Asymptotic Theory for Sample Covariance Matrix under Cross–Sectional Dependence [1]

By Guangming Pan[12] and Jiti Gao[2]

[1] *Nanyang Technological University and* [2] *The University of Adelaide*

Consider a sample covariance matrix of the form

$$\sqrt{\frac{n}{p}}\left(\frac{1}{n}\mathbf{T}^{1/2}\mathbf{X}\mathbf{X}^T\mathbf{T}^{1/2}-\mathbf{T}\right),$$

where $\mathbf{X}=(X_{ij})_{p\times n}$ consists of independent and identically distributed real random variables and $\mathbf{T}$ is a symmetric nonnegative definite nonrandom matrix. When $p\to\infty$ and $n\to\infty$ with $\frac{p}{n}\to 0$, it is shown that its empirical spectral distribution converges to a fixed distribution under the fourth moment condition. It is then discussed that such an asymptotic distribution can be used to derive an asymptotically consistent test for the hypothesis testing of cross–sectional independence.

AMS 2000 subject classifications: 15A52, 60F15, 62E20; 60F17.

Key words: Eigenvalues, Sample covariance matrix, Stieltjes transform, Strong convergence, Semicircle law.

[1] G. M. Pan's research was supported by a grant M58110052 at the Nanyang Technological University, Singapore; The second author would like to acknowledge the financial support from the Australian Research Council Discovery Grants Program under Grant Number: DP0879088.

[2] *Corresponding Author*: Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. Email: gmpan@ntu.edu.sg

1. Introduction

Recently, there has been a great interest in investigating the empirical spectral distribution function of the eigenvalues of large dimensional random matrices. The majority of the literature focuses on the asymptotic theory (see, for example, [3]–[7], [11]–[15] and [18]–[19]). In the paper by [12], the author provides an extensive discussion about the need to study sample covariance matrices and their large sample theory.

This paper motives such a discussion from a different aspect. Suppose that Z_{ij} are real–valued random variables. For $1 \leq j \leq p$, let $\mathbf{Z}_j = (Z_{j1}, \cdots, Z_{jn})^T$ denote the j–th time series and $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_p)^T$ be a panel of p time series, where n usually denotes the sample size in each of the time series data. In both theory and practice, it is not uncommon to assume that each of the time series $(Z_{j1}, Z_{j2}, \cdots, Z_{jn})$ is statistically dependent, but it may be unrealistic to assume that $\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_p$ are independent and even uncorrelated. This is because there is no natural ordering for cross–sectional indices. There are such cases in various disciplines. In economics and finance, for example, it is not unreasonable to expect that there is significant evidence of cross–sectional dependence in output innovations across p countries and regions in the World. In the field of climatology, there is also some evidence to show that climatic variables in different stations may be cross–sectionally dependent and the level of cross–sectional dependence may be determined by some kind of physical distance. Moreover, one would expect that climatic variables, such as temperature and rainfall variables, in a station in Australia have higher–level dependence with the same type of climatic variables in a station in New Zealand than those in the United States.

In such situations, it may be necessary to test whether $\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_p$ are uncorrelated before a statistical model is used to model such data. In the econometrics and statistics literature, several papers have considered testing for cross–sectional independence for the residuals involved in some specific regression models. Such studies include [16] for the parametric linear model case, [9] for the parametric nonlinear case, and [8] for the nonparametric nonlinear case. As the main motivation of this paper, we will propose using an empirical spectral distribution function based test statistic for cross–sectional uncorrelatedness of $\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_p$.

In the discussion of different types of hypothesis testing problems, existing studies include [10], [14] and [20]. Their common feature is to assume that the components of $\mathbf{Z}$ are all independent random variables.

The main contribution of this paper is summarized as follows:

- This paper establishes an asymptotic theory for the empirical spectral distribution function of the eigenvalues of a large dimensional random matrix A under a general dependent structure for the case of $\frac{p}{n} \to 0$. Such an asymptotic theory complements the main theory by [17] and [4] for the case where there is some dependence structure in the columns of a matrix and $\frac{p}{n} \to c \in (0, \infty)$.
- Because of the involvement of a symmetric deterministic matrix, the main structure of this paper covers some special but important cases. As a consequence, some existing results in the field become corollaries of the main theorem of this paper.
- In addition to the contribution to the theoretical development, we discuss the applicability of the empirical spectral distribution function in the construction of a general test for cross–sectional uncorrelatedness for a panel of time series.

The organization of this paper is as follows. Section 2 establishes the almost sure convergence of the empirical spectral distribution function to a given distribution function. Section 3 discusses how such an asymptotic convergence may be used to establish an asymptotically consistent test for cross–sectional uncorrelatedness. Conclusions and discussion are given in Section 4. The mathematical proof is given in Section 5.

2. Large sample theory

Suppose that X_{ij} are independent and identically distributed (i.i.d.) real–valued random variables. Let $\mathbf{s}_j = (X_{1j}, \cdots, X_{nj})^T$ denote the j–th column vector of random variables and $\mathbf{X} = (\mathbf{s}_1, \cdots, \mathbf{s}_n)$, where n usually denotes the sample size.

For any $p \times p$ matrix $\mathbf{A}$ with real eigenvalues, define its empirical spectral distribution function by

$$F^{\mathbf{A}}(x) = \frac{1}{n} \sum_{k=1}^{n} I(\lambda_k \leq x), \tag{2.1}$$

where $\lambda_k,\ k = 1, \cdots, p$ denote the eigenvalues of $\mathbf{A}$.

When $p \to \infty$ and $n \to \infty$ with $\frac{p}{n} \to c > 0$, matrices of the form $\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$ have been investigated in [15] and [11] and it has been shown that $F^{\mathbf{S}_n}$ converges to Marcenko and Pastur law's with probability one or in probability. For more detailed reading of the recent literature up to the year of 2005, see the monograph by [3].

Surprisingly, in the setting of $p \to \infty$ and $n \to \infty$ with $\frac{p}{n} \to 0$, Bai and Yin [2] prove that for the matrix $\frac{1}{2\sqrt{np}}(\mathbf{X}\mathbf{X}' - n\mathbf{I})$, its empirical spectral distribution converges, with probability one, to the semicircle law with density

$$f(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, & |x| < 1 \\ 0, , & |x| > 1 \end{cases}. \tag{2.2}$$

This density is also the limit of the empirical spectral distribution of a symmetric random matrix whose diagonal are i.i.d. random variables and above diagonal elements are also i.i.d. (see [19]).

In this paper under the setting of $p \to \infty$ and $n \to \infty$ with $\frac{p}{n} \to 0$, we consider the following matrix

$$\mathbf{S}_n = \frac{1}{n}\mathbf{T}^{1/2}\mathbf{X}\mathbf{X}^T\mathbf{T}^{1/2}, \tag{2.3}$$

where $\mathbf{T}$ is a $p \times p$ symmetric nonnegative definite matrix and $(\mathbf{T}^{1/2})^2 = \mathbf{T}$. To develop the limiting spectral distribution for $\mathbf{S}_n$ we then re-normalize it as follows:

$$\mathbf{A} = \sqrt{\frac{n}{p}}\left(\mathbf{S}_n - \mathbf{T}\right). \tag{2.4}$$

The moment method, in conjunction with sophisicated graph theory and combinatorial argument, was used in [2] to establish the semi-circle law. Instead, we use another popular tool in random matrix theory, Stieltjes transform, in this paper. The Stieltjes transform for any function $G(x)$ is given by

$$m_G(z) = \int \frac{1}{\lambda - z} dG(\lambda), \quad z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C},\ v = Imz > 0\}, \tag{2.5}$$

where $Im(\cdot)$ stands for the imaginary part of a complex number. The main result is listed as below.

Theorem 1. *Suppose that*

1) $\{X_{ij}\}$ *are i.i.d. real random variables with* $E[X_{11}] = 0$, $E[X_{11}^2] = 1$ *and* $E[X_{11}^4] < \infty$.

2) $\frac{p}{n} \to 0$ *with* $p \to \infty$ *and* $n \to \infty$.

3) $\mathbf{T}$ *is a symmetric nonnegative definite matrix with* $F^{\mathbf{T}}(x) \xrightarrow{D} H(x)$, *a probability distribution function as* $p \to \infty$.

Then $F^{\mathbf{A}}(\cdot)$ *converges, with probability one, to a fixed distribution function,* $F(\cdot)$, *whose Stieltjes transform satisfies*

$$s_1(z) = -\int \frac{dH(t)}{z + ts_2(z)}, \tag{2.6}$$

where $s_2(z)$ *is the unique solution in* $\mathbb{C}^+$ *to*

$$s_2(z) = -\int \frac{tdH(t)}{z + ts_2(z)}. \tag{2.7}$$

The proof of the theorem is given in Section 5 below.

Remark 1. *Apparently, this result recovers Theorem in* [5] *when* $\mathbf{T} = \mathbf{I}$.

3. Hypothesis testing

Let Z_{ij} be real–valued random variables, $\mathbf{Z}_j = (Z_{j1}, \cdots, Z_{jn})^T$ denote the j–th column vector for $1 \le j \le p$ and $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_p)^T$ be a panel of p vectors. Consider testing the null hypothesis H_0 versus an alternative H_1 of the form:

$$H_0: \ E[Z_{i1}Z_{j1}] = 0 \quad \text{for all } 1 \le i \ne j \le p \text{ versus} \tag{3.1}$$

$$H_1: \ E[Z_{i1}Z_{j1}] = \rho_{ij} \ne 0 \quad \text{for at least one pair } (i,j)\text{: } 1 \le i \ne j \le p, \tag{3.2}$$

where $\{\rho_{ij}\}$ is a set of real numbers.

Let $\mathbf{X}$ and $\mathbf{T}$ be as defined in Section 2 above. Let $\mathbf{Z} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$. Then we have

$$E[Z_{i1}Z_{j1}] = t_{ij}, \tag{3.3}$$

where $\{t_{ij}\}$ is the (i,j)–th element of matrix $\mathbf{T}$. In this case, equations (3.1) and (3.2) correspond to

$$H_0: \ \mathbf{T} = \mathbf{I} \quad \text{versus} \quad H_1: \ \mathbf{T} \ne \mathbf{I}. \tag{3.4}$$

Let $F_i^{\mathbf{A}}(\cdot)$ and $F_i(\cdot)$ correspond to $F^{\mathbf{A}}(\cdot)$ and $F(\cdot)$, respectively, under H_i for $i = 0, 1$.

Consider a Cramér-von Mises type of test statistic of the form

$$L_n = \int \left(F_1^{\mathbf{A}}(x) - F_0^{\mathbf{A}}(x)\right)^2 dF_0^{\mathbf{A}}(x). \tag{3.5}$$

Theorem 1 then implies the following proposition.

Proposition 3.1. *Under the conditions of Theorem 1, we have with probability one*

$$L_n \to \begin{cases} \int \left(F_0(x) - F_0(x)\right)^2 dF_0(x)dx = 0 & \textit{under } H_0 \\ \int \left(F_1(x) - F_0(x)\right)^2 dF_0(x)dx > 0 & \textit{under } H_1. \end{cases} \tag{3.6}$$

where $F_i(\cdot)$ corresponds to the limit of $F^{\mathbf{A}}(\cdot)$ with $\mathbf{T} = \mathbf{I}$ under H_0 and $\mathbf{T} \ne \mathbf{I}$ under H_1, respectively.

Equation (3.6) may suggest that there is some $C_n \to \infty$ such that

$$M_n \equiv C_n L_n \to_D \begin{cases} Z & \text{under } H_0 \\ \infty & \text{under } H_1, \end{cases} \tag{3.7}$$

where Z is a random variable.

Since the proof of (3.7) is quite challenging, we have not been able to include a rigorous proof in this paper. Hopefully, it may be given in a future paper.

4. Conclusions and discussion

This paper has considered establishing the empirical spectral distribution of a sample covariance matrix of the form $\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{T}^{1/2}\mathbf{X}\mathbf{X}^T\mathbf{T}^{1/2} - \mathbf{T})$, where $\mathbf{X} = (X_{ij})_{p\times n}$ consists of independent and identically distributed real random variables and $\mathbf{T}$ is a symmetric nonnegative definite nonrandom matrix. Theorem 1 has established the almost sure convergence of the empirical spectral distribution function to a fixed distribution function for the case where $p \to \infty$, $n \to \infty$ and $\frac{p}{n} \to 0$.

It has been discussed that such an asymptotic convergence may be used to derive the asymptotic consistency of a test statistic for cross–sectional uncorrelatedness. Future topics include a rigorous proof of equation (3.7) and the discussion of the size and power properties of the resulting test.

5. Proof of Theorem 1

The whole argument consists of four steps. The first step deals with the tightness of $F^{\mathbf{A}}$ and almost sure convergence of the random part of the Stieltjes transform of $F^{\mathbf{A}}$. The main difficulty is to prove that the limit of the Stieltjes transform of $EF^{\mathbf{A}}$ satisfies equations (2.6) and (2.7). To do that, we first investigate the corresponding matrix with Gaussian elements and then finish the proof by Lindeberg's method, along with the proof of uniqueness. These are accomplished, respectively, in steps 2–4.

Throughout the paper, M denotes a constant which may stand for different values at different places and the limits are taken as p goes to infinity (n may be viewed as $n(p)$, the function of p). Additionally, let $\|\cdot\|$ denote the Euclidean norm of a vector or the spectral norm of a matrix.

5.1. **Step 1: Almost sure convergence of the random part.** In this subsection, we prove that $F^{\mathbf{A}}$ is tight with probability one. In addition, we establish a general and useful result of the form

$$E\left|\frac{1}{p}tr\left(\mathbf{A}^{-1}(z)\mathbf{D}\right)-E\left[\frac{1}{p}tr\left(\mathbf{A}^{-1}(z)\mathbf{D}\right)\right]\right|^2\leq\frac{M}{p^2}\tag{5.1}$$

for an application at a late stage, where $(\mathbf{A}-z\mathbf{I})^{-1}$ is denoted by $\mathbf{A}^{-1}(z)$ and $\mathbf{D}$ is some non-random matrix with the spectral norm $\|\mathbf{D}\|\leq M$. Here $z=u+iv$ with $v>0$.

We start with the truncation of the spectral of the matrix $\mathbf{T}$ and of the elements of $\mathbf{X}$. Denote the spectral decomposition of $\mathbf{T}$ by $\mathbf{U}^T\Lambda\mathbf{U}$ where $\Lambda=diag(\lambda_1,\cdots,\lambda_p)$ is a diagonal matrix, $\lambda_1,\cdots,\lambda_n$ are eigenvalues of $\mathbf{T}$ and $\mathbf{U}$ is the corresponding eigenvector matrix. Then

$$\mathbf{T}^{1/2}=\mathbf{U}^T\Lambda^{1/2}\mathbf{U},\tag{5.2}$$

where $\Lambda^{1/2}=diag(\sqrt{\lambda_1},\cdots,\sqrt{\lambda_p})$.

Moreover, with τ being a pre-chosen positive constant such that τ is a continuity point of $F^{\mathbf{T}}(t)$, define $\Lambda_\tau=diag(\lambda_1 I(\lambda_1\leq\tau),\cdots,\lambda_p I(\lambda_p\leq\tau))$ and $\sqrt{\Lambda}_\tau=diag(\sqrt{\lambda_1}I(\lambda_1\leq\tau),\cdots,\sqrt{\lambda}_\tau I(\lambda_p\leq\tau))$. Set $\mathbf{T}_\tau=\mathbf{U}^T\Lambda_\tau\mathbf{U}$ and $\mathbf{T}_\tau^{1/2}=\mathbf{U}^T\sqrt{\Lambda}_\tau\mathbf{U}$. Then, by Lemmas 2.4 and 2.5 in [18]

$$\begin{aligned}
&\left\|F^{\mathbf{A}}-F^{\sqrt{\frac{n}{p}}(\mathbf{S}_n-\mathbf{T}_\tau)}\right\|\leq\frac{1}{p}rank(\mathbf{T}-\mathbf{T}_\tau)\to 1-F^{\mathbf{T}}(\tau),\\
&\left\|F^{\sqrt{\frac{n}{p}}(\mathbf{S}_n-\mathbf{T}_\tau)}-F^{\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{T}_\tau^{1/2}\mathbf{X}\mathbf{X}^T\mathbf{T}_\tau^{1/2}-\mathbf{T}_\tau)}\right\|\leq\frac{2}{p}rank(\mathbf{T}^{1/2}-\mathbf{T}_\tau^{1/2})\\
&\to\ 2(1-F^{\mathbf{T}}(\tau)).
\end{aligned}$$

In addition,

$$F^{\mathbf{T}_\tau}\to\int_0^x I(u\leq\tau)dH(u)+1-F^{\mathbf{T}}(\tau)\triangleq H_\tau(x).\tag{5.3}$$

The value of $1-F^{\mathbf{T}}(\tau)$ can be arbitrary small if τ is sufficiently large. Therefore by Propositions 3.1 and 3.2 in [1] in order to finish Theorem 1 it suffices to prove that $F^{\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{T}_\tau^{1/2}\mathbf{X}\mathbf{X}^T\mathbf{T}_\tau^{1/2}-\mathbf{T}_\tau)}$ converges with probability one to a nonrandom distribution function $F^{\tau}(x)$ whose Stieljtes transform satisfies (2.6) and (2.7) with $H(x)$ being replaced by $H_\tau(x)$. Consequently, we may assume that the spectral $\lambda_1, \cdots, \lambda_n$ are bounded, say by τ. To simplify the notation we still use $\mathbf{T}$ instead of using $\mathbf{T}_\tau$.

Additionally, let $\check{X}_{ij} = X_{ij} I(|X_{ij}| \le n^{1/4}\varepsilon_p)$ and $\tilde{X}_{ij} = \check{X}_{ij} - E\left[\check{X}_{ij}\right]$, where ε_p is chosen such that $\varepsilon_p \to 0$, $\varepsilon_p p^{1/4} \to \infty$ and $P(|X_{11}| \ge \varepsilon_p p^{1/4}) \le \varepsilon_p/n$.

Set $\tilde{\mathbf{X}} = (\tilde{X}_{ij})$ and $\tilde{\mathbf{A}} = \sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{T}^{1/2}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T\mathbf{T}^{1/2} - \mathbf{T})$.

Then, as in [5], one may prove that

$$\left|\left|F^{\tilde{\mathbf{A}}} - F^{\mathbf{A}}\right|\right| \xrightarrow{a.s.} 0. \tag{5.4}$$

In addition, we may also show that re-normalization of $\tilde{X}_{ij}$ does not affect the limiting spectral distribution of $\tilde{\mathbf{A}}$ with probability one. In view of the truncation above we may assume that

$$\|\mathbf{T}\| \le \tau,\ |X_{ij}| \le n^{1/4}\varepsilon_p,\ E\left[X_{ij}\right] = 0,\ E\left[X_{ij}^2\right] = 1. \tag{5.5}$$

Also, we use X_{ij} for $\tilde{X}_{ij}$ to simplify the notation.

We now verify that $F^{\mathbf{A}}$ is tight with probability one. Note that

$$\frac{1}{p} tr\left(\mathbf{A}^2\right) \le \frac{\tau^2}{p} tr\left[\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{X}\mathbf{X}^T - \mathbf{I})\right]^2 = \frac{\tau^2}{np^2}\sum_{i\neq j}(\hat{\mathbf{s}}_i^T\hat{\mathbf{s}}_j)^2 + \frac{\tau^2}{np^2}\sum_{i=1}^{p}(\hat{\mathbf{s}}_i^T\hat{\mathbf{s}}_i - n)^2, \tag{5.6}$$

where $\tilde{\mathbf{s}}_j^T$ denotes the j–th row of $\mathbf{X}$. It is easy to verify that the expectation of the term on the right hand above converges to one. It follows from Burkholder's inequality that

$$\begin{aligned}
& E\left|\frac{1}{np^2}\sum_{i=1}^{p}(\hat{\mathbf{s}}_i^T\hat{\mathbf{s}}_i - n)^2 - E(\hat{\mathbf{s}}_i^T\hat{\mathbf{s}}_i - n)^2\right|^2 \\
= & \ \frac{1}{n^2p^4}\sum_{i=1}^{p} E\left|(\hat{\mathbf{s}}_i^T\hat{\mathbf{s}}_i - n)^2 - E(\hat{\mathbf{s}}_i^T\hat{\mathbf{s}}_i - n)^2\right|^2 \\
\le & \ \frac{M}{n^2p^4}\sum_{i=1}^{p} E\left(\hat{\mathbf{s}}_i^T\hat{\mathbf{s}}_i - n\right)^4 \le \frac{M}{n^2p^4}\sum_{i=1}^{p}\left(\sum_{j=1}^{n} E\left[X_{ij}^4\right]\right)^2 \\
+ & \ \frac{M}{n^2p^4}\sum_{i=1}^{p}\sum_{j=1}^{n} E\left[X_{ij}^8\right] \le \frac{M}{p^2}.
\end{aligned}$$

A direct calculation indicates that

$$
\begin{aligned}
&E\left(\frac{\tau^2}{np^2}\sum_{i\neq j}(\hat{\mathbf{s}}_i^T\hat{\mathbf{s}}_j)^2 - E(\hat{\mathbf{s}}_i^T\hat{\mathbf{s}}_j)^2\right)^2 \\
&= \frac{1}{n^2p^4}\sum_{i_1\neq j_1, i_2\neq j_2} E\Big(\left[(\hat{\mathbf{s}}_{i_1}^T\hat{\mathbf{s}}_{j_1})^2 - E(\hat{\mathbf{s}}_{i_1}^T\hat{\mathbf{s}}_{j_1})^2\right]\left[(\hat{\mathbf{s}}_{i_2}^T\hat{\mathbf{s}}_{j_2})^2 - E(\hat{\mathbf{s}}_{i_2}^T\hat{\mathbf{s}}_{j_2})^2\right]\Big) \leq \frac{M}{p^2},
\end{aligned}
$$

which may be obtained by distinguishing different cases for $i_1 \neq j_1, i_2 \neq j_2$. We then conclude that

$$
\frac{1}{p}tr\left[\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{X}\mathbf{X}^T - \mathbf{I})\right]^2 \xrightarrow{a.s.} 1, \tag{5.7}
$$

which ensures that $F^{\mathbf{A}}$ is tight with probability one.

We then turn to the proof of (5.1). To this end, let $\mathcal{F}_k$ denote the σ-field generated by $\mathbf{s}_1, \cdots, \mathbf{s}_k$, $E_k = E(\cdot|\mathcal{F}_k)$ denote conditional expectation and E_0 unconditional expectation. Denote by $\mathbf{X}_k$ the matrix obtained from $\mathbf{X}$ with the k–th column deleted.

Moreover, to simplify the notation, set $\mathbf{A}_k = \sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{T}^{1/2}\mathbf{X}_k\mathbf{X}_k^T\mathbf{T}^{1/2} - \mathbf{T})$, $(\mathbf{A}_k - z\mathbf{I})^{-1} = \mathbf{A}_k^{-1}(z)$ and $\mathbf{r}_k^T = \mathbf{s}_k^T\mathbf{T}^{1/2}$. We will frequently use the following formulas throughout this paper:

$$
(\mathbf{C} + \mathbf{u}_k\mathbf{v}_k^T)^{-1} = \mathbf{C}^{-1} - \frac{\mathbf{C}^{-1}\mathbf{u}_k\mathbf{v}_k^T\mathbf{C}^{-1}}{1 + \mathbf{v}_k^T\mathbf{C}^{-1}\mathbf{u}_k}, \tag{5.8}
$$

$$
(\mathbf{C} + \mathbf{u}_k\mathbf{v}_k^T)^{-1}\mathbf{u}_k = \frac{\mathbf{C}^{-1}\mathbf{u}_k}{1 + \mathbf{v}_k^T\mathbf{C}^{-1}\mathbf{u}_k}, \tag{5.9}
$$

and

$$
\mathbf{C}^{-1} - \mathbf{B}^{-1} = \mathbf{C}^{-1}(\mathbf{B} - \mathbf{C})\mathbf{B}^{-1}, \tag{5.10}
$$

holding for any two invertible matrices $\mathbf{C}$ and $\mathbf{B}$ of size $p \times p$, and $\mathbf{u}_k, \mathbf{v}_k \in \mathbb{R}^p$.

We then apply (5.10) and (5.9) to write

$$\frac{1}{p}tr\mathbf{A}^{-1}(z)\mathbf{D} - E\frac{1}{p}tr\mathbf{A}^{-1}(z)\mathbf{D} = \frac{1}{p}\sum_{k=1}^{n} E_k(tr\mathbf{A}^{-1}(z)\mathbf{D}) - E_{k-1}(tr\mathbf{A}^{-1}(z)\mathbf{D})$$
$$= \frac{1}{p}\sum_{k=1}^{n}(E_k - E_{k-1})tr\Big(\mathbf{A}^{-1}(z)\mathbf{D} - tr\mathbf{A}_k^{-1}(z)\mathbf{D}\Big)$$
$$= \frac{1}{p}\sum_{k=1}^{n}(E_k - E_{k-1})tr\Big(\mathbf{A}^{-1}(z)(\mathbf{A}_k - \mathbf{A})\mathbf{A}_k^{-1}(z)\mathbf{D}\Big)$$
$$= -\frac{1}{p}\sum_{k=1}^{n}(E_k - E_{k-1})\Big(\frac{\frac{1}{\sqrt{np}}\mathbf{r}_k^T\mathbf{A}_k^{-1}(z)\mathbf{D}\mathbf{A}_k^{-1}(z)\mathbf{r}_k}{1+\frac{1}{\sqrt{np}}\mathbf{r}_k^T\mathbf{A}_k^{-1}(z)\mathbf{r}_k}\Big),$$
$$= -\frac{1}{p}\sum_{k=1}^{n}(E_k - E_{k-1})(f_{n1}+f_{n2}),$$

where

$$f_{n1} = \Big(\frac{1}{\sqrt{np}}\mathbf{r}_k^T\mathbf{A}_k^{-1}(z)\mathbf{D}\mathbf{A}_k^{-1}(z)\mathbf{r}_k - \frac{1}{\sqrt{np}}tr\mathbf{A}_k^{-1}(z)\mathbf{D}\mathbf{A}_k^{-1}(z)\mathbf{T}\Big)\frac{1}{1+\frac{1}{\sqrt{np}}tr\mathbf{A}_k^{-1}(z)\mathbf{T}},$$
$$f_{n2} = \frac{1}{\sqrt{np}}\mathbf{r}_k^T\mathbf{A}_k^{-1}(z)\mathbf{D}\mathbf{A}_k^{-1}(z)\mathbf{r}_k\frac{\frac{1}{\sqrt{np}}\mathbf{r}_k^T\mathbf{A}_k^{-1}(z)\mathbf{r}_k - \frac{1}{\sqrt{np}}tr\mathbf{A}_k^{-1}(z)\mathbf{T}}{(1+\frac{1}{\sqrt{np}}tr\mathbf{A}_k^{-1}(z)\mathbf{T})(1+\frac{1}{\sqrt{np}}\mathbf{r}_k^T\mathbf{A}_k^{-1}(z)\mathbf{r}_k)}.$$

In the last step above we use

$$\frac{1}{1+\frac{1}{\sqrt{np}}\mathbf{r}_k^T\mathbf{A}_k^{-1}(z)\mathbf{r}_k} = \frac{1}{1+\frac{1}{\sqrt{np}}tr\mathbf{A}_k^{-1}(z)\mathbf{T}}$$
$$- \frac{\frac{1}{\sqrt{np}}\mathbf{r}_k^T\mathbf{A}_k^{-1}(z)\mathbf{r}_k - \frac{1}{\sqrt{np}}tr\mathbf{A}_k^{-1}(z)\mathbf{T}}{(1+\frac{1}{\sqrt{np}}\mathbf{r}_k^T\mathbf{A}_k^{-1}(z)\mathbf{r}_k)(1+\frac{1}{\sqrt{np}}tr\mathbf{A}_k^{-1}(z)\mathbf{T})},$$
$$E_k\left(\frac{\frac{1}{\sqrt{np}}tr\mathbf{A}_k^{-1}(z)\mathbf{D}\mathbf{A}_k^{-1}(z)\mathbf{T}}{1+\frac{1}{\sqrt{np}}tr\mathbf{A}_k^{-1}(z)\mathbf{T}}\right) = E_{k-1}\left(\frac{\frac{1}{\sqrt{np}}tr\mathbf{A}_k^{-1}(z)\mathbf{D}\mathbf{A}_k^{-1}(z)\mathbf{T}}{1+\frac{1}{\sqrt{np}}tr\mathbf{A}_k^{-1}(z)\mathbf{T}}\right).$$

Note that

$$\left|\frac{\frac{1}{\sqrt{np}}\mathbf{r}_k^T\mathbf{A}_k^{-1}(z)\mathbf{D}\mathbf{A}_k^{-1}(z)\mathbf{r}_k}{1+\frac{1}{\sqrt{np}}\mathbf{r}_k^T\mathbf{A}_k^{-1}(z)\mathbf{r}_k}\right| \le \frac{\frac{1}{\sqrt{np}}\|\mathbf{r}_k^T\mathbf{A}_k^{-1}(z)\|^2\|\mathbf{D}\|}{Im(1+\frac{1}{\sqrt{np}}\mathbf{r}_k^T\mathbf{A}_k^{-1}(z)\mathbf{r}_k)} \le \frac{1}{v}. \tag{5.11}$$

Also, since

$$\left|\frac{1}{\sqrt{np}}tr\mathbf{A}_k^{-1}(z)\mathbf{T}\right| \le \sqrt{\frac{p}{n}}\frac{M}{v}, \tag{5.12}$$

we have

$$\left|\frac{1}{1+\frac{1}{\sqrt{np}}tr\mathbf{A}_k^{-1}(z)\mathbf{T}}\right| \leq \frac{1}{1-\sqrt{\frac{p}{n}}\frac{M}{v}} \to 1. \tag{5.13}$$

Therefore by Burkholder's inequality for the martingale differences in [6] and Lemma 2.7 in [2] we obtain

$$E\left|\frac{1}{p}\sum_{k=1}^{n}(E_k - E_{k-1})f_{n2}\right|^2 \leq \frac{M}{p^2}E\sum_{k=1}^{n}E\left[|f_{n2}|^2\right] \leq \frac{M}{p^2}. \tag{5.14}$$

Similarly, one can also obtain

$$E\left|\frac{1}{p}\sum_{k=1}^{n}(E_k - E_{k-1})f_{n1}\right|^2 \leq \frac{M}{p^2}. \tag{5.15}$$

Thus the proof of (5.1) is completed. It follows from Borel–Cantelli's lemma and (5.1) that

$$\frac{1}{p}\left(tr\left(\mathbf{A}^{-1}(z)\right) - E\left[tr\left(\mathbf{A}^{-1}(z)\right)\right]\right) \xrightarrow{a.s.} 0. \tag{5.16}$$

5.2. **Step 2: Convergence of $E\left[\frac{1}{p}tr\mathbf{A}^{-1}(z)\right]$ with the Gaussian elements.** The aim in this subsection is to find the limit of $E\left[\frac{1}{p}tr\mathbf{A}^{-1}(z)\right]$ when X_{ij}'s are i.i.d. Gaussian random variables with $E\left[X_{ij}\right]=0$ and $E\left[X_{ij}^2\right]=1$.

Recalling $\mathbf{T}=\mathbf{U}^T\Lambda\mathbf{U}$, pre-multiplying and post-multiplying $\mathbf{A}^{-1}(z)$, respectively, by $\mathbf{U}$ and $\mathbf{U}^T$ we obtain a key identity

$$E\left[\frac{1}{p}tr\mathbf{A}^{-1}(z)\right] = E\left[\frac{1}{p}tr\left(\sqrt{\frac{n}{p}}\left(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T-\Lambda\right)-z\mathbf{I}_p\right)^{-1}\right], \tag{5.17}$$

where $\mathbf{Y}=(\hat{\mathbf{y}}_1,\cdots,\hat{\mathbf{y}}_n)=\left(\mathbf{y}_1^T,\cdots,\mathbf{y}_p^T\right)^T=(Y_{kj})_{p\times n}$ and $\hat{\mathbf{y}}_k$ are independent Gaussian vectors with covariance matrix Λ. In addition, we remark that $\mathbf{y}_k$'s are also independent Gaussian vectors and, moreover, the components of each $\mathbf{y}_k$ are i.i.d Gaussian random variables with $E\left[Y_{kk}^2\right]=\lambda_k$. Here we would remind the reader that $E\left[Y_{kk}^4\right]\leq M$. Consequently, it is enough to investigate the matrix on the right hand side of (5.17).

Before proceeding, let us introduce some notation. Let $\mathbf{e}_k$ be the $p\times 1$ vector with the k-th element being 1 and others zero and

$$\begin{aligned}
\mathbf{h}_k^T &= \frac{1}{\sqrt{np}}(\mathbf{y}_k^T\mathbf{y}_1,\cdots,\mathbf{y}_k^T\mathbf{y}_{k-1},\mathbf{y}_k^T\mathbf{y}_k-n\lambda_k,\mathbf{y}_k^T\mathbf{y}_{k+1}\cdots,\mathbf{y}_k^T\mathbf{y}_p),\\
\mathbf{Y}^{-1}(z) &= (\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T-\Lambda)-z\mathbf{I}_p)^{-1},\ \mathbf{Y}_k^{-1}(z)=(\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{Y}_k\mathbf{Y}^T-\Lambda_k)-z\mathbf{I}_p)^{-1},\\
\mathbf{Y}_{(k)}^{-1}(z) &= (\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{Y}_k\mathbf{Y}_k^T-\Lambda_k)-z\mathbf{I}_p)^{-1},\ a_n=\frac{1}{p}E\Big(\sum_{k=1}^p\frac{\lambda_k}{z(1+\mathbf{h}_k^T\mathbf{Y}_k^{-1}(z)\mathbf{e}_k)}\Big),
\end{aligned}$$

where the matrix $\mathbf{Y}_k$ is obtained from $\mathbf{Y}$ with the entries on its k-th row being replaced by zero, Λ_k obtained from Λ with the k-th diagonal element being replaced by zero and $\mathbf{I}_p$ is the identity matrix of size p.

Apparently, we have

$$\mathbf{Y}=\mathbf{Y}_k+\mathbf{e}_k\mathbf{y}_k^T. \tag{5.18}$$

With respect to the above notation we would make the following remarks: $\mathbf{h}_k^T$ is the k-th row of $\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T-\Lambda)$; $\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{Y}_k\mathbf{Y}^T-\Lambda_k)$ is obtained from $\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T-\Lambda)$ with the entries on its k-th row being replaced by zero; and $\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{Y}_k\mathbf{Y}_k^T-\Lambda_k)$ is obtained from $\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{Y}_k\mathbf{Y}^T-\Lambda_k)$ with the entries on its k-th column being replaced by zero.

Write

$$\sqrt{\frac{n}{p}}\left(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T-\Lambda\right)=\sum_{k=1}^p\mathbf{e}_k\mathbf{h}_k^T. \tag{5.19}$$

Then, we conclude from (5.10) and (5.9) that

$$\begin{aligned}
&\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\right)-\frac{1}{p}tr\left((a_n\Lambda-z\mathbf{I}_p)^{-1}\right)\\
&= \frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\Big(a_n\Lambda-\sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T-\Lambda)\Big)(a_n\Lambda-z\mathbf{I}_p)^{-1}\right)\\
&= \frac{a_n}{p}tr\left(\mathbf{Y}^{-1}(z)\Lambda(a_n\Lambda-z\mathbf{I}_p)^{-1}\right)-\frac{1}{p}\sum_{k=1}^p\mathbf{h}_k^T(a_n\Lambda-z\mathbf{I}_p)^{-1}\mathbf{Y}^{-1}(z)\mathbf{e}_k\\
&= \frac{a_n}{p}tr\left(\mathbf{Y}^{-1}(z)\Lambda(a_n\Lambda-z\mathbf{I}_p)^{-1}\right)-\frac{1}{p}\sum_{k=1}^p\frac{\mathbf{h}_k^T(a_n\Lambda-z\mathbf{I}_p)^{-1}\mathbf{Y}_k^{-1}(z)\mathbf{e}_k}{1+\mathbf{h}_k^T\mathbf{Y}_k^{-1}(z)\mathbf{e}_k}.
\end{aligned}$$

First taking expectation on both sides of the equality above and then using the definition of a_n we obtain

$$E\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\right) - E\left[\frac{1}{p}tr\left(a_n\Lambda - z\mathbf{I}_p\right)^{-1}\right]\right] \tag{5.20}$$
$$= -\frac{1}{p}\sum_{k=1}^{p} E\left(\frac{z\mathbf{h}_k^T(a_n\Lambda - z\mathbf{I}_p)^{-1}\mathbf{Y}_k^{-1}(z)\mathbf{e}_k - \lambda_k E\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\Lambda(a_n\Lambda - z\mathbf{I}_p)^{-1}\right)\right]}{z(1+\mathbf{h}_k^T\mathbf{Y}_k^{-1}(z)\mathbf{e}_k)}\right).$$

We then investigate $z\mathbf{h}_k^T\mathbf{C}\mathbf{Y}_k^{-1}(z)\mathbf{e}_k$ with $\mathbf{C}$ equal to $(a_n\Lambda - z\mathbf{I}_p)^{-1}$ or $\mathbf{I}$. By definition of $\mathbf{h}_k^T$ we have

$$\mathbf{h}_k^T = \frac{1}{\sqrt{np}}\mathbf{y}_k^T\mathbf{Y}_k^T + \sqrt{\frac{n}{p}}\left(\frac{\mathbf{y}_k^T\mathbf{y}_k}{n} - \lambda_k\right)\mathbf{e}_k^T. \tag{5.21}$$

This, together with (5.8) and (5.18), ensures that

$$z\mathbf{h}_k^T\mathbf{C}\mathbf{Y}_k^{-1}(z)\mathbf{e}_k = \frac{z}{\sqrt{np}}\mathbf{y}_k^T\mathbf{Y}_k^T\mathbf{C}\mathbf{Y}_k^{-1}(z)\mathbf{e}_k \tag{5.22}$$
$$+ z\sqrt{\frac{n}{p}}(\frac{\mathbf{y}_k^T\mathbf{y}_k}{n} - \lambda_k)\mathbf{e}_k^T\mathbf{C}\mathbf{Y}_k^{-1}(z)\mathbf{e}_k$$
$$= \frac{z}{\sqrt{np}}\mathbf{y}_k^T\mathbf{Y}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{e}_k - \frac{z}{np}\frac{\mathbf{y}_k^T\mathbf{Y}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k\mathbf{e}_k^T\mathbf{Y}_{(k)}^{-1}(z)\mathbf{e}_k}{1+\mathbf{e}_k^T\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k/\sqrt{np}}$$
$$+ z\sqrt{\frac{n}{p}}(\frac{\mathbf{y}_k^T\mathbf{y}_k}{n} - \lambda_k)\mathbf{e}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{e}_k - \frac{z}{p}\frac{(\frac{\mathbf{y}_k^T\mathbf{y}_k}{n} - \lambda_k)\mathbf{e}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k\mathbf{e}_k^T\mathbf{Y}_{(k)}^{-1}(z)\mathbf{e}_k}{1+\mathbf{e}_k^T\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k/\sqrt{np}}$$
$$= -\frac{1}{\sqrt{np}}\mathbf{y}_k^T\mathbf{Y}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{e}_k + \frac{1}{np}\mathbf{y}_k^T\mathbf{Y}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k$$
$$-\sqrt{\frac{n}{p}}(\frac{\mathbf{y}_k^T\mathbf{y}_k}{n} - \lambda_k)\mathbf{e}_k^T\mathbf{C}\mathbf{e}_k + \frac{1}{p}(\frac{\mathbf{y}_k^T\mathbf{y}_k}{n} - \lambda_k)\mathbf{e}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k.$$

The last step is based on the following observation. Since the entries on the k-th row and k-th column of $\sqrt{\frac{n}{p}}\left(\frac{1}{n}\mathbf{Y}_k\mathbf{Y}_k^T - \Lambda_k\right) - z\mathbf{I}_p$ are all zero except that the entry on the (k,k) position is $-z$, we have

$$\mathbf{Y}_{(k)}^{-1}(z)\mathbf{e}_k = -\frac{1}{z}\mathbf{e}_k \quad \text{and} \quad \mathbf{e}_k^T\mathbf{Y}_{(k)}^{-1}(z)\mathbf{e}_k = -\frac{1}{z}. \tag{5.23}$$

Also, by the structure of $\mathbf{Y}_k$ we have

$$\mathbf{e}_k^T\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k/\sqrt{np} = 0. \tag{5.24}$$

Applying (5.22) with $\mathbf{C} = \mathbf{I}$ yields that the imaginary part of $z\mathbf{h}_k^T\mathbf{Y}_k^{-1}(z)\mathbf{e}_k$ is nonnegative. That is

$$Im(z\mathbf{h}_k^T\mathbf{Y}_k^{-1}(z)\mathbf{e}_k) \geq 0. \tag{5.25}$$

This implies

$$Im(-a_n) \geq 0. \tag{5.26}$$

Thus we have

$$\|\mathbf{C}\| \leq \max(1/v, 1).$$

As will be seen, the second term on the right hand side of the equality (5.22) contributes to the limit and all the remaining terms are negligible.

We now demonstrate the details. A simple calculation implies

$$E\left|\sqrt{\frac{n}{p}}(\frac{\mathbf{y}_k^T\mathbf{y}_k}{n} - \lambda_k)\mathbf{e}_k^T\mathbf{C}\mathbf{e}_k\right|^2 \leq \frac{nM}{p}E\left|\frac{\mathbf{y}_k^T\mathbf{y}_k}{n} - \lambda_k\right|^2 \leq \frac{M}{p}. \tag{5.27}$$

With $\mathbf{x} = (x_1, \cdots, x_n)^T = \mathbf{e}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k$, we obtain

$$\begin{aligned}
& E\left|\mathbf{e}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k\right|^2 = \sum_{j=1}^n E\left[x_j^2 Y_{kj}^2\right] \leq M\ E\left[\left|\mathbf{x}^T\mathbf{x}\right|^2\right] \\
=\ & M\ E\left[\left|\mathbf{e}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{Y}_k^T\mathbf{Y}_{(k)}^{-1}(\bar{z})\bar{\mathbf{C}}\mathbf{e}_k\right|\right] \\
\leq\ & M\ E\left(\left|\left|\mathbf{e}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\|\|\mathbf{Y}_k\mathbf{Y}_k^T\mathbf{Y}_{(k)}^{-1}(\bar{z})\right|\right| \cdot ||\bar{\mathbf{C}}\mathbf{e}_k||\right) \\
\leq\ & M\sqrt{np}E\left[\left|\left|\mathbf{Y}_{(k)}^{-1}(\bar{z})\frac{\mathbf{Y}_k\mathbf{Y}_k^T}{\sqrt{np}}\right|\right|\right] \\
\leq\ & M\sqrt{np}E\left[||I_p||\right] + M\sqrt{np}E\left[\left|\left|\mathbf{Y}_{(k)}^{-1}(z)\left(\sqrt{\frac{n}{p}}\Lambda_k + z\mathbf{I}_p\right)\right|\right|\right], \\
\leq\ & M\ n, \qquad (5.28)
\end{aligned}$$

where $\bar{\mathbf{C}}$ is the complex conjugate of $\mathbf{C}$ and $\mathbf{Y}_{(k)}^{-1}(\bar{z})$ the complex conjugate of $\mathbf{Y}_{(k)}^{-1}(z)$.

This, together with Holder's inequality, implies

$$
\begin{aligned}
& E\left|\frac{1}{p}(\frac{\mathbf{y}_k^T\mathbf{y}_k}{n}-\lambda_k)\mathbf{e}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k\right| \\
\leq\ & \frac{1}{p}\left(E\left[\left|\frac{\mathbf{y}_k^T\mathbf{y}_k}{n}-\lambda_k\right|^2\right]\cdot E\left[\left|\mathbf{e}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k\right|^2\right]\right)^{1/2} \\
\leq\ & \frac{M}{p}. \qquad (5.29)
\end{aligned}
$$

The argument for (5.28) also gives $E\left[\left|\frac{1}{\sqrt{np}}\mathbf{y}_k^T\mathbf{Y}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{e}_k\right|^2\right]\leq nM$. Thus all terms except the second term in (5.22) are negligible, as claimed.

Consider the second term in (5.22) now. We conclude from Lemma 2.7 in [4] that

$$
\begin{aligned}
& E\left[\left|\frac{1}{np}\mathbf{y}_k^T\mathbf{Y}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k-\frac{\lambda_k}{np}tr\mathbf{Y}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\right|^2\right] \\
\leq\ & \frac{M}{n^2p^2}E\left(tr\left(\mathbf{Y}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{Y}_k^T\mathbf{Y}_{(k)}^{-1}(\bar{z})\bar{\mathbf{C}}\mathbf{Y}_k\right)\right)\leq\frac{M}{p}, \qquad (5.30)
\end{aligned}
$$

because of

$$
\begin{aligned}
& \frac{1}{n^2p^2}tr\left(C\mathbf{Y}_k^{-1}(z)\mathbf{Y}_k\mathbf{Y}_k^T\mathbf{Y}_k^{-1}(\bar{z})\bar{\mathbf{C}}\mathbf{Y}_k\mathbf{Y}_k^T\right) \\
\leq\ & \frac{1}{n^2p}\left|\left|C\mathbf{Y}_k^{-1}(z)\mathbf{Y}_k\mathbf{Y}_k^T\right|\right|\cdot\left|\left|\mathbf{Y}_k^{-1}(\bar{z})\bar{\mathbf{C}}\mathbf{Y}_k\mathbf{Y}_k^T\right|\right| \\
\leq\ & \frac{\|\mathbf{C}\|^2}{n^2p}\left|\left|\mathbf{Y}_k^{-1}(z)\mathbf{Y}_k\mathbf{Y}_k^T\right|\right|^2\leq\frac{\|\mathbf{C}\|^2}{n}\|\mathbf{I}_p\|^2+\frac{\|\mathbf{C}\|^2}{n}\left|\left|\mathbf{Y}_k^{-1}(z)\left(\sqrt{\frac{n}{p}}\Lambda_k+z\mathbf{I}_p\right)\right|\right|^2\leq\frac{M}{p}.
\end{aligned}
$$

Meanwhile, we also have

$$
\frac{\lambda_k}{np}tr\left(\mathbf{Y}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\right)=\frac{\lambda_k}{\sqrt{np}}tr\left(\mathbf{C}\right)+\frac{\lambda_k}{p}tr\left(\mathbf{C}\mathbf{Y}_k^{-1}(z)\Lambda_k\right)+z\frac{\lambda_k}{\sqrt{np}}tr\left(\mathbf{C}\mathbf{Y}_k^{-1}(z)\right),
$$

which implies

$$
\left|\frac{\lambda_k}{np}tr\left(\mathbf{Y}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\right)-\frac{\lambda_k}{p}tr\left(\mathbf{C}\mathbf{Y}_k^{-1}(z)\Lambda_k\right)\right|\leq\frac{M\sqrt{p}}{\sqrt{n}}\to 0.
$$

The next aim is to prove that

$$
E\left|\frac{1}{p}tr\left(\mathbf{C}\mathbf{Y}_k^{-1}(z)\Lambda_k\right)-\frac{1}{p}tr\left(\mathbf{C}\mathbf{Y}^{-1}(z)\Lambda\right)\right|\leq\frac{M}{p}. \qquad (5.31)
$$

Evidently,

$$
\left|\frac{1}{p}tr\left(\mathbf{C}\mathbf{Y}_k^{-1}(z)\Lambda_k\right)-\frac{1}{p}tr\left(\mathbf{C}\mathbf{Y}_k^{-1}(z)\Lambda\right)\right|\leq\frac{\lambda_k\|\mathbf{C}\|}{pv}\leq\frac{M}{pv}.
$$

Moreover, we conclude from (5.10) and (5.18) that

$$\begin{aligned}
&\frac{1}{p}tr\left(\mathbf{C}\mathbf{Y}_k^{-1}(z)\Lambda\right)-\frac{1}{p}tr\left(\mathbf{C}\mathbf{Y}^{-1}(z)\Lambda\right)\\
=\ &\frac{1}{p}tr\left(\Lambda\mathbf{C}\mathbf{Y}_k^{-1}(z)\right)\left[\sqrt{\frac{n}{p}}\left(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T-\Lambda\right)-\sqrt{\frac{n}{p}}\left(\frac{1}{n}\mathbf{Y}_k\mathbf{Y}_k^T-\Lambda_k\right)\right]\mathbf{Y}^{-1}(z)\\
=\ &b_{n1}+b_{n2}+b_{n3},
\end{aligned}$$

where

$$\begin{aligned}
b_{n1}&=\frac{1}{p}\sqrt{\frac{n}{p}}\left(\frac{\mathbf{y}_k^T\mathbf{y}_k}{n}-\lambda_k\right)\mathbf{e}_k^T\mathbf{Y}^{-1}(z)\Lambda\mathbf{C}\mathbf{Y}_k^{-1}(z)\mathbf{e}_k\\
&=\frac{1}{p}\sqrt{\frac{n}{p}}\left(\frac{\mathbf{y}_k^T\mathbf{y}_k}{n}-\lambda_k\right)\left[\mathbf{e}_k^T\mathbf{Y}^{-1}(z)\Lambda\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{e}_k+\frac{1}{z\sqrt{np}}\mathbf{e}_k^T\mathbf{Y}^{-1}(z)\Lambda\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k\right],\\
b_{n2}&=\frac{1}{p\sqrt{np}}\mathbf{y}_k^T\mathbf{Y}_k^T\mathbf{Y}^{-1}(z)\Lambda\mathbf{C}\mathbf{Y}_k^{-1}(z)\mathbf{e}_k\\
&=\frac{1}{p\sqrt{np}}\mathbf{y}_k^T\mathbf{Y}_k^T\mathbf{Y}^{-1}(z)\Lambda\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{e}_k+\frac{1}{znp^2}\mathbf{y}_k^T\mathbf{Y}_k^T\mathbf{Y}^{-1}(z)\Lambda\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k,\\
b_{n3}&=\frac{1}{p\sqrt{np}}\mathbf{e}_k^T\mathbf{Y}^{-1}(z)\Lambda\mathbf{C}\mathbf{Y}_k^{-1}(z)\mathbf{Y}_k\mathbf{y}_k=\frac{1}{p\sqrt{np}}\mathbf{e}_k^T\mathbf{Y}^{-1}(z)\Lambda\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k.
\end{aligned}$$

Here the further simplified expressions for $b_{nj}, j=1,2,3$ are obtained by (5.23) and (5.24), as in (5.22). The arguments for (5.28) and (5.29) imply that the first absolute moments of b_{n1}, b_{n3} and the first term of b_{n2} have an order of $1/p$.

As for the second term of b_{n2}, we have

$$\begin{aligned}
&E\left||\frac{1}{np^2}\mathbf{y}_k^T\mathbf{Y}_k^T\mathbf{Y}^{-1}(z)\Lambda\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k\right|\\
\leq\ &\frac{M}{np^2}E\Big(\|\mathbf{y}_k^T\mathbf{Y}_k^T\|^2\cdot\|\mathbf{Y}^{-1}(z)\Lambda\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\|\Big)\leq\frac{M}{np^2}E\|\mathbf{y}_k^T\mathbf{Y}_k^T\|^2\\
\leq\ &\frac{M\alpha_p^2}{np^3}E\left[\left|\mathbf{y}_k^T\mathbf{Y}_k^T\mathbf{Y}_k\mathbf{y}_k-tr\mathbf{Y}_k^T\mathbf{Y}_k\right|\right]+\frac{M}{np^3}E\left[\left|tr\left(\mathbf{Y}_k^T\mathbf{Y}_k\right)\right|\right]\\
\leq\ &\frac{M}{np^2}\left(E\left[tr\left(\mathbf{Y}_k^T\mathbf{Y}_k\right)\right]^2\right)^{1/2}\leq\frac{M}{np^2}\left(p\sum_{j=1}^{p}E\left(\mathbf{y}_j^T\mathbf{y}_j\right)^2\right)^{1/2}\leq\frac{M}{p}.
\end{aligned}\tag{5.32}$$

Thus, equation (5.31) follows.

Repeating the argument for (5.1) we may obtain

$$E\left|\frac{1}{p}tr\left(\mathbf{C}\mathbf{Y}^{-1}(z)\Lambda\right)-E\left[\frac{1}{p}tr\left(\mathbf{C}\mathbf{Y}^{-1}(z)\Lambda\right)\right]\right|^2\leq\frac{M}{p^2}.\tag{5.33}$$

Thus, summarizing the argument from (5.30)-(5.33) we have proved that

$$E\left|-\frac{1}{np}\mathbf{y}_k^T\mathbf{Y}_k^T\mathbf{C}\mathbf{Y}_{(k)}^{-1}(z)\mathbf{Y}_k\mathbf{y}_k + E\left[\frac{\lambda_k}{p}tr\mathbf{C}\mathbf{Y}^{-1}(z)\Lambda\right]\right| \leq \frac{M}{\sqrt{p}}. \tag{5.34}$$

It follows from from (5.22), (5.27), (5.28) and (5.34) that

$$E\left|z\mathbf{h}_k^T\mathbf{C}\mathbf{Y}_k^{-1}(z)\mathbf{e}_k - E\left[\frac{\lambda_k}{p}tr\left(\mathbf{C}\mathbf{Y}^{-1}(z)\Lambda\right)\right]\right| \leq \frac{M}{\sqrt{p}}. \tag{5.35}$$

We then conclude from (5.25), (5.35) and (5.20) that

$$E\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\right)\right] - E\left[\frac{1}{p}tr\left((a_n\Lambda - z\mathbf{I}_p)^{-1}\right)\right] \to 0 \quad \text{as} \quad p\to\infty. \tag{5.36}$$

Moreover, denote the spectral decomposition of $\mathbf{Y}^{-1}(z)$ by

$$\mathbf{V}_n^T\mathbf{Y}^{-1}(z)\mathbf{V}_n = diag\left(\frac{1}{\mu_1 - z}, \cdots, \frac{1}{\mu_p - z}\right),$$

where $\mu_1, \cdots, \mu_p$ are eigenvalues of $\sqrt{\frac{n}{p}}\left(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T - \Lambda\right)$ and $\mathbf{V}_n$ is the corresponding eigenvector matrix. It follows that

$$\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\Lambda\right) = \frac{1}{p}\sum_{k=1}^{p}\frac{(\mathbf{V}^T\Lambda\mathbf{V})_{kk}}{\mu_k - z},$$

where $(\cdot)_{kk}$ is the k-th diagonal element of $\mathbf{V}^T\Lambda\mathbf{V}$. This implies

$$Im(z + \lambda_k E\frac{1}{p}tr\mathbf{Y}^{-1}(z)\Lambda) = v + v\frac{\lambda_k}{p}\sum_{k=1}^{p}E\left[\frac{(\mathbf{V}^T\Lambda\mathbf{V})_{kk}}{|\mu_k - z|^2}\right] \geq v, \tag{5.37}$$

because for each k

$$(\mathbf{V}^T\Lambda\mathbf{V})_{kk} \geq \lambda_{\min}(\mathbf{V}^T\Lambda\mathbf{V}) \geq 0,$$

where $\lambda_{\min}(\mathbf{V}^T\Lambda\mathbf{V})$ stands for the minimum eigenvalue of $\mathbf{V}^T\Lambda\mathbf{V}$.

Thus, applying (5.35) with $\mathbf{C} = \mathbf{I}$ and (5.25) we have

$$a_n - \frac{1}{p}\sum_{k=1}^{p}\frac{\lambda_k}{z + \lambda_k E\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\Lambda\right)\right]} \to 0. \tag{5.38}$$

It is necessary to have one more equation to find a solution from (5.36) and (5.38). To this end, as in (5.19), write

$$\sqrt{\frac{n}{p}}\left(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T - \Lambda\right) - z\mathbf{I} = \sum_{k=1}^{p}\mathbf{e}_k\mathbf{h}_k^T - z\mathbf{I}$$

Post–multiplying both sides of the above equality by $\mathbf{Y}^{-1}(z)$, then taking trace and expectation, and finally dividing by p on both sides of the above equality, we obtain

$$1 = \frac{1}{p}\sum_{k=1}^{p} E\Big(\mathbf{h}_k^T\mathbf{Y}^{-1}(z)\mathbf{e}_k\Big) - zE\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\right)\right].$$

Furthermore, equation (5.9) yields

$$1 = \frac{1}{p}\sum_{k=1}^{p} E\Big(\frac{\mathbf{h}_k^T\mathbf{Y}_k^{-1}(z)\mathbf{e}_k}{1+\mathbf{h}_k^T\mathbf{Y}_k^{-1}(z)\mathbf{e}_k}\Big) - zE\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\right)\right],$$

which is equivalent to

$$\frac{1}{p}\sum_{k=1}^{p} E\Big(\frac{1}{z(1+\mathbf{h}_k^T\mathbf{Y}_k^{-1}(z)\mathbf{e}_k)}\Big) = -E\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\right)\right].$$

Applying (5.35) with $\mathbf{C} = \mathbf{I}$, together with (5.25) and (5.37), ensures that as $p \to \infty$,

$$E\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\right)\right] + \frac{1}{p}\sum_{k=1}^{p} E\left(\frac{1}{z+\lambda_k E\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\Lambda\right)\right]}\right) \to 0. \tag{5.39}$$

Since $E\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\right)\right]$ and $E\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\Lambda\right)\right]$ are both bounded we may choose a subsequence p' such that $E\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\right)\right]$ and $E\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\Lambda\right)\right]$ converge to their respective limits, say $s_1(z)$ and $s_2(z)$, as $p' \to \infty$.

In addition, by (5.26)

$$Im(-a_n t + z) \geq v$$

and it is verified in the next subsection that

$$Im(-t\int\frac{xdH(x)}{z+xs_2(z)} + z) \geq v. \tag{5.40}$$

Thus

$$\left|\frac{1}{a_n t - z} - \frac{1}{t\int\frac{xdH(x)}{z+xs_2(z)} - z}\right| \leq \frac{M}{v^2}\left|a_n - \int\frac{xdH(x)}{z+xs_2(z)}\right|$$

and by (5.37)

$$\left|\frac{x}{z+E\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\Lambda\right)\right]} - \frac{x}{z+xs_2(z)}\right| \leq \frac{M}{v^2}\left|E\left[\frac{1}{p}\left(tr\mathbf{Y}^{-1}(z)\Lambda\right)\right] - s_2(z)\right|.$$

It follows from (5.36), (5.38) and (5.39) that

$$s_1(z) = \int \frac{dH(t)}{t\int \frac{xdH(x)}{z+xs_2(z)} - z} \tag{5.41}$$

and

$$s_1(z) = -\int \frac{dH(t)}{z+ts_2(z)}. \tag{5.42}$$

When

$$H(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases},$$

equation (5.41) or (5.42) determines $s_1(z) = -1/z$.

In what follows, suppose that $H(t)$ is not a degenerate distribution at the point zero. By (5.41) and (5.42), $s_2(z)$ satisfies

$$\int \frac{dH(t)}{z+ts_2(z)} = -\int \frac{dH(t)}{t\int \frac{xdH(x)}{z+xs_2(z)} - z}. \tag{5.43}$$

This is equivalent to

$$\left(s_2(z) + \int \frac{xdH(x)}{z+xs_2(z)}\right)\left(\int \frac{t}{(z+ts_2(z))(t\int \frac{xdH(x)}{z+xs_2(z)} - z)} dH(t)\right) = 0.$$

Moreover, it is shown in the next subsection that $s_2(z)$ is the unique solution to (2.7) in $\mathbb{C}^+$ and that

$$\int \frac{t}{(z+ts_2(z))(t\int \frac{xdH(x)}{z+xs_2(z)} - z)} dH(t) \neq 0. \tag{5.44}$$

Therefore, we have

$$E\left[\frac{1}{p} tr\left(\mathbf{A}^{-1}(z)\right)\right] \to -\int \frac{dH(t)}{z+ts_2(z)}, \tag{5.45}$$

where $s_2(z)$ is the unique solution in $\mathbb{C}^+$ to the equation below

$$s_2(z) = -\int \frac{xdH(x)}{z+xs_2(z)}. \tag{5.46}$$

5.3. **Step 3: Proof of (5.44) and uniqueness of solution of (5.46).** In this section, we verify (5.44) and (5.40), and show that $s_2(z)$ is the unique solution to (5.46) in $\mathbb{C}^+$. We should keep in mind that $H(t)$ is not a degenerate distribution at the point zero.

We first verify (5.44). Let $z = u+iv$ and $s_2(z) = m_1+im_2$. From (5.37) we see that $Im\left(E\left[\frac{1}{p}tr\left(\mathbf{Y}^{-1}(z)\Lambda\right)\right]\right) \geq 0$ and hence $m_2 \geq 0$. It follows that $v + tm_2 \geq v > 0$ and $t\int \frac{(v+xm_2)xdH(x)}{|z+xs_2(z)|^2} + v \geq v > 0$, which implies (5.40). We calculate the complex number involved in (5.44) as follow.

$$
\begin{aligned}
g(z) &\triangleq \left(\bar{z} + t\overline{s_2(z)}\right)\left(\overline{t\int \frac{xdH(x)}{z+xs_2(z)}} - \bar{z}\right) \\
&= [u + tm_1 - i(v+tm_2)]\left[t\int \frac{(u+xm_1)xdH(x)}{|z+xs_2(z)|^2} - u + i\left(t\int \frac{(v+xm_2)xdH(x)}{|z+xs_2(z)|^2} + v\right)\right] \\
&= (u+tm_1)\left[t\int \frac{(u+xm_1)xdH(x)}{|z+xs_2(z)|^2} - u\right] + (v+tm_2)\left(t\int \frac{(v+xm_2)xdH(x)}{|z+xs_2(z)|^2} + v\right) \\
&\quad + i\left[(u+tm_1)\left(t\int \frac{(v+xm_2)xdH(x)}{|z+xs_2(z)|^2} + v\right) - (v+tm_2)\left(t\int \frac{(u+xm_1)xdH(x)}{|z+xs_2(z)|^2} - u\right)\right],
\end{aligned}
$$

where the symbol " $\overline{x}$ " denotes the complex conjugate of complex number x.

If $(u+tm_1)$ and $\left[t\int \frac{(u+xm_1)xdH(x)}{|z+xs_2(z)|^2} - u\right]$ are both nonnegative or both negative, then the real part of $g(z)$ is positive.

If

$$
(u+tm_1) \geq 0, \qquad \left[t\int \frac{(u+xm_1)xdH(x)}{|z+xs_2(z)|^2} - u\right] < 0,
$$

or

$$
(u+tm_1) < 0, \qquad \left[t\int \frac{(u+xm_1)xdH(x)}{|z+xs_2(z)|^2} - u\right] \geq 0,
$$

then the absolute value of the imaginary part of $g(z)$ is positive. Also, note that the imaginary parts of $z + ts_2(z)$ and $-t\int \frac{xdH(x)}{z+xs_2(z)} + z$ are both greater than v.

Therefore, we have obtained

$$
\begin{aligned}
&\int \frac{t}{(z+ts_2(z))(t\int \frac{xdH(x)}{z+xs_2(z)} - z)}dH(t) \\
&= \int \frac{tg(z)}{|z+ts_2(z)|^2|t\int \frac{xdH(x)}{z+xs_2(z)} - z|^2}dH(t) \neq 0,
\end{aligned}
$$

as claimed.

We next prove uniqueness. Suppose that there is $s_3(z) \in \mathbb{C}^+$ satisfying (5.46). Then, we have

$$\begin{aligned} s_2(z) - s_3(z) &= -\int \frac{xdH(x)}{z + xs_2(z)} + \int \frac{xdH(x)}{z + xs_3(z)} \\ &= (s_2(z) - s_3(z)) \int \frac{x^2 dH(x)}{(z + xs_2(z))(z + xs_3(z))}. \end{aligned} \tag{5.47}$$

Considering the imaginary part of the both sides of (5.46), we have

$$m_2 = \int \frac{xv + x^2 m_2}{|z + xs_2(z)|^2} dH(x) > m_2 \int \frac{x^2}{|z + xs_2(z)|^2} dH(x), \tag{5.48}$$

which implies

$$1 > \int \frac{x^2}{|z + xs_2(z)|^2} dH(x). \tag{5.49}$$

Here one should note that $\int \frac{x}{|z+xs_2(z)|^2} dH(x) \neq 0$ and hence the equality in (5.48) implies $m_2 > 0$. By Holder's inequality

$$\left| \int \frac{x^2 dH(x)}{(z + xs_2(z))(z + xs_3(z))} \right|^2 \leq \int \frac{x^2 dH(x)}{|z + xs_2(z)|^2} \int \frac{x^2 dH(x)}{|z + xs_3(z)|^2} < 1.$$

Therefore, in view of (5.47), $s_3(z)$ must be equal to $s_2(z)$.

5.4. **Step 4: From Gaussian distribution to general distributions.** This subsection is devoted to showing that the limit found for the Gaussian random matrices in the last subsection also applies for the nonGaussian distributions.

Define

$$\frac{1}{p} tr\left((\mathbf{D} - z\mathbf{I})^{-1}\right) = \frac{1}{n} tr\left(\left(\sqrt{\frac{n}{p}}\left(\frac{1}{n}\mathbf{T}^{1/2}\mathbf{W}\mathbf{W}^T\mathbf{T}^{1/2} - \mathbf{T}\right) - z\mathbf{I}\right)^{-1}\right),$$

where $\mathbf{W} = (W_{ij})_{p\times n}$ consists of i.i.d Gaussian random variables with $E[W_{11}] = 0$ and $E\left[W_{11}^2\right] = 1$, and W_{ij} are independent of X_{ij}.

The aim in this subsection is to prove that as $p \to \infty$

$$E\left[\frac{1}{p} tr\left(\mathbf{A}^{-1}(z)\right)\right] - E\left[\frac{1}{p} tr\left((\mathbf{D} - z\mathbf{I})^{-1}\right)\right] \to 0. \tag{5.50}$$

Inspired by [7] and [13], we use Lindeberg's method to prove (5.50). In what follows, to simplify the notation, denote $(\mathbf{A} - z\mathbf{I})^{-2}$ by $\mathbf{A}^{-2}(z)$,

$$X_{11}, \cdots, X_{1n}, X_{21}, \cdots, X_{pn} \text{ by } \hat{X}_1, \cdots, \hat{X}_n, \hat{X}_{n+1}, \cdots, \hat{X}_{pn}$$

and $W_{11}, \cdots, W_{1n}, W_{21}, \cdots, W_{pn}$ by $\hat{W}_1, \cdots, \hat{W}_n, \hat{W}_{n+1}, \cdots, \hat{W}_{pn}$.

For each j, $0 \le j \le pn$, define

$$\mathbf{Z}_j = (\hat{X}_1, \cdots, \hat{X}_j, \hat{W}_{j+1}, \cdots, \hat{W}_{pn}) \text{ and } \mathbf{Z}_j^0 = (\hat{X}_1, \cdots, \hat{X}_{j-1}, 0, \hat{W}_{j+1}, \cdots, \hat{W}_{pn}).$$

Note that all random variables in $\mathbf{A}$ constitute the random vector $\mathbf{Z}_{pn}$ and so denote $\frac{1}{p}tr\left(\mathbf{A}^{-1}(z)\right)$ by $\frac{1}{p}tr\left((\mathbf{A}(\mathbf{Z}_{pn}) - z\mathbf{I})^{-1}\right)$. We then define the mapping f from R^{np} to C as

$$f(\mathbf{Z}_{pn}) = \frac{1}{p}tr\left((\mathbf{A}(\mathbf{Z}_{pn}) - z\mathbf{I})^{-1}\right).$$

Moreover, we use the components of $\mathbf{Z}_j, j = 0, 1, \cdots, pn-1$, respectively, to replace $\hat{X}_1, \cdots, \hat{X}_{pn}$, the corresponding components of $\mathbf{Z}_{pn}$, in $\mathbf{A}$ to form a series of new matrices. For these new matrices, we define $f(\mathbf{Z}_j), j = 0, 1, \cdots, pn-1$ as $f(\mathbf{Z}_{pn})$ is defined for the matrix $\mathbf{A}$. For example, $\frac{1}{p}tr\left((\mathbf{D} - z\mathbf{I})^{-1}\right) = f(\mathbf{Z}_0)$. Then, we write

$$E\left[\frac{1}{p}tr\left(\mathbf{A}^{-1}(z)\right)\right] - E\left[\frac{1}{p}tr\left((\mathbf{D} - z\mathbf{I})^{-1}\right)\right] = \sum_{j=1}^{pn} E\Big(f(\mathbf{Z}_j) - f(\mathbf{Z}_{j-1})\Big).$$

In addition, a third Taylor expansion gives

$$f(\mathbf{Z}_j) = f(\mathbf{Z}_j^0) + \hat{X}_j \partial_j f(\mathbf{Z}_j^0) + \frac{1}{2}\hat{X}_j^2 \partial_j^2 f(\mathbf{Z}_j^0) + \frac{1}{2}\hat{X}_j^3 \int_0^1 (1-t)^2 \partial_j^3 f\left(\mathbf{Z}_j^{(1)}(t)\right) dt,$$

$$f(\mathbf{Z}_{j-1}) = f(\mathbf{Z}_j^0) + \hat{W}_j \partial_j f(\mathbf{Z}_j^0) + \frac{1}{2}\hat{W}_j^2 \partial_j^2 f(\mathbf{Z}_j^0) + \frac{1}{2}\hat{W}_j^3 \int_0^1 (1-t)^2 \partial_j^3 f\left(\mathbf{Z}_{j-1}^{(2)}(t)\right) dt,$$

where $\partial_j^r f(\cdot), r = 1, 2, 3$, denote the r-fold derivative of the function f in the jth coordinate,

$$\mathbf{Z}_j^{(1)}(t) = (\hat{X}_1, \cdots, \hat{X}_{j-1}, t\hat{X}_j, \hat{W}_{j+1}, \cdots, \hat{W}_{pn}) \quad \text{and}$$
$$\mathbf{Z}_{j-1}^{(2)}(t) = (\hat{X}_1, \cdots, \hat{X}_{j-1}, t\hat{W}_j, \hat{W}_{j+1}, \cdots, \hat{W}_{pn}).$$

Note that $\hat{X}_j$ and $\hat{W}_j$ are both independent of $\mathbf{Z}_j^0$, and that $E\left[\hat{X}_j\right] = E\left[\hat{W}_j\right] = 0$ and $E\left[\hat{X}_j^2\right] = E\left[\hat{W}_j^2\right] = 1$. Thus

$$\begin{aligned} (5.51) \quad & E\left[\frac{1}{p}tr\left(\mathbf{A}^{-1}(z)\right)\right] - E\left[\frac{1}{p}tr\left((\mathbf{D} - z\mathbf{I})^{-1}\right)\right] \\ = \quad & \frac{1}{2}\sum_{j=1}^{pn} E\Big[\hat{X}_j^3 \int_0^1 (1-t)^2 \partial_j^3 f\left(\mathbf{Z}_j^{(1)}(t)\right) dt - \hat{W}_j^3 \int_0^1 (1-t)^2 \partial_j^3 f\left(\mathbf{Z}_{j-1}^{(2)}(t)\right) dt\Big]. \end{aligned}$$

Evaluate $\partial_j^3 f\left(\mathbf{Z}_j^{(1)}(t)\right)$ next. In what follows, we make the use of the following results:

$$\frac{1}{p}\frac{\partial tr\left(\mathbf{A}^{-1}(z)\right)}{\partial X_{ij}} = -\frac{1}{p}tr\left(\frac{\partial \mathbf{A}}{\partial X_{ij}}\mathbf{A}^{-2}(z)\right) \tag{5.52}$$

and

$$\frac{\partial \mathbf{A}}{\partial X_{ij}} = \frac{1}{\sqrt{np}}\mathbf{T}^{1/2}\mathbf{e}_i\mathbf{e}_j^T\mathbf{X}^T\mathbf{T}^{1/2} + \frac{1}{\sqrt{np}}\mathbf{T}^{1/2}\mathbf{X}\mathbf{e}_j\mathbf{e}_i^T\mathbf{T}^{1/2}. \tag{5.53}$$

It follows that

$$\frac{1}{p}\frac{\partial^2 tr\left(\mathbf{A}^{-1}(z)\right)}{\partial X_{ij}} = \frac{2}{p}tr\left(\frac{\partial \mathbf{A}}{\partial X_{ij}}\mathbf{A}^{-1}(z)\frac{\partial \mathbf{A}}{\partial X_{ij}}\mathbf{A}^{-2}(z)\right) - \frac{2}{p\sqrt{np}}\mathbf{e}_i^T\mathbf{T}^{1/2}\mathbf{A}^{-2}(z)\mathbf{T}^{1/2}\mathbf{e}_i$$

and

$$\begin{aligned}\frac{1}{p}\frac{\partial^3 tr\left(\mathbf{A}^{-1}(z)\right)}{\partial X_{ij}} &= \frac{8}{p\sqrt{np}}\mathbf{e}_i^T\mathbf{T}^{1/2}\mathbf{A}^{-2}(z)\frac{\partial \mathbf{A}}{\partial X_{ij}}\mathbf{A}^{-1}(z)\mathbf{T}^{1/2}\mathbf{e}_i \\ &\quad -\frac{6}{p}tr\left(\frac{\partial \mathbf{A}}{\partial X_{ij}}\mathbf{A}^{-1}(z)\frac{\partial \mathbf{A}}{\partial X_{ij}}\mathbf{A}^{-1}(z)\frac{\partial \mathbf{A}}{\partial X_{ij}}\mathbf{A}^{-2}(z)\right).\end{aligned} \tag{5.54}$$

Recalling the definition of $\mathbf{s}_j$ given in the introduction, we have

$$\mathbf{e}_j^T\mathbf{X}^T\mathbf{T}^{1/2} = \mathbf{s}_j^T\mathbf{T}^{1/2} = \mathbf{r}_j^T.$$

Let $\hat{\mathbf{e}}_i = \mathbf{T}^{1/2}\mathbf{e}_i$. Then, using (5.53) and (5.9), we further write

$$\frac{1}{p\sqrt{np}}\hat{\mathbf{e}}_i^T\mathbf{A}^{-2}(z)\frac{\partial \mathbf{A}}{\partial X_{ij}}\mathbf{A}^{-1}(z)\hat{\mathbf{e}}_i = c_{n1} + c_{n2} + c_{n3}, \tag{5.55}$$

where

$$\begin{aligned} c_{n1} &= \frac{1}{np^2}\hat{\mathbf{e}}_i^T\mathbf{A}^{-2}(z)\hat{\mathbf{e}}_i\frac{\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\hat{\mathbf{e}}_i}{1+\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j}, \\ c_{n2} &= \frac{1}{np^2}\frac{\hat{\mathbf{e}}_i^T\mathbf{A}_j^{-2}(z)\mathbf{r}_j}{1+\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j}\hat{\mathbf{e}}_i^T\mathbf{A}^{-1}(z)\hat{\mathbf{e}}_i, \\ c_{n3} &= -\frac{1}{n^{3/2}p^{5/2}}\frac{\hat{\mathbf{e}}_i^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^T\mathbf{A}_j^{-2}(z)\mathbf{r}_j}{(1+\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j)^2}\hat{\mathbf{e}}_i^T\mathbf{A}^{-1}(z)\hat{\mathbf{e}}_i, \end{aligned}$$

where the definition of $\mathbf{A}_j^{-1}(z)$ is given in the subsection 2.1, equation (5.8) is also used to obtain c_{n2} and c_{n3}, and define $(\mathbf{A}_j - z\mathbf{I})^{-2}$ by $\mathbf{A}_j^{-2}(z)$.

We then claim that

$$\frac{\|\mathbf{A}_j^{-1}(z)\mathbf{r}_j\|/(np)^{1/4}}{\left|1+\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j\right|} \le \frac{M}{v}\frac{\|\mathbf{r}_j\|}{(np)^{1/4}}. \tag{5.56}$$

To this end, we need a result which states the relationship between the real part and the imaginary part of the Stieltjes transform, say $m(z)$, of any probability distribution function:

$$|Re(m(z))| \le v^{-1/2}\sqrt{Im(m(z))}, \tag{5.57}$$

whose proof is straightforward or one may refer to Theorem 12.8 in[3].

Note that $\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j/\frac{\|\mathbf{r}_j\|^2}{\sqrt{np}}$ can be viewed as the Stieltjes transform of a probability distribution function. It follows from (5.57) that

$$\left|Re\Big(\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j\Big)\right| \le \frac{\|\mathbf{r}_j\|}{\sqrt{v}(np)^{1/4}}\sqrt{Im\Big(\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j\Big)}.$$

Therefore, it follows

$$\left|1+\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j\right| \ge 1-\left|Re(\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j)\right| \ge 2/3,$$

if $\frac{\|\mathbf{r}_j\|}{\sqrt{v}(np)^{1/4}}\sqrt{Im\Big(\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j\Big)} \le \frac{1}{3}$.

This implies

$$\frac{\|\mathbf{A}_j^{-1}(z)\mathbf{r}_j\|/(np)^{1/4}}{\left|1+\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j\right|} \le \frac{3\|\mathbf{r}_j\|}{2v(np)^{1/4}}.$$

If $\frac{\|\mathbf{r}_j\|}{\sqrt{v}(np)^{1/4}}\sqrt{Im\Big(\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j\Big)} > \frac{1}{3}$, then

$$\frac{\|\mathbf{A}_j^{-1}(z)\mathbf{r}_j\|/(np)^{1/4}}{\left|1+\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j\right|} \le \frac{1}{\sqrt{vIm(\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j)}} \le \frac{3\|\mathbf{r}_j\|}{v(np)^{1/4}},$$

which completes the proof of (5.56).

Applying (5.56) gives

$$E\left[\left|X_{ij}^3c_{n1}\right|\right] \le \frac{M}{np^2}E\left[\left|X_{ij}^3\|\mathbf{r}_j\|\right|\right] \le \frac{M}{np^{3/2}}(E\left[X_{11}^4\right]+\Big(E\left[X_{11}^2\right])^{1/2}\Big). \tag{5.58}$$

Similarly

$$E\left[\left|X_{ij}^3 c_{n2}\right|\right] \le \frac{M}{np^{3/2}} \quad \text{and} \quad E\left[\left|X_{ij}^3 c_{n3}\right|\right] \le \frac{M}{np^{3/2}}, \tag{5.59}$$

because, as in (5.11),

$$\left|\frac{\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-2}(z)\mathbf{r}_j}{1+\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j}\right| \le \frac{1}{v}. \tag{5.60}$$

Consider the second term in (5.54), which, by (5.9) and (5.53), equals to

$$\begin{aligned} &\frac{1}{n^{3/2}p^{5/2}} tr\Big[(\hat{\mathbf{e}}_i\mathbf{r}_j^T + \mathbf{r}_j\hat{\mathbf{e}}_i^T)\mathbf{A}^{-1}(z)(\hat{\mathbf{e}}_i\mathbf{r}_j^T + \mathbf{r}_j\hat{\mathbf{e}}_i^T)\mathbf{A}^{-1}(z)(\hat{\mathbf{e}}_i\mathbf{r}_j^T + \mathbf{r}_j\hat{\mathbf{e}}_i^T)\mathbf{A}^{-2}(z)\Big] \\ &= \ 2d_{n1} + 2d_{n2} + 2d_{n3} + 2d_{n4}, \end{aligned} \tag{5.61}$$

where

$$\begin{aligned} d_{n1} &= \frac{1}{n^{3/2}p^{5/2}}\left(\frac{\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\hat{\mathbf{e}}_i}{1+\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j}\right)^2 \mathbf{r}_j^T\mathbf{A}^{-2}(z)\hat{\mathbf{e}}_i, \\ d_{n2} &= \frac{1}{n^{3/2}p^{5/2}}\frac{\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\hat{\mathbf{e}}_i}{\left(1+\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j\right)^2}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j\hat{\mathbf{e}}_i^T\mathbf{A}^{-2}(z)\hat{\mathbf{e}}_i, \\ d_{n3} &= \frac{1}{n^{3/2}p^{5/2}}\frac{\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j}{\left(1+\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j\right)}\ \hat{\mathbf{e}}_i^T\mathbf{A}^{-1}(z)\hat{\mathbf{e}}_i\mathbf{r}_j^T\mathbf{A}^{-2}(z)\hat{\mathbf{e}}_i, \\ d_{n4} &= \frac{1}{n^{3/2}p^{5/2}}\hat{\mathbf{e}}_i^T\mathbf{A}^{-1}(z)\hat{\mathbf{e}}_i\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\hat{\mathbf{e}}_i\ \frac{\mathbf{r}_j^T\mathbf{A}_j^{-2}(z)\mathbf{r}_j}{\left(1+\frac{1}{\sqrt{np}}\mathbf{r}_j^T\mathbf{A}_j^{-1}(z)\mathbf{r}_j\right)^3}. \end{aligned}$$

By (5.56) and recalling that $|X_{ij}| \le n^{1/4}\varepsilon_p$, we have

$$\begin{aligned} &E\left[\left|X_{ij}^3 d_{n1}\right|\right] \le \frac{M}{n^{3/2}p^{5/2}}E\left[\left|X_{ij}^3\|\mathbf{r}_j^T\|^3\right|\right] \\ &\le \ \frac{M}{n^{3/2}p^{5/2}}\left(E\left[|X_{ij}|^6\right] + E\left[\left|X_{ij}^3\right|\right]\ E\left[\left|\sum_{k\neq i}^p X_{kj}^2\right|^{3/2}\right]\right) \\ &\le \ \frac{M}{np^{5/2}} + \frac{M}{n^{3/2}p} \le \frac{M}{np^{3/2}}. \end{aligned}$$

Obviously, this argument also gives for $k = 2, 3, 4$

$$E\left[\left|X_{ij}^3 d_{nk}\right|\right] \le \frac{M}{n^{3/2}p^{5/2}}E\left[\left|X_{ij}^3\|\mathbf{r}_j^T\|^3\right|\right] \le \frac{M}{np^{3/2}}.$$

Summarizing the above, we have proved that

$$E\left[\left|X_{ij}^3 \frac{1}{p} \frac{\partial^3 tr\left(\mathbf{A}^{-1}(z)\right)}{\partial X_{ij}}\right|\right] \leq \frac{M}{np^{3/2}}. \tag{5.62}$$

Moreover, in the derivation above, we only use the facts that X_{ij} are independent with mean zero and finite fourth moment. In the meantime, note that X_{ij} and W_{ij} play the same role in their corresponding matrices. Additionally, all these random variables are independent with mean zero and finite fourth moment. Therefore, the above argument apparently works for all matrices.

We finally conclude from (5.62) and (5.51) that

$$\begin{aligned}
&\left|E\left[\frac{1}{p} tr\left(\mathbf{A}^{-1}(z)\right)\right] - E\left[\frac{1}{p} tr\left((\mathbf{D} - z\mathbf{I})^{-1}\right)\right]\right| \\
&\leq M \sum_{j=1}^{pn} \left[\int_0^1 (1-t)^2 \left| E\left(\hat{X}_j^3 \partial_j^3 f\left(\mathbf{Z}_j^{(1)}(t)\right)\right)\right| dt \right. \\
&\left. + \int_0^1 (1-t)^2 E\left|\left(\hat{W}_j^3 \partial_j^3 f\left(\mathbf{Z}_{j-1}^{(2)}(t)\right)\right)\right| dt\right] \leq \frac{M}{\sqrt{p}}.
\end{aligned}$$

Therefore, the proof of Theorem 1 is completed.

References

[1] Bai, Z. D., Miao, B. Q. & Jin, B. S. (2007). On limit theomrem of the eigenvalues of product of two random matrices. *J. Multivariate Anal.*, **98**, 76-101.

[2] Bai, Z. D. and Silverstein, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large dimensional random matrices. *Ann. Probab.* **26**, 316-345.

[3] Bai, Z. D. and Silverstein, J. W. (2006). *Spectral Analysis of Large Dimensional Random Matrices*, Mathematics Monograph Series 2, Science Press, Beijing.

[4] Bai, Z. D. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statistica Sinica* **18**, 425–442.

[5] Bai, Z. D. and Yin, Y. Q. (1986). A convergence to the semicircle law. *Ann. Probab.* **2**, 863-875.

[6] Burkholder, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probab.*, **1**, 19-42.

[7] Chatterrjee, S. (2006). A generazation of Lindeberg principle. *Ann. Probab.* **6**, 2061-2076.

[8] Chen, J., Gao, J. and Li, D. (2009). A new diagnostic test for cross–sectional uncorrelatedness in nonparametric panel data models. Working paper available at http://www.adelaide.edu.au/directory/jiti.gao

[9] Hsiao, C., Pesaran, M. H. and Pick, A. (2007). Diagnostic tests of cross section independence for nonlinear panel data models. *IZA discussion paper* No. 2756.

[10] Jiang. T. F. (2004). The asymptotic distributions of largest entries of sample correlation matrices. *Ann. Applied Prob.* **14**, 865-880.

[11] Jonsson, D. (1982). Some limit theorems of the eigenvalues of sample convariance matrix, *J. Multivariate Anal.* **12**, 1-38.

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

[13] Lindeberg, J. W. (1922). Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung. *Math. Z.*, **15**, 211–225.

[14] Liu, W. D., Lin, Z. Y. and Shao, Q. M. (2008). The asymptotic distribution and Berry–Esseen bound for a new test for independence in high dimension with an application to stochastic optimization. *Ann. Applied Prob.* **6**, 2337-2366.

[15] Marčenkon V. A. and Pastur, L. A. (1967). Distribution for some sets of random matrices, *Math. USSR-Sb.* **1**, 457-483.

[16] Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. Cambridge Working Paper in Economics No. 0435.

[17] Silverstein, J. W. (1995). Strong convergence of the limiting distribution of the eigenvalues of large dimensional random matrices. *J. Multivariate Anal.*, **55**, 331-339.

[18] Silverstein, J. W. and Bai, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.*, **55**, 331-339.

[19] Wigner, E. P. (1958). On the distributions of the roots fo certain symmetric matrices, *Ann. Math.* **67**, 325-327.

[20] Zhou, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.* **359**, 5345-5363.