{B}$ with a bounded spectral norm. This implies that

$$E\left|\mathbf{s}_1^T\mathbf{A}_j^{-2}(z)\bar{\mathbf{s}}_1\right|^k=E\left|\mathbf{x}_1^T\mathbf{T}_n^{1/2}\mathbf{A}_j^{-2}(z)\mathbf{T}_n^{1/2}\bar{\mathbf{x}}_1\right|^k=\mathrm{O}\left(n^{(k/2)-2}\varepsilon_n^{k-4}\right)\quad\text{for } k\geq 4, \tag{2.31}$$

where yields

$$\sum_{j=1}^n(E_j-E_{j-1})d_{n3}\xrightarrow{\text{i.p.}}0.$$

Furthermore, write

$$d_{n1}=d_{n1}^{(1)}+d_{n1}^{(2)}+d_{n1}^{(3)}+d_{n1}^{(4)},$$

where

$$d_{n1}^{(1)}=\frac{1}{n}\mathbf{s}_j^T\mathbf{A}_j^{-2}(z)\bar{\mathbf{s}}_j,\qquad d_{n1}^{(2)}=\frac{1}{n^2}\mathbf{s}_j^T\mathbf{A}_j^{-2}(z)\mathbf{s}_j$$

and

$$d_{n1}^{(3)}=\frac{1}{n}\mathbf{s}_j^T\left(\mathbf{A}^{-2}(z)-\mathbf{A}_j^{-2}(z)\right)\bar{\mathbf{s}}_j,\qquad d_{n1}^{(4)}=\frac{1}{n^2}\mathbf{s}_j^T\left(\mathbf{A}^{-2}(z)-\mathbf{A}_j^{-2}(z)\right)\mathbf{s}_j.$$

From (2.31) we have

$$\sum_{j=1}^n(E_j-E_{j-1})d_{n1}^{(1)}\xrightarrow{\text{i.p.}}0.$$

By Lemma 2.7 in [3] and (2.21)

$$n^{-k}E\left|\mathbf{s}_1^T\mathbf{B}(z)\mathbf{s}_1-\operatorname{tr}\mathbf{BT}\right|^k=\mathrm{O}\left(\varepsilon_n^{2k-4}n^{-1}\right),\quad k\geq 2. \tag{2.32}$$

This gives

$$E\left|\sum_{j=1}^n(E_j-E_{j-1})d_{n1}^{(2)}\right|^2=\frac{1}{n^4}\sum_{j=1}^nE\left|(E_j-E_{j-1})\left(\mathbf{s}_j^T\mathbf{A}_j^{-2}(z)\mathbf{s}_j-\operatorname{tr}\mathbf{T}_n\mathbf{A}_j^{-2}(z)\right)\right|^2=\mathrm{O}\left(n^{-2}\right).$$

Since $\|\mathbf{A}^{-2}(z)\|\leq 1/v^2$ we obtain

$$E\left|\sum_{j=1}^n(E_j-E_{j-1})d_{n1}^{(4)}\right|^2\leq\frac{M}{n^4v^4}\sum_{j=1}^nE\|\mathbf{s}_j\|^4\leq\frac{M}{n}.$$

Note that

$$\mathbf{A}^{-1}(z)-\mathbf{A}_j^{-1}(z)=-\frac{1}{n}\mathbf{A}_j^{-1}(z)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-1}(z)\beta_j(z). \tag{2.33}$$

Applying it we may write

$$\begin{aligned}d_{n1}^{(3)}=&-\frac{1}{n^2}\mathbf{s}_j^T\mathbf{A}_j^{-1}(z)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-2}(z)\bar{\mathbf{s}}_j\beta_j(z)+\frac{1}{n^3}\mathbf{s}_j^T\mathbf{A}_j^{-1}(z)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-2}(z)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-1}(z)\bar{\mathbf{s}}_j\beta_j^2(z)\\&-\frac{1}{n^2}\mathbf{s}_j^T\mathbf{A}_j^{-2}(z)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-1}(z)\bar{\mathbf{s}}_j\beta_j(z).\end{aligned}$$

It follows from (2.31), (2.32) and the fact that $|\beta_j(z)| \le 1/v$ that

$$E\left|\sum_{j=1}^{n}(E_j - E_{j-1})d_{n1}^{(3)}\right|^2 \le \frac{M}{n}.$$

So far we have proved that

$$\sum_{j=1}^{n}(E_j - E_{j-1})d_{n1} \xrightarrow{\text{i.p.}} 0.$$

As in dealing with d_{n1}, one may verify that

$$\sum_{j=1}^{n}(E_j - E_{j-1})d_{n2} \xrightarrow{\text{i.p.}} 0.$$

Summarizing the above we have obtained

$$\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}} - E\big(\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}\big) \xrightarrow{\text{i.p.}} 0.$$

Applying the argument for $\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}$ to $\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}$ yields

$$\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}} - E\big(\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}\big) \xrightarrow{\text{i.p.}} 0.$$

2.3. *Tightness of* $(\bar{\mathbf{s}}^T\mathbf{A}^{-j}(z)\bar{\mathbf{s}} - E(\bar{\mathbf{s}}^T\mathbf{A}^{-j}(z)\bar{\mathbf{s}}))$, $j = 1, 2$

This section is to prove tightness of $K_n^{(j)}(z)$, $j = 1, 2$ where $K_n^{(j)}(z) = \bar{\mathbf{s}}^T\mathbf{A}^{-j}(z)\bar{\mathbf{s}} - E(\bar{\mathbf{s}}^T\mathbf{A}^{-j}(z)\bar{\mathbf{s}})$. We shall use Theorem 12.3 of [8]. In what follows, we consider the case $j = 2$ only and the case $j = 1$ can be handled similarly, even simpler. As pointed out in [3], condition (i) of Theorem 12.3 of [8] can be replaced with the assumption of tightness at any point in the interval. From (3.18) of [15] we see that

$$E\left|\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}\right|^2 \le M, \tag{2.34}$$

which ensures condition (i) of Theorem 12.3 of [8]. We next verify condition (ii) of Theorem 12.3 of [8] by proving

$$\sup_{n, z_1, z_2 \in \in \mathcal{C}_n} \frac{E|K_n^{(2)}(z_1) - K_n^{(2)}(z_2)|^2}{|z_1 - z_2|^2} < \infty. \tag{2.35}$$

Below we only prove the above inequality on $\mathcal{C}_n^+$ and the remaining cases can be verified similarly.

Since the truncation steps are the same as those in [23] we have

$$P\big(\|\lambda_{\max}(\mathbf{S})\| \ge \zeta_r\big) = \mathrm{o}\big(n^{-l}\big), \qquad P\big(\lambda_{\min}(\mathbf{S}) \le \zeta_l\big) = \mathrm{o}\big(n^{-l}\big) \tag{2.36}$$

for any l (see (1.9a) and (1.9b) in [3]), where $\limsup_n \lambda_{\max}(\mathbf{T}_n)(1 + \sqrt{c})^2 < \zeta_r$ and $0 < \zeta_l < \liminf_n \lambda_{\min}(\mathbf{T}_n)(1 - \sqrt{c})^2$. It is also proved in section 3 of [3] that for any positive k on $\mathcal{C}_n^+$

$$\max\big(E\big\|\mathbf{A}^{-1}(z)\big\|^k, E\big\|\mathbf{A}_j^{-1}(z)\big\|^k\big) \le M. \tag{2.37}$$

As in the last section we write

$$\begin{aligned}\frac{K_n^{(2)}(z_1) - K_n^{(2)}(z_2)}{z_1 - z_2} &= \big[\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}} - E\big(\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}}\big)\big] \\ &\quad + \big[\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z_1)\mathbf{A}^{-2}(z_2)\bar{\mathbf{s}} - E\big(\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z_1)\mathbf{A}^{-2}(z_2)\bar{\mathbf{s}}\big)\big].\end{aligned}$$

Since the above two terms on the right-hand side are similar we only prove the tightness of the first term. To this end, write

$$\begin{aligned}
&\bar{\mathbf{s}}^T \mathbf{A}^{-2}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}} - E\big(\bar{\mathbf{s}}^T \mathbf{A}^{-2}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}}\big)\\
&\quad = \sum_{j=1}^{n}(E_j - E_{j-1})\big[\bar{\mathbf{s}}^T \mathbf{A}^{-2}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}} - \bar{\mathbf{s}}_j^T \mathbf{A}_j^{-2}(z_1)\mathbf{A}_j^{-1}(z_2)\bar{\mathbf{s}}_j\big]\\
&\quad = \sum_{j=1}^{n}(E_j - E_{j-1})[a_{n1} + a_{n2} + a_{n3}],
\end{aligned}$$

where

$$a_{n1} = (\bar{\mathbf{s}} - \bar{\mathbf{s}}_j)^T \mathbf{A}^{-2}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}}, \qquad a_{n2} = \bar{\mathbf{s}}_j^T\big(\mathbf{A}^{-2}(z_1)\mathbf{A}^{-1}(z_2) - \mathbf{A}_j^{-2}(z_1)\mathbf{A}_j^{-1}(z_2)\big)\bar{\mathbf{s}}$$

and

$$a_{n3} = \bar{\mathbf{s}}_j^T \mathbf{A}_j^{-2}(z_1)\mathbf{A}_j^{-1}(z_2)(\bar{\mathbf{s}} - \bar{\mathbf{s}}_j).$$

It is straightforward to check that $E\|\mathbf{s}_j^T\|^4 = E(\mathbf{s}_j^T\mathbf{s}_j)^2 = \mathrm{O}(n^2)$. As in (3.17) of [15] one can prove that

$$E\big|\bar{\mathbf{s}}^T\bar{\mathbf{s}}\big|^k = E\big|\bar{\mathbf{x}}^T\mathbf{T}_n\bar{\mathbf{x}}\big|^k \le ME\big|\bar{\mathbf{x}}^T\bar{\mathbf{x}}\big|^k = \mathrm{O}(1) \quad \text{for } k \ge 4. \tag{2.38}$$

This, together with (2.37), implies that

$$\begin{aligned}
E\Big|\sum_{j=1}^{n}(E_j - E_{j-1})(a_{n1})\Big|^2 &\le \frac{M}{n^2}\sum_{j=1}^{n} E\big\|\mathbf{s}_j^T \mathbf{A}^{-2}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}}\big\|^2\\
&\le \frac{M}{n^2}\sum_{j=1}^{n}\big(E\|\mathbf{s}_j^T\|^4\big)^{1/2}\big(E\big\|\mathbf{A}^{-2}(z_1)\big\|^{12}E\big\|\mathbf{A}^{-1}(z_2)\big\|^{12}E\|\bar{\mathbf{s}}\|^{12}\big)^{1/6} \le M.
\end{aligned}$$

This argument also works for a_{n3}.

We further write

$$a_{n2} = a_{n2}^{(1)} + a_{n2}^{(2)} + a_{n2}^{(3)},$$

where

$$\begin{aligned}
a_{n2}^{(1)} &= \bar{\mathbf{s}}_j^T\big(\mathbf{A}^{-1}(z_1) - \mathbf{A}_j^{-1}(z_1)\big)\mathbf{A}^{-1}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}}\\
&= -\frac{1}{n}\bar{\mathbf{s}}_j^T\mathbf{A}_j^{-1}(z_1)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-1}(z_1)\mathbf{A}^{-1}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}}\beta_j(z),\\
a_{n2}^{(2)} &= \bar{\mathbf{s}}_j^T\mathbf{A}_j^{-1}(z_1)\big(\mathbf{A}^{-1}(z_1) - \mathbf{A}_j^{-1}(z_1)\big)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}}\\
&= -\frac{1}{n}\bar{\mathbf{s}}_j^T\mathbf{A}_j^{-2}(z_1)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-1}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}}\beta_j(z)
\end{aligned}$$

and

$$\begin{aligned}
a_{n2}^{(3)} &= \bar{\mathbf{s}}_j^T\mathbf{A}_j^{-2}(z_1)\big(\mathbf{A}^{-1}(z_2) - \mathbf{A}_j^{-1}(z_2)\big)\bar{\mathbf{s}}\\
&= -\frac{1}{n}\bar{\mathbf{s}}_j^T\mathbf{A}_j^{-2}(z_1)\mathbf{A}_j^{-1}(z_2)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-1}(z_2)\bar{\mathbf{s}}\beta_j(z).
\end{aligned}$$

In the second equality of each $a_{n2}^{(j)}$, $j = 1, 2, 3$ above we also use (2.33). Note that (one may see the remark to (3.2) in [3])

$$\left|\beta_j(z)\right| = \left|1 - n^{-1}\mathbf{s}_j^T \mathbf{A}^{-1}(z)\mathbf{s}_j\right| \le 1 + M\eta_r + Mn^{2+\alpha} I\left(\|\mathbf{S}\| \ge \eta_r \text{ or } \lambda_{\min}(\mathbf{S}) \le \eta_l\right), \tag{2.39}$$

where $\limsup_n \lambda_{\max}(\mathbf{T}_n)(1+\sqrt{c})^2 < \eta_r < u_r$ and $u_l < \eta_l < \limsup_n \lambda_{\min}(\mathbf{T}_n) I_{(0,1)}(c)(1-\sqrt{c})^2$. Also we have

$$\begin{aligned} n^{-1/2}\left|\mathbf{s}_1^T \mathbf{A}_1^{-1}(z_1)\mathbf{A}^{-1}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}}\right| &\le n^{-1/2}\left\|\mathbf{s}_1^T\right\| \|\bar{\mathbf{s}}\| \left\|\mathbf{A}_1^{-1}(z_1)\mathbf{A}^{-1}(z_1)\mathbf{A}^{-1}(z_2)\right\| \\ &\le M\eta_r \|\bar{\mathbf{s}}\| + Mn^{6+4\alpha} I\left(\|\mathbf{S}\| \ge \eta_r \text{ or } \lambda_{\min}(\mathbf{S}_1) \le \eta_l\right), \end{aligned} \tag{2.40}$$

where we use (2.24) and the fact that $n^{-1/2}\|\mathbf{s}_1\|^2$ is smaller than η_r if $\|\mathbf{S}\| < \eta_r$, and $n^{-1/2}\|\mathbf{s}_1\|^2 \le n$ otherwise. Here $\mathbf{S}_1 = \mathbf{S} - n^{-1}\mathbf{s}_1\mathbf{s}_1^T$. One should note that (2.31) still holds when $\mathbf{B}$ is replaced by $\mathbf{A}_j^{-1}(z)$ or $\mathbf{A}_j^{-2}(z)$ due to (2.37). We then conclude from (2.36), (2.39), (2.31), (2.38) and (2.40) that

$$\begin{aligned} E\left|\sum_{j=1}^n (E_j - E_{j-1})\left(a_{n2}^{(1)}\right)\right|^2 &\le M\sum_{j=1}^n E\left|a_{n2}^{(1)}\right|^2 \\ &\le M\left(E\left|\bar{\mathbf{s}}_j^T \mathbf{A}_j^{-1}(z_1)\mathbf{s}_j\right|^4 E\|\bar{\mathbf{s}}\|^4\right)^{1/2} \\ &\quad + Mn^{20+12\alpha} P\left(\|\mathbf{S}\| \ge \eta_r \text{ or } \lambda_{\min}(\mathbf{S}_1) \le \eta_l\right) \le M. \end{aligned}$$

This argument apparently applies to the terms $a_{n2}^{(j)}$, $j = 2, 3$. Therefore (2.35) is satisfied and $K_n^{(2)}(z)$ is tight.

2.4. *Convergence of* $E(\bar{\mathbf{s}}^T \mathbf{A}^{-2}(z)\bar{\mathbf{s}})$ *and* $E(\bar{\mathbf{s}}^T \mathbf{A}^{-1}(z)\bar{\mathbf{s}})$

The aim in this section is to find the limits of $E(\bar{\mathbf{s}}^T \mathbf{A}^{-2}(z)\bar{\mathbf{s}})$ and $E(\bar{\mathbf{s}}^T \mathbf{A}^{-1}(z)\bar{\mathbf{s}})$. In what follows, all bounds including $\mathrm{O}(\cdot)$ and $\mathrm{o}(\cdot)$ expressions hold uniformly on $\mathcal{C}_n$.

Consider $E(\bar{\mathbf{s}}^T \mathbf{A}^{-2}(z)\bar{\mathbf{s}})$ first. Applying $\bar{\mathbf{s}} = \frac{1}{n}\sum_{j=1}^n \mathbf{s}_j$, (2.26) and (2.33) we obtain

$$\begin{aligned} E\left[\bar{\mathbf{s}}^T \mathbf{A}^{-2}(z)\bar{\mathbf{s}}\right] &= \frac{1}{n}\sum_{j=1}^n E\left[\mathbf{s}_j^T \mathbf{A}_j^{-1}(z)\mathbf{A}^{-1}(z)\bar{\mathbf{s}}\beta_j(z)\right] \\ &= \frac{1}{n}\sum_{j=1}^n E\left[\mathbf{s}_j^T \mathbf{A}_j^{-2}(z)\bar{\mathbf{s}}\beta_j(z)\right] - \frac{1}{n^2}\sum_{j=1}^n E\left[\mathbf{s}_j^T \mathbf{A}_j^{-2}(z)\mathbf{s}_j\mathbf{s}_j^T \mathbf{A}_j^{-1}(z)\bar{\mathbf{s}}\beta_j^2(z)\right] \\ &= q_{n1} + q_{n2} + q_{n3} + q_{n4}, \end{aligned}$$

where

$$q_{n1} = \frac{1}{n}\sum_{j=1}^n E\left[\mathbf{s}_j^T \mathbf{A}_j^{-2}(z)\bar{\mathbf{s}}_j\beta_j(z)\right], \qquad q_{n3} = -\frac{1}{n^2}\sum_{j=1}^n E\left[\mathbf{s}_j^T \mathbf{A}_j^{-2}(z)\mathbf{s}_j\mathbf{s}_j^T \mathbf{A}_j^{-1}(z)\bar{\mathbf{s}}_j\beta_j^2(z)\right]$$

and

$$q_{n2} = \frac{1}{n^2}\sum_{j=1}^n E\left[\mathbf{s}_j^T \mathbf{A}_j^{-2}(z)\mathbf{s}_j\beta_j(z)\right], \qquad q_{n4} = -\frac{1}{n^3}\sum_{j=1}^n E\left[\mathbf{s}_j^T \mathbf{A}_j^{-2}(z)\mathbf{s}_j\mathbf{s}_j^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j\beta_j^2(z)\right].$$

Using

$$\beta_j(z) = b_1(z) - \beta_j(z)b_1(z)\gamma_j(z) \tag{2.41}$$

we further obtain

$$q_{n1} = -\frac{1}{n}\sum_{j=1}^{n} E\big[\mathbf{s}_j^T\mathbf{A}_j^{-2}(z)\bar{\mathbf{s}}_j\beta_j(z)b_1(z)\gamma_j(z)\big].$$

Moreover, it is proved in [3] (see (3.5) and three lines below it in [3]) that

$$E\big|\gamma_1(z)\big|^k \le Mn^{-1}\varepsilon_n^{2k-4}, \qquad E\big|\beta_1(z)\big|^k \le M, \quad k\ge 2, \qquad \big|b_1(z)\big| \le M. \tag{2.42}$$

We then conclude from (2.39), (2.31), (2.40) and (2.42) that

$$|q_{n1}| \le M\big(E\big|\mathbf{s}_1^T\mathbf{A}_1^{-2}(z)\bar{\mathbf{s}}_1\big|^2 E\big|\gamma_1(z)\big|^2\big)^{1/2} + Mn^{7+4\alpha}P\big(\|\mathbf{S}\| \ge \eta_r \text{ or } \lambda_{\min}(\mathbf{S}_1) \le \eta_l\big) \le \frac{M}{\sqrt{n}}. \tag{2.43}$$

Applying a similar approach one can prove that

$$|q_{n3}| \le \frac{M}{\sqrt{n}}.$$

By (2.41) we have

$$\begin{aligned} q_{n2} &= b_1(z)E\left[\frac{1}{n}\operatorname{tr}\mathbf{A}_1^{-2}(z)\mathbf{T}_n\right] + \frac{b_1(z)}{n^2}\sum_{j=1}^{n}E\big[\mathbf{s}_j^T\mathbf{A}_j^{-2}(z)\mathbf{s}_j\beta_j(z)\gamma_j(z)\big] \\ &= b_1(z)E\left[\frac{1}{n}\operatorname{tr}\mathbf{A}_1^{-2}(z)\mathbf{T}_n\right] + \mathrm{O}\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \tag{2.44}$$

where the last step uses the fact that the absolute value of the second term of q_{n2} is bounded by $M/\sqrt{n}$ by the same approach as that used in (2.43).

From (2.41) we may write

$$q_{n4} = -b_1^2(z)E\left[\frac{1}{n}\operatorname{tr}\mathbf{A}_1^{-1}(z)\mathbf{T}_n\right]E\left[\frac{1}{n}\operatorname{tr}\mathbf{A}_1^{-2}(z)\mathbf{T}_n\right] + q_{n4}^{(1)} + q_{n4}^{(2)} + q_{n4}^{(3)},$$

where

$$q_{n4}^{(1)} = -\frac{1}{n^3}\sum_{j=1}^{n}E\big[\mathbf{s}_j^T\mathbf{A}_j^{-2}(z)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-1}(z)\mathbf{s}_j\beta_j^2(z)b_1^2(z)\gamma_j^2(z)\big],$$

$$q_{n4}^{(2)} = \frac{2}{n^3}\sum_{j=1}^{n}E\big[\mathbf{s}_j^T\mathbf{A}_j^{-2}(z)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-1}(z)\mathbf{s}_j\beta_j(z)b_1^2(z)\gamma_j(z)\big]$$

and

$$q_{n4}^{(3)} = -\frac{1}{n^2}\sum_{j=1}^{n}E\big[\big(n^{-1}\mathbf{s}_j^T\mathbf{A}_j^{-2}(z)\mathbf{s}_j - n^{-1}Etr\mathbf{A}_j^{-2}(z)\big)\gamma_j(z)b_1^2(z)\big].$$

By (3.2) in [3], as in (2.43), we may prove that $|q_{n4}^{(k)}| \le M/\sqrt{n}, k = 1, 2, 3$. It follows from that

$$q_{n4} = -b_1^2(z)E\left[\frac{1}{n}\operatorname{tr}\mathbf{A}_1^{-1}(z)\mathbf{T}_n\right]E\left[\frac{1}{n}\operatorname{tr}\mathbf{A}_1^{-2}(z)\mathbf{T}_n\right] + \mathrm{O}\left(\frac{1}{\sqrt{n}}\right). \tag{2.45}$$

This, together with (2.44), (2.43), (2.33) and (2.42), yields

$$E\big[\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}\big]=q_{n1}+q_{n2}+q_{n2}+q_{n4}=b_1^2(z)E\bigg[\frac{1}{n}\operatorname{tr}\mathbf{A}^{-2}(z)\mathbf{T}_n\bigg]+\mathrm{O}\bigg(\frac{1}{\sqrt{n}}\bigg). \tag{2.46}$$

It was proved in Section 4 of [3] (see three lines above (4.14) there) that

$$b_1(z)+z\underline{m}(z)\longrightarrow 0. \tag{2.47}$$

Consider $E[\frac{1}{n}\operatorname{tr}\mathbf{A}^{-2}(z)\mathbf{T}_n]$ now. To this end we need formula (4.13) in [3], which states

$$\mathbf{A}^{-1}(z)=-\mathbf{H}^{-1}(z)+b_1(z)A(z)+B(z)+C(z), \tag{2.48}$$

where

$$\mathbf{H}^{-1}(z)=\big(z\mathbf{I}-b_1(z)\mathbf{T}_n\big)^{-1},\qquad A(z)=\sum_{j=1}^n\mathbf{H}^{-1}(z)\big(\mathbf{s}_j\mathbf{s}_j^*-n^{-1}\mathbf{T}_n\big)\mathbf{A}_j^{-1}(z),$$

$$B(z)=\sum_{j=1}^n\big(\beta_j(z)-b_1(z)\big)\mathbf{H}^{-1}(z)\mathbf{s}_j\mathbf{s}_j^*\mathbf{A}_j^{-1}(z)$$

and

$$C(z)=n^{-1}b_1(z)\mathbf{H}^{-1}(z)\mathbf{T}_n\sum_{j=1}^n\beta_j(z)\mathbf{A}_j^{-1}(z)\mathbf{s}_j\mathbf{s}_j^*\mathbf{A}_j^{-1}(z).$$

It was proved in (4.15) and (4.16) of [3] that

$$\frac{1}{n}\operatorname{tr}B(z)\mathbf{M}\le Cn^{-1/2}\big(E\|\mathbf{M}\|^4\big)^{1/4},\qquad \frac{1}{n}\operatorname{tr}C(z)\mathbf{M}\le Cn^{-1}\big(E\|\mathbf{M}\|^4\big)^{1/4}. \tag{2.49}$$

Taking the matrix $\mathbf{M}=\mathbf{I}$ in (4.17) and (4.19) of [3] yields

$$En^{-1}A_1(z)=-b_1(z)\big(En^{-1}\operatorname{tr}\mathbf{A}^{-1}(z)\mathbf{T}_n\mathbf{A}^{-1}(z)\mathbf{T}_n\big)\big(En^{-1}\operatorname{tr}\mathbf{A}^{-1}(z)\mathbf{H}^{-1}(z)\mathbf{T}_n\big)+\mathrm{o}(1). \tag{2.50}$$

From (2.48) and (2.49) we have

$$\begin{aligned}En^{-1}\operatorname{tr}\mathbf{A}^{-1}(z)\mathbf{H}^{-1}(z)\mathbf{T}_n&=n^{-1}\operatorname{tr}\big[\big(-\mathbf{H}^{-1}(z)+EB(z)+EC(z)\big)\mathbf{H}^{-1}(z)\mathbf{T}_n\big]\\&=-\frac{c_n}{z^2}\int\frac{t\,\mathrm{d}H_n(t)}{(1+tE\underline{m}_n)^2}+\mathrm{o}(1).\end{aligned} \tag{2.51}$$

It follows from (2.48) and (2.50) that

$$\begin{aligned}E\bigg[\frac{1}{n}\operatorname{tr}\mathbf{A}^{-2}(z)\mathbf{T}_n\bigg]&=-E\bigg[\frac{1}{n}\operatorname{tr}\mathbf{A}^{-1}(z)\mathbf{H}^{-1}(z)\mathbf{T}_n\bigg]\\&\quad-b_1^2(z)E\big[n^{-1}\operatorname{tr}\mathbf{A}^{-1}(z)\mathbf{T}_n\mathbf{A}^{-1}(z)\mathbf{T}_n\big]E\big[n^{-1}\operatorname{tr}\mathbf{A}^{-1}(z)\mathbf{H}^{-1}(z)\mathbf{T}_n\big]+\mathrm{o}(1).\end{aligned}$$

By (4.22) of [3] we obtain

$$E\big[n^{-1}\operatorname{tr}\mathbf{A}^{-1}(z)\mathbf{T}_n\mathbf{A}^{-1}(z)\mathbf{T}_n\big]=\frac{\frac{c_n}{z^2}\int\frac{t^2\,\mathrm{d}H_n(t)}{(1+tE\underline{m}_n)^2}}{1-c_n\int\frac{(E\underline{m}_n)^2t^2\,\mathrm{d}H_n(t)}{(1+tE\underline{m}_n)^2}}+\mathrm{o}(1).$$

This, together with (2.51), yields

$$E\left[\frac{1}{n}\operatorname{tr}\mathbf{A}^{-2}(z)\mathbf{T}_n\right]=\frac{\frac{c_n}{z^2}\int\frac{t\,\mathrm{d}H_n(t)}{(1+tE\underline{m}_n)^2}}{1-c_n\int\frac{(E\underline{m}_n)^2t^2\,\mathrm{d}H_n(t)}{(1+tE\underline{m}_n)^2}}+\mathrm{o}(1).$$

It was proved in (4.1) of [3] that

$$\sup_{z\in\mathcal{C}_n}\left|E\underline{m}_n(z)-\underline{m}(z)\right|\to 0.$$

This, (2.46) and (2.47) ensure that

$$E\left[\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}\right]=\frac{c\underline{m}^2(z)\int\frac{t\,\mathrm{d}H(t)}{(1+t\underline{m})^2}}{1-c\int\frac{(\underline{m})^2t^2\,\mathrm{d}H(t)}{(1+t\underline{m})^2}}+\mathrm{o}(1).$$

Consider $E(\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}})$ next. As in dealing with the term $E(\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}})$, write

$$E\left(\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}\right)=\frac{1}{n}\sum_{j=1}^n E\left[\mathbf{s}_j^T\mathbf{A}_j^{-1}(z)\bar{\mathbf{s}}_j\beta_j(z)\right]+\frac{1}{n^2}\sum_{j=1}^n E\left[\mathbf{s}_j^T\mathbf{A}_j^{-1}(z)\mathbf{s}_j\beta_j(z)\right].$$

By (2.41) and (2.42) the absolute value of the fist term on the right-hand side of the above equality converges to zero as $n\to\infty$. From (2.41), (2.42) and (2.47) the above second term becomes

$$b_1(z)E\frac{1}{n}\operatorname{tr}\mathbf{A}_1^{-1}(z)\mathbf{T}_n=1-b_1(z)\to 1+z\underline{m}(z).$$

Therefore

$$E\left(\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}\right)=1-b_1(z)\to 1+z\underline{m}(z).\tag{2.52}$$

2.5. *Convergence of* $\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}/(1-\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}})$

From sections 2.2 to 2.4 we see that after centering the stochastic processes $\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-2}(z)}\bar{\mathbf{s}}$ and $\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-1}(z)}\bar{\mathbf{s}}$ converge in distribution in the C space, the space of continuous functions, to zero for $z\in\mathcal{C}$. This implies that

$$\sup_{z\in\mathcal{C}}\left|\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-2}(z)}\bar{\mathbf{s}}-\frac{c\underline{m}^2(z)\int\frac{t\,\mathrm{d}H(t)}{(1+t\underline{m}(z))^2}}{1-c\int\frac{(\underline{m}(z))^2t^2\,\mathrm{d}H(t)}{(1+t\underline{m}(z))^2}}\right|\xrightarrow{\text{i.p.}}0$$

and that

$$\sup_{z\in\mathcal{C}}\left|\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-1}(z)}\bar{\mathbf{s}}-1-z\underline{m}(z)\right|\xrightarrow{\text{i.p.}}0.\tag{2.53}$$

Thus we conclude from (2.25), (2.28) and (2.59) below that

$$\sup_{z\in\mathcal{C}}\left|\frac{\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-2}(z)}\bar{\mathbf{s}}}{1-\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-1}(z)}\bar{\mathbf{s}}}+\frac{c\underline{m}(z)\int\frac{t\,\mathrm{d}H(t)}{(1+t\underline{m}(z))^2}}{z(1-c\int\frac{(\underline{m}(z))^2t^2\,\mathrm{d}H(t)}{(1+t\underline{m}(z))^2})}\right|\xrightarrow{\text{i.p.}}0.\tag{2.54}$$

Theorem 1 follows from Lemma 1.1 of [3], the argument of Theorem 1.4 of [16], Theorems 4.4 and 5.1 of [8], (2.22), (2.23), (2.29) and (2.54). Formulaes (1.8) and (1.9) follow from (1.18) and (1.19) of [16].

2.6. *Derivations of* (1.10) *and* (1.11)

We now verify (1.10) and (1.11). Consider (1.10) first. As in [3], we select the contour to be the rectangle with sides parallel to the axes. It intersects the real axis at $a_1 \neq 0$ and b_1 (the support of $\underline{F}_{c,H}$ is a subset of (a_1, b_1)), and the horizontal sides are a distance v away from the real axis. We let $v \to 0$. By (1.3) we obtain

$$\frac{\mathrm{d}\underline{m}(z)}{\mathrm{d}z} = \frac{\underline{m}^2(z)}{1 - c\int \frac{(\underline{m}(z))^2 t^2 \,\mathrm{d}H(t)}{(1+t\underline{m}(z))^2}}.$$

This, together with (1.3) and integration by parts, ensures that

$$\begin{aligned}
&\frac{1}{2\pi \mathrm{i}} \int g(z) \frac{c\underline{m}(z)\int \frac{t\,\mathrm{d}H(t)}{(1+t\underline{m}(z))^2}}{z(1 - c\int \frac{(\underline{m}(z))^2 t^2\,\mathrm{d}H(t)}{(1+t\underline{m}(z))^2})}\,\mathrm{d}z \\
&\quad = \frac{1}{2\pi\mathrm{i}} \int g(z) \frac{\mathrm{d}}{\mathrm{d}z}\bigl(\mathrm{Log}\bigl(z\underline{m}(z)\bigr)\bigr)\,\mathrm{d}z = -\frac{1}{2\pi\mathrm{i}}\int g'(z)\,\mathrm{Log}\bigl(z\underline{m}(z)\bigr)\,\mathrm{d}z,
\end{aligned} \tag{2.55}$$

where Log denotes any branch of the logarithm. By the fact that $|\underline{m}(z)| \leq 1/v$ and (5.1) in [3] the integrals on the two vertical sides in (2.55) are bounded in absolute value by $Mv\log v^{-1}$, which converges to zero. Therefore the integral in (2.55) equals

$$-\frac{1}{\pi}\int_{a_1}^{b_1} g_i'(x+\mathrm{i}v)\log\bigl|(x+\mathrm{i}v)\underline{m}(x+\mathrm{i}v)\bigr|\,\mathrm{d}x - \frac{1}{\pi}\int_{a_1}^{b_1} g_r'(x+\mathrm{i}v)\arg\bigl[-(x+\mathrm{i}v)\underline{m}(x+\mathrm{i}v)\bigr]\,\mathrm{d}x. \tag{2.56}$$

By the fact that $|\underline{m}(z)| \leq 1/v$, (5.6) and (5.1) in [3] the first term in (2.56) is bounded in absolute value by $Mv\log v^{-1}$, converging to zero. Therefore we conclude from the dominated convergence theorem that

$$-\frac{1}{2\pi\mathrm{i}}\int g'(z)\,\mathrm{Log}\bigl(z\underline{m}(z)\bigr)\,\mathrm{d}z \to -\frac{1}{\pi}\int g'(x)\arg\bigl[x\underline{m}(x)\bigr]\,\mathrm{d}x. \tag{2.57}$$

Consider (1.11) now. Keep in mind that $\mathbf{T}_n = \mathbf{I}$ in this case. We select the same contour for evaluating (1.11) as that for (1.10) but with a_1 and b_1 replaced by a and b (a and b are defined in the Introduction). The simplified formula for the first term on the right-hand of (1.8) is given in [3]. By remark 2 the second term on the right-hand of (1.8) then becomes

$$-\frac{c(EX_{11}^4 - 3)}{2\pi\mathrm{i}}\int \frac{g(z)\underline{m}^3(z)/(1+\underline{m}(z))^3}{1 - c\underline{m}^2(z)/(1+\underline{m}(z))^2}\,\mathrm{d}z. \tag{2.58}$$

To calculate (2.58), solving (1.3) gives

$$\underline{m}(z) = \frac{-(z+1-c)+\sqrt{(z-1-c)^2-4c}}{2z}.$$

It follows that $|\underline{m}(z)|$ and $|m(z)|$ are bounded on the contour $\mathcal{C}$. Hence by (1.3) and (1.5)

$$\left|\frac{1}{1+cm(z)}\right| = \bigl|1 - c - czm(z)\bigr| \leq M \tag{2.59}$$

and by (1.3)

$$\left|\frac{c\underline{m}(z)}{1+\underline{m}(z)}\right| = \bigl|1 + z\underline{m}(z)\bigr| \leq M.$$

One can verify that

$$1 - c\frac{\underline{m}^2(z)}{(1+\underline{m}(z))^2} = \frac{\sqrt{(z-1-c)^2-4c}}{2c}\frac{(1+cm(z))}{cm(z)},$$

where we use (1.5) and (1.3). Note that

$$\left|\sqrt{(z-1-c)^2-4c}\right| = \left|\sqrt{(z-a)(z-b)}\right|$$

and that

$$\int_a^b \frac{1}{\sqrt{(x-a)(b-x)}}\,\mathrm{d}x < \infty. \tag{2.60}$$

We then conclude from (2.59) and (2.60) that the integrals on the vertical lines in (2.58) are bounded in absolute value by

$$Mv\int_a^b \frac{1}{\sqrt{(x-a)(b-x)}}\,\mathrm{d}x,$$

converging to zero. The integral on the two horizontal lines equals

$$\begin{aligned}
&-\frac{c(EX_{11}^4-3)}{\pi}\int g_i(z)\,\mathrm{Re}\left[\frac{\underline{m}^3(z)/(1+\underline{m}(z))^3}{1-c\underline{m}^2(z)/(1+\underline{m}(z))^2}\right]\mathrm{d}z\\
&\quad-\frac{c(EX_{11}^4-3)}{\pi}\int g_r(z)\Im\left[\frac{\underline{m}^3(z)/(1+\underline{m}(z))^3}{1-c\underline{m}^2(z)/(1+\underline{m}(z))^2}\right]\mathrm{d}z,
\end{aligned}\tag{2.61}$$

where $g_i'(x+\mathrm{i}v)$ and $g_r'(x+\mathrm{i}v)$ denote, respectively, the imaginary part and real part of $g'(x+\mathrm{i}v)$ and $\mathrm{Re}(\cdot)$ and $\Im(\cdot)$ also denote, respectively, the imaginary part and real part of the corresponding function. It follows from (5.6) in [3] and (2.59) and (2.60) that the first term in (2.61) is bounded in absolute value by

$$Mv\int_a^b \frac{1}{\sqrt{(x-a)(b-x)}}\,\mathrm{d}x,$$

converging to zero. Applying the generalized dominated convergence theorem and (2.59) and (2.60) to the second term in (2.61) yields the third term in (1.11).

3. Proof of Theorem 2

Let $\|\cdot\|$ denote the spectral norm of matrices or the Euclidean norm of vectors. Let $z = u + \mathrm{i}v$, $v > 0$. For $K > 0$, let $\tilde{\mathbf{X}}_n = (\tilde{X}_{ij})$, $\bar{\tilde{\mathbf{s}}} = \frac{1}{n}\sum_{j=1}^n \tilde{\mathbf{s}}_j$ and $\tilde{\mathcal{S}} = \frac{1}{n}\tilde{\mathbf{X}}_n\tilde{\mathbf{X}}_n^T - \bar{\tilde{\mathbf{s}}}\bar{\tilde{\mathbf{s}}}^T$, where $\tilde{X}_{ij} = X_{ij}I(|X_{ij}| \le K) - EX_{ij}I(|X_{ij}| \le K)$ and $\tilde{\mathbf{s}}_j = (\tilde{X}_{1j}, \dots, \tilde{X}_{pj})^T$. Note that

$$\begin{aligned}
\left|\mathbf{x}_n^T(\mathcal{S}-zI)^{-1}\mathbf{x}_n - \mathbf{x}_n^T(\tilde{\mathcal{S}}-zI)^{-1}\mathbf{x}_n\right| &\le \|\mathbf{x}_n\|^2\left\|(\mathcal{S}-zI)^{-1} - (\tilde{\mathcal{S}}-zI)^{-1}\right\|\\
&\le \frac{1}{v^2}\left\|\frac{1}{n}\mathbf{X}_n\mathbf{X}_n^T - \frac{1}{n}\tilde{\mathbf{X}}_n\tilde{\mathbf{X}}_n^T\right\| + \frac{1}{v^2}\left\|\bar{\mathbf{s}}\bar{\mathbf{s}}^T - \bar{\tilde{\mathbf{s}}}\bar{\tilde{\mathbf{s}}}^T\right\|.
\end{aligned}$$

Moreover, it is proven in [4] that $\|\frac{1}{n}\mathbf{X}_n\mathbf{X}_n^T - \frac{1}{n}\tilde{\mathbf{X}}_n\tilde{\mathbf{X}}_n^T\|$ can be arbitrary small if K is sufficiently large. Thus, it suffices to investigate $\|(\bar{\mathbf{s}} - \bar{\tilde{\mathbf{s}}})(\bar{\mathbf{s}} - \bar{\tilde{\mathbf{s}}}^T)\|$ and $\|(\bar{\mathbf{s}} - \bar{\tilde{\mathbf{s}}})\bar{\tilde{\mathbf{s}}}^T\|$.

Define $\hat{\mathbf{s}}_j = \mathbf{T}_n^{1/2}(X_{1j} - \tilde{X}_{1j}, \ldots, X_{pj} - \tilde{X}_{pj})^T$ and then obtain

$$\|(\bar{\mathbf{s}} - \bar{\tilde{\mathbf{s}}})(\bar{\mathbf{s}} - \bar{\tilde{\mathbf{s}}})^T\| = \frac{1}{n^2}\sum_{j=1}^{n}\big(\hat{\mathbf{s}}_j^T\hat{\mathbf{s}}_j - E\big(\hat{\mathbf{s}}_j^T\hat{\mathbf{s}}_j\big)\big) + \frac{1}{n^2}\sum_{j_1\neq j_2}^{n}\hat{\mathbf{s}}_{j_1}^T\hat{\mathbf{s}}_{j_2} + \frac{1}{n^2}\sum_{j=1}^{n}E\big(\hat{\mathbf{s}}_j^T\hat{\mathbf{s}}_j\big). \tag{3.1}$$

Since

$$E\left|\frac{1}{n^2}\sum_{j=1}^{n}\big(\hat{\mathbf{s}}_j^T\hat{\mathbf{s}}_j - E\big(\hat{\mathbf{s}}_j^T\hat{\mathbf{s}}_j\big)\big)\right|^2 = \frac{1}{n^4}\sum_{j=1}^{n}E\big|\big(\hat{\mathbf{s}}_j^T\hat{\mathbf{s}}_j - E\big(\hat{\mathbf{s}}_j^T\hat{\mathbf{s}}_j\big)\big)\big|^2 \le \frac{M}{n^2} \tag{3.2}$$

we have $\frac{1}{n^2}\sum_{j=1}^n(\hat{\mathbf{s}}_j^T\hat{\mathbf{s}}_j - E\hat{\mathbf{s}}_j^T\hat{\mathbf{s}}_j) \xrightarrow{\text{a.s.}} 0$ by the Borel–Cantelli lemma. Obviously $\frac{1}{n^2}\sum_{j=1}^n E(\hat{\mathbf{s}}_j^T\hat{\mathbf{s}}_j)$ can be arbitrarily small by choosing K large sufficiently. A direct calculation indicates that

$$\begin{aligned}
E\left|\frac{1}{n^2}\sum_{j_1\neq j_2}^{n}\hat{\mathbf{s}}_{j_1}^T\hat{\mathbf{s}}_{j_2}\right|^4 &= \frac{1}{n^8}\sum_{\substack{j_1\neq j_2, j_3\neq j_4,\\ j_5\neq j_6, j_7\neq j_8}} E\big[\hat{\mathbf{s}}_{j_1}^T\hat{\mathbf{s}}_{j_2}\hat{\mathbf{s}}_{j_3}^T\hat{\mathbf{s}}_{j_4}\hat{\mathbf{s}}_{j_5}^T\hat{\mathbf{s}}_{j_6}\hat{\mathbf{s}}_{j_7}^T\hat{\mathbf{s}}_{j_8}\big] \\
&\le \frac{M}{n^8}\sum_{j_1\neq j_2}E\big|\hat{\mathbf{s}}_{j_1}^T\hat{\mathbf{s}}_{j_2}\big|^4 + \frac{M}{n^8}\sum_{j_1\neq j_2, j_2\neq j_3, j_3\neq j_4, j_4\neq j_1}E\big|\hat{\mathbf{s}}_{j_1}^T\hat{\mathbf{s}}_{j_2}\hat{\mathbf{s}}_{j_2}^T\hat{\mathbf{s}}_{j_3}\hat{\mathbf{s}}_{j_3}^T\hat{\mathbf{s}}_{j_4}\hat{\mathbf{s}}_{j_4}^T\hat{\mathbf{s}}_{j_1}\big| \\
&\quad + \frac{M}{n^8}\sum_{j_1\neq j_2, j_2\neq j_3, j_3\neq j_1}E\big|\big(\hat{\mathbf{s}}_{j_1}^T\hat{\mathbf{s}}_{j_2}\big)^2\hat{\mathbf{s}}_{j_2}^T\hat{\mathbf{s}}_{j_3}\hat{\mathbf{s}}_{j_3}^T\hat{\mathbf{s}}_{j_1}\big| + \frac{M}{n^8}\sum_{j_1\neq j_2, j_3\neq j_2}E\big|\big(\hat{\mathbf{s}}_{j_1}^T\hat{\mathbf{s}}_{j_2}\big)^2\big(\hat{\mathbf{s}}_{j_2}^T\hat{\mathbf{s}}_{j_3}\big)^2\big| \\
&\quad + \frac{M}{n^8}\sum_{j_1\neq j_2, j_3\neq j_4}E\big|\hat{\mathbf{s}}_{j_1}^T\hat{\mathbf{s}}_{j_2}\big|^2E\big|\hat{\mathbf{s}}_{j_3}^T\hat{\mathbf{s}}_{j_4}\big|^2 = \mathrm{O}\left(\frac{1}{n^2}\right).
\end{aligned} \tag{3.3}$$

Therefore $\|(\bar{\mathbf{s}} - \bar{\tilde{\mathbf{s}}})(\bar{\mathbf{s}} - \bar{\tilde{\mathbf{s}}}^T)\|$ can be arbitrary small with probability one by choosing K large sufficiently. Likewise, one can also verify that $\|(\bar{\mathbf{s}} - \bar{\tilde{\mathbf{s}}})\bar{\tilde{\mathbf{s}}}^T\|$ can be arbitrary small by choosing K large sufficiently. The re-scaling of $\tilde{X}_{ij}$ can be treated similarly, because $\lim_{n\to\infty}E|\bar{X}_{11}|^2 = 1$. Hence, in what follows, we may assume $|X_{ij}| \le K$, $EX_{11} = 0$ and $E|X_{11}|^2 = 1$ (for simplicity, suppressing all super- or sub-scripts on the variables X_{ij}).

Recalling $\mathbf{A}^{-1}(z) = (\mathbf{S} - zI)^{-1}$, it is observed that

$$\mathbf{x}_n^T(\mathcal{S} - zI)^{-1}\mathbf{x}_n = \mathbf{x}_n^T\mathbf{A}^{-1}(z)\mathbf{x}_n + \frac{\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\mathbf{x}_n}{1 - \bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}}. \tag{3.4}$$

To prove Theorem 2, according to the argument of Theorem 1 in [4] it is sufficient to show that

$$\frac{\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}\bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\mathbf{x}_n}{1 - \bar{\mathbf{s}}^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}} \xrightarrow{\text{a.s.}} 0. \tag{3.5}$$

To this end, we first show that

$$\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}} - E\big(\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}\big) \xrightarrow{\text{a.s.}} 0. \tag{3.6}$$

We use the same notation as in the proof of Theorem 1. Write

$$\begin{aligned}
\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}} - E\big(\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}\big) &= \sum_{j=1}^{n}E_j\big(\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}\big) - E_{j-1}\big(\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}\big) \\
&= \sum_{j=1}^{n}(E_j - E_{j-1})\big(\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}} - \mathbf{x}_n^T\mathbf{A}_j^{-1}(z)\bar{\mathbf{s}}_j\big).
\end{aligned} \tag{3.7}$$

Furthermore, we have

$$\mathbf{x}_n^T \mathbf{A}^{-1}(z)\bar{\mathbf{s}} - \mathbf{x}_n^T \mathbf{A}_j^{-1}(z)\bar{\mathbf{s}}_j = c_{n1} + c_{n2} + c_{n3}, \tag{3.8}$$

where, via (2.33),

$$c_{n1} = \mathbf{x}_n^T \big(\mathbf{A}^{-1}(z) - \mathbf{A}_j^{-1}(z)\big)(\bar{\mathbf{s}} - \bar{\mathbf{s}}_j) = -\frac{1}{n^2}\mathbf{x}_n^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j\mathbf{s}_j^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j\beta_j(z),$$

$$c_{n2} = \mathbf{x}_n^T \big(\mathbf{A}^{-1}(z) - \mathbf{A}_j^{-1}(z)\big)\bar{\mathbf{s}}_j = -\frac{1}{n}\mathbf{x}_n^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j\mathbf{s}_j^T \mathbf{A}_j^{-1}(z)\bar{\mathbf{s}}_j\beta_j(z)$$

and

$$c_{n3} = \mathbf{x}_n^T \mathbf{A}_j^{-1}(z)(\bar{\mathbf{s}} - \bar{\mathbf{s}}_j) = \frac{1}{n}\mathbf{x}_n^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j.$$

Observe that $|\beta_j(z)| \le |z|/v$ and $|b_j(z)| \le |z|/v$ (see (3.4) in [2]). Using an argument similar to (3.19) of [15] one can prove that

$$E\big|\mathbf{x}_n^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j\big|^4 = \mathrm{O}(1). \tag{3.9}$$

By the Burkholder inequality and the fact that $|\mathbf{s}_j^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j/n| \le M/v$ we have

$$\begin{aligned} E\left|\sum_{j=1}^n (E_j - E_{j-1})c_{n1}\right|^4 &\le \frac{M}{n^4}E\left[\sum_{j=1}^n \left|\frac{1}{n}\mathbf{x}_n^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j\mathbf{s}_j^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j\right|^2\right]^2 \\ &\le \frac{M}{n^3}\sum_{j=1}^n E\big|\mathbf{x}_n^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j\big|^4 = \mathrm{O}\left(\frac{1}{n^2}\right), \end{aligned} \tag{3.10}$$

which implies that $\sum_{j=1}^n (E_j - E_{j-1})c_{n1} \xrightarrow{\text{a.s.}} 0$. The above argument also ensures that $\sum_{j=1}^n (E_j - E_{j-1})c_{n3} \xrightarrow{\text{a.s.}} 0$. Similarly, by the Burkholder inequality, Lemma 2.7 in [2] and (2.38) we have

$$\begin{aligned} E\left|\sum_{j=1}^n (E_j - E_{j-1})c_{n2}\right|^4 &\le \frac{M}{n^3}\sum_{j=1}^n E\big|\mathbf{s}_j^T \mathbf{A}_j^{-1}(z)\bar{\mathbf{s}}_j\mathbf{x}_n^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j\big|^4 \\ &\le \frac{M}{n^3}\sum_{j=1}^n \big[E\big|\mathbf{s}_j^T \mathbf{A}_j^{-1}(z)\bar{\mathbf{s}}_j\mathbf{x}_n^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j - \mathbf{x}_n^T \mathbf{A}_j^{-2}(z)\bar{\mathbf{s}}_j\big|^4 + E\big|\mathbf{x}_n^T \mathbf{A}_j^{-2}(z)\bar{\mathbf{s}}_j\big|^4\big] \\ &\le \frac{M}{n^3}\sum_{j=1}^n E\big\|\bar{\mathbf{s}}_j^T \bar{\mathbf{s}}_j\big\|^2 = \mathrm{O}\left(\frac{1}{n^2}\right). \end{aligned} \tag{3.11}$$

Thus we prove (3.6), as expected.

Finally applying $\bar{\mathbf{s}} = \sum_j s_j/n$ gives

$$\begin{aligned} E\big(\mathbf{x}_n^T \mathbf{A}^{-1}(z)\bar{\mathbf{s}}\big) &= \frac{1}{n}\sum_{j=1} E\big(\mathbf{x}_n^T \mathbf{A}^{-1}(z)\mathbf{s}_j\big) = \frac{1}{n}\sum_{j=1} E\big(\mathbf{x}_n^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j\beta_j(z)\big) \\ &= -\frac{1}{n}\sum_{j=1} E\big(\mathbf{x}_n^T \mathbf{A}_j^{-1}(z)\mathbf{s}_j b_1(z)\beta_j(z)\gamma_j(z)\big) = \mathrm{O}\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \tag{3.12}$$

where the step before to the last one uses (2.41) and the last step uses Holder's inequality, (2.42) and (3.9). Therefore (3.5) follows from (3.6), (3.12), (2.28), and the fact that $\bar{\mathbf{s}}^T\bar{\mathbf{s}} \le M\bar{\mathbf{x}}^T\bar{\mathbf{x}} \xrightarrow{\text{a.s.}} Mc$, which can be verified by an argument similar to (3.1)–(3.3).

4. Proof of Theorem 3

The proof of Theorem 3 is similar to that of Theorem 1. The only difference is that the extra random variable here converges in probability in the C space to zero.

Similar to (2.19) we can also truncate the underlying random variables at $\varepsilon_n\sqrt{n}$ with ε_n defined as before. Thus, in view of the structure of $\mathcal{S}$ we may assume that the following additional conditions hold:

$$|X_{ij}| \le \sqrt{n}\varepsilon_n, \quad EX_{ij} = 0, \qquad EX_{11}^2 = 1 + \mathrm{o}\bigl(p^{-1}\bigr) \tag{4.1}$$

and under assumption (b) of Theorem 3

$$E(X_{11} - \mu)^4 = 3 + \mathrm{o}(1). \tag{4.2}$$

It is proved in section 10.7 of [5] that Lemma 2 in [4] holds under conditions (4.1) and (4.2). Also, we see that Theorem 1.3 of [16] is true under conditions (4.1) and (4.2) by carefully checking on the argument of Theorem 1.3 of [16]. Thus, to prove Theorem 3, by Theorems 4.4 and 5.1 of [8], (2.53) and (3.4) it is sufficient to prove that on the contour $\mathcal{C}$ ($\mathcal{C}$ is defined in Theorem 1)

$$n^{1/4}\mathbf{x}_n^T\widehat{\mathbf{A}^{-1}(z)}\bar{\mathbf{s}} \xrightarrow{\text{i.p.}} 0, \tag{4.3}$$

where the truncated process $\mathbf{x}_n^T\widehat{\mathbf{A}^{-1}(z)}\bar{\mathbf{s}}$ is obtained from $\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}$ like the truncated process $\bar{\mathbf{s}}^T\widehat{\mathbf{A}^{-2}(z)}\bar{\mathbf{s}}$ is obtained from $\bar{\mathbf{s}}^T\mathbf{A}^{-2}(z)\bar{\mathbf{s}}$ in Theorem 1.

To prove (4.3), as in Theorem 1, it is sufficient to prove that

$$n^{1/4}\mathbf{x}_n^T\widehat{\mathbf{A}^{-1}(z)}\bar{\mathbf{s}} \xrightarrow{d} 0. \tag{4.4}$$

For each $z \in \mathbb{C}_u$ from the definition of c_{n1} in (3.8) we have

$$E\left|n^{1/4}\sum_{j=1}^n (E_j - E_{j-1})c_{n1}\right|^2 \le M\sqrt{n}\sum_{j=1}^n E|c_{n1}|^2 \le Mn^{-1/2},$$

because by Holder's inequality, (2.42), (2.20) and (3.9)

$$E|c_{n1}|^2 \le M\left(E\left|\frac{1}{n}\mathbf{x}_n^T\mathbf{A}_j^{-1}(z)\mathbf{s}_j\right|^4 E\left|\frac{1}{n}\mathbf{s}_j^T\mathbf{A}_j^{-1}(z)\mathbf{s}_j\right|^4\right)^{1/2} = \mathrm{O}\bigl(n^{-2}\bigr).$$

Appealing to (3.9) yields

$$E\left|n^{1/4}\sum_{j=1}^n (E_j - E_{j-1})c_{n3}\right|^2 \le Mn^{-1/2}.$$

By (2.31), (2.20) and (3.9) we obtain

$$E\left|n^{1/4}\sum_{j=1}^n (E_j - E_{j-1})c_{n2}\right|^2 \le Mn^{-1/2}.$$

With $M_n(z) = n^{1/4}(\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}} - E\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}})$ we thus have

$$E\bigl|M_n(z)\bigr|^2 = \mathrm{O}\bigl(n^{-1/2}\bigr). \tag{4.5}$$

By (2.37) and (2.40), (3.9) holds for $z \in \mathbb{C}_n^+$. For $z \in \mathbb{C}_n^+$, it follows from (3.12), (3.9), (2.20), (2.36), (2.39) and (2.42) that

$$\begin{aligned}\left|E\left(n^{1/4}\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}\right)\right| &= \left|\frac{1}{n^{3/4}}\sum_{j=1} E\left(\mathbf{x}_n^T\mathbf{A}_j^{-1}(z)\mathbf{s}_j b_1(z)\beta_j(z)\gamma_j(z)\right)\right| \\ &\le Mn^{1/4}\left(E\left|\mathbf{x}_n^T\mathbf{A}_1^{-1}(z)\mathbf{s}_1\right|^2 E\left|\gamma_1(z)\right|^2\right)^{1/2} + Mv^{-4}n^3 P\left(\lambda_{\max}(\mathbf{S}) > \eta_r \text{ or } \lambda_{\min}(\mathbf{S}_1) \le \eta_l\right) \\ &\le M/n^{1/4}. \end{aligned} \tag{4.6}$$

From (3.12) one can verify that for $\Im(z) = v_0$ (v_0 is defined in the proof of Theorem 1)

$$E\left|n^{1/4}\mathbf{x}_n^T\mathbf{A}^{-1}(z)\bar{\mathbf{s}}\right|^2 \le M,$$

which implies condition (i) of Theorem 12.3 of [8]. Thus, in view of (4.5) and (4.6), to ensure that convergence in (4.4) is true on the contour $\mathcal{C}$, it suffices to show that

$$E\frac{|M_n(z_1) - M_n(z_2)|^2}{|z_1 - z_2|^2} \le M, \quad \text{if } z_1, z_2 \in \mathcal{C}_n^+ \cup \mathcal{C}_n^-, \tag{4.7}$$

where $\mathcal{C}_n^+$ and $\mathcal{C}_n^-$ are defined in Theorem 1. Write

$$\begin{aligned}&\frac{M_n(z_1) - M_n(z_2)}{z_1 - z_2} \\ &\quad = n^{1/4}\left(\mathbf{x}_n^T\mathbf{A}^{-1}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}} - E\left(\mathbf{x}_n^T\mathbf{A}^{-1}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}}\right)\right) \\ &\quad = n^{1/4}\sum_{j=1}^n (E_j - E_{j-1})\left(\mathbf{x}_n^T\mathbf{A}^{-1}(z_1)\mathbf{A}^{-1}(z_2)\bar{\mathbf{s}} - \mathbf{x}_n^T\mathbf{A}_j^{-1}(z_1)\mathbf{A}_j^{-1}(z_2)\bar{\mathbf{s}}_j\right) \\ &\quad = d_{n1} + d_{n2} + d_{n3} + d_{n4} + d_{n5} + d_{n6},\end{aligned}$$

where

$$d_{n1} = \frac{n^{1/4}}{n^3}\sum_{j=1}^n (E_j - E_{j-1})\beta_j(z_1)\beta_j(z_2)\mathbf{s}_j^T\mathbf{A}_j^{-1}(z_1)\mathbf{A}_j^{-1}(z_2)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-1}(z_2)\mathbf{s}_j\mathbf{x}_n^T\mathbf{A}_j^{-1}(z_1)\mathbf{s}_j,$$

$$d_{n2} = -\frac{n^{1/4}}{n^2}\sum_{j=1}^n (E_j - E_{j-1})\beta_j(z_1)\mathbf{s}_j^T\mathbf{A}_j^{-1}(z_1)\mathbf{A}_j^{-1}(z_2)\mathbf{s}_j\mathbf{x}_n^T\mathbf{A}_j^{-1}(z_1)\mathbf{s}_j,$$

$$d_{n3} = -\frac{n^{1/4}}{n^2}\sum_{j=1}^n (E_j - E_{j-1})\beta_j(z_2)\mathbf{x}_n^T\mathbf{A}_j^{-1}(z_1)\mathbf{A}_j^{-1}(z_2)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-1}(z_2)\mathbf{s}_j,$$

$$d_{n4} = \frac{n^{1/4}}{n^2}\sum_{j=1}^n (E_j - E_{j-1})\beta_j(z_1)\beta_j(z_2)\mathbf{s}_j^T\mathbf{A}_j^{-1}(z_1)\mathbf{A}_j^{-1}(z_2)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-1}(z_2)\bar{\mathbf{s}}_j\mathbf{x}_n^T\mathbf{A}_j^{-1}(z_1)\mathbf{s}_j,$$

$$d_{n5} = -\frac{n^{1/4}}{n}\sum_{j=1}^n (E_j - E_{j-1})\beta_j(z_1)\mathbf{s}_j^T\mathbf{A}_j^{-1}(z_1)\mathbf{A}_j^{-1}(z_2)\bar{\mathbf{s}}_j\mathbf{x}_n^T\mathbf{A}_j^{-1}(z_1)\mathbf{s}_j$$

and

$$d_{n6} = -\frac{n^{1/4}}{n}\sum_{j=1}^n (E_j - E_{j-1})\beta_j(z_2)\mathbf{x}_n^T\mathbf{A}_j^{-1}(z_1)\mathbf{A}_j^{-1}(z_2)\mathbf{s}_j\mathbf{s}_j^T\mathbf{A}_j^{-1}(z_2)\bar{\mathbf{s}}_j.$$

Here we also use the fact that $\bar{\mathbf{s}} = \bar{\mathbf{s}}_j + \mathbf{s}_j/n$ and the identity above (3.7) in [3]. By (2.36), (2.40), (2.39), (3.9) and (2.39) we obtain

$$E|d_{n1}|^2 \le \frac{M}{\sqrt{n}} E\left|\mathbf{x}_n^T \mathbf{A}_1^{-1}(z_1)\mathbf{s}_1\right|^2 + Mv^{-12}n^{11/2}P\left(\lambda_{\max}(\mathbf{S}) > \eta_r \text{ or } \lambda_{\min}(\mathbf{S}_1) \le \eta_l\right) \le \frac{M}{\sqrt{n}}.$$

This argument of course handles the terms d_{n2} and d_{n3}. Furthermore, we conclude from (2.36), (2.40), (2.39), (3.9), (2.39), (2.31) and Holder's inequality that

$$\begin{aligned} E|d_{n4}|^2 &\le \frac{M}{\sqrt{n}}\left(E\left|\mathbf{s}_1^T \mathbf{A}_1^{-1}(z_2)\bar{\mathbf{s}}_1\right|^2 E\left|\mathbf{x}_n^T \mathbf{A}_1^{-1}(z_1)\mathbf{s}_1\right|^2\right)^{1/2} \\ &\quad + Mv^{-12}n^{11/2}P\left(\lambda_{\max}(\mathbf{S}) > \eta_r \text{ or } \lambda_{\min}(\mathbf{S}_1) \le \eta_l\right) \le \frac{M}{\sqrt{n}}. \end{aligned}$$

Obviously, the argument for d_{n4} also applies to d_{n5} and d_{n6}. Thus, the proof of (4.3) is complete.

Acknowledgments

The author would like to thank the Editor and two referees for their constructive comments, which have helped to improve the paper a great deal. G. M. Pan is partially supported by the Ministry of Education, Singapore, under a Grant #ARC 14/11.

References

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*, 2nd edition. Wiley, New York, 1984. MR0771294
[2] Z. D. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large dimensional random matrices. *Ann. Probab.* **26** (1998) 316–345. MR1617051
[3] Z. D. Bai and J. W. Silverstein. CLT for linear spectral statistics of large dimensional sample covariance matrices. *Ann. Probab.* **32** (2004) 553–605. MR2040792
[4] Z. D. Bai, B. Q. Miao and G. M. Pan. On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab.* **35** (2007) 1532–1572. MR2330979
[5] Z. D. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*, 2nd edition. Springer, New York, 2010. MR2567175
[6] Z. D. Bai and G. M. Pan. Limiting behavior of eigenvectors of large Wigner matrices. *J. Stat. Phys.* **146** (2012) 519–549. MR2880031
[7] F. Benaych-Georges. Eigenvectors of Wigner matrices: Universality of global fluctuations. Available at http://arxiv.org/abs/1104.1219.
[8] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968. MR0233396
[9] C. Donati-Martin and A. Rouault. Truncations of Haar unitary matrices, traces and bivariate Brownian bridge. *RMTA* **1**(1) (2012) 1150007.
[10] I. Johnstone. On the distribution of the largest eigenvalue in principal component analysis. *Ann. Statist.* **29** (2001) 295–327. MR1863961
[11] T. F. Jiang. The limiting distriubtions of eigenvalues of sample correlation matrices. *Sankhya* **66** (2004) 35–48. MR2082906
[12] D. Jonsson. Some limit theorems of the eigenvalues of sample convariance matrix. *J. Multivariate Anal.* **12** (1982) 1–38. MR0650926
[13] O. Ledoit and S. Peche. Eigenvectors of some large sample covariance matrix ensembles. *Probab. Theory Related Fields* **151** (2011) 233–264. MR2834718
[14] V. A. Marčenko and L. A. Pastur. Distribution for some sets of random matrices. *Math. USSR-Sb.* **1** (1967) 457–483.
[15] G. M. Pan and W. Zhou. Central limit theorem for Hotelling's T^2 statistic under large dimension. *Ann. Appl. Probab.* **21** (2011) 1860–1910. MR2884053
[16] G. M. Pan and W. Zhou. Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab*. **18** (2008) 1232–1270. MR2418244
[17] G. M. Pan, Q. M. Shao and W. Zhou. Universality of sample covariance matrices: CLT of the smoothed empirical spectral distribution. Preprint.
[18] J. W. Silverstein. On the eigenvectors of large dimensional sample covariance matrices. *J. Multivariate Anal.* **30** (1989) 1–16. MR1003705
[19] J. W. Silverstein. Strong convergence of the limiting distribution of the eigenvalues of large dimensional random matrices. *J. Multivariate Anal.* **55** (1995) 331–339. MR1370408
[20] J. W. Silverstein. Weak convergence of random functions defined by the eigenvectors of sample covariance matrices. *Ann. Probab.* **18** (1990) 1174–1194. MR1062064
[21] T. Tao and V. Vu. Random matrices: Universal properties of eigenvectors. Available at arXiv:1103.2801v2.
[22] H. Xiao and W. Zhou. On the limit of the smallest eigenvalue of some sample covariance matrix. *J. Theoret. Probab*. **23** (2010) 1–20. MR2591901
[23] Y. Q. Yin, Z. D. Bai and P. R. Krishanaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** (1988) 509–521. MR0950344