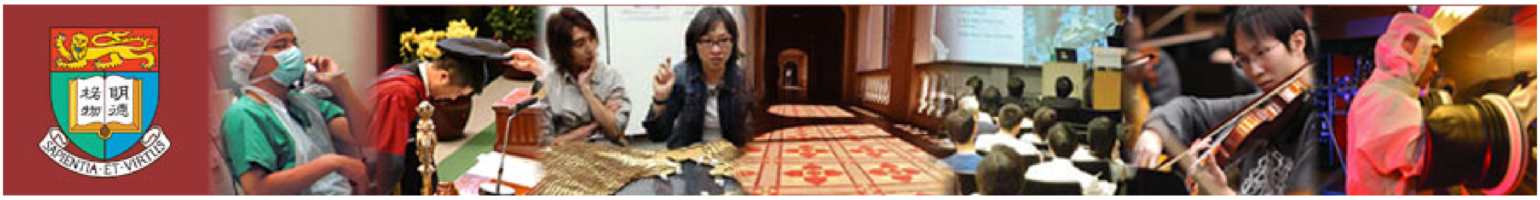
View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by HKU Scholars Hub

The HKU Scholars Hub The University of Hong Kong 香港大學學術庫



Title	**CLT for linear spectral statistics of a rescaled sample precision matrix**
Author(s)	**Zheng, S; Bai, Z; Yao, JJ**
Citation	**Random Matrices: Theory and Applications, 2015, v. 4, article no. 1550014**
Issued Date	**2015**
URL	**http://hdl.handle.net/10722/231310**
Rights	**Electronic version of an article published as [Random Matrices: Theory and Applications, 2015, v. 4, article no. 1550014] [10.1142/S2010326315500148] © [copyright World Scientific Publishing Company] [http://www.worldscinet.com/rmta]; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.**

Random Matrices: Theory and Applications
© World Scientific Publishing Company

CLT for linear spectral statistics of a rescaled sample precision matrix

Shurong Zheng
School of Mathematics and Statistics and KLAS, Northeast Normal University, P. R. China
zhengsr@nenu.edu.cn

Zhidong Bai
School of Mathematics and Statistics and KLAS, Northeast Normal University, P. R. China
baizd@nenu.edu.cn

Jianfeng Yao
Department of Statistics and Actuarial Science, The University of Hong Kong, P. R. China
jeffyao@hku.hk

Received (-)

Central limit theorems (CLT) of linear spectral statistics of general Fisher matrices $\mathbf{F}$ are widely used in multivariate statistical analysis where $\mathbf{F} = \mathbf{S}_y\mathbf{M}\mathbf{S}_x^{-1}\mathbf{M}^*$ with a deterministic complex matrix $\mathbf{M}$ and two sample covariance matrices $\mathbf{S}_x$ and $\mathbf{S}_y$ from two independent samples with sample sizes m and n. As the first step to obtain the CLT, it is necessary to establish the CLT for linear spectral statistics of the random matrix $\mathbf{M}\mathbf{S}_x^{-1}\mathbf{M}^*$, or equivalently that of $\mathbf{S}_x^{-1}\mathbf{T}$, that is a sample precision matrix rescaled by a general non-negative definite Hermitian matrix $\mathbf{T} = \mathbf{M}^*\mathbf{M}$. Because the scaling matrix $\mathbf{T}$ in many large-dimensional problems may not be invertible, the result does not simply follow from the celebrated CLT by Bai and Silverstein (2004). Thus we have to alternatively derive the CLT of linear spectral statistics of $\mathbf{S}_x^{-1}\mathbf{T}$ where the inverse of $\mathbf{T}$ may not exist, thus extending Bai and Silverstein's CLT. As a further innovation of the paper, general populations for the sample covariance matrix $\mathbf{S}_x$ are covered requiring the existence a fourth order moment of arbitrary value, that is not necessarily matching the values of the the Gaussian case.

Keywords: linear spectral statistics, central limit theorem, precision matrix, rescaled precision matrix, Fisher matrix

Mathematics Subject Classification 2000: 15B52, 60F05, 60B20

1. Introduction

For a $p \times p$ random matrix $\mathbf{A}_n$ with eigenvalues $\{\lambda_j\}_{j=1}^p$, the linear spectral statistics (LSS) defined by $p^{-1}\sum_{j=1}^p f(\lambda_j)$ for the function f are of central importance in the theory of random matrices and its applications. Two main questions arise for such LSS: (1). determine their point limit in term of a *limiting spectral distribution* (LSD), say G, such that they converge to $G(f) = \int f(x)dG(x)$ (in an appropriate sense and for an appropriate family of functions f); (2). characterise the fluctuations

$p^{-1}\sum_{j=1}^{p} f(\lambda_j) - G(f)$ in term of an appropriate central limit theorem (CLT). Both questions on such LSS of large dimensional random matrices have a long history, and received considerable attention in recent years. They have important applications in various domains such as the number theory, high-dimensional multivariate statistics, wireless communication networks and signal processing, e.g. ; for more information, the readers are referred to the recent survey paper [14], [11] and [25]. The rescaled precision matrices considered in this paper belong to a particular class of random matrices involving products of two independent matrices, and the early works [29] and [30] about Voiculescu's S-transform provide tools based on free probability for the study of the LSD of such products. Other references on LSD of product of random matrices include [1], [12], [20], [21] and [26]. Concerning the so-called Fisher matrices (see below for examples), their LSD are derived in [27] and [8]. As for the second question on fluctuations, the CLT for $(\mathrm{tr}(\mathbf{A}_n), \cdots, \mathrm{tr}(\mathbf{A}_n^k))$ is established in [15] for a sequence of Wishart matrices $(\mathbf{A}_n)$, where k is a fixed number, and the dimension p of the matrices grows proportionally to the sample size n. Subsequent works [10] and [16] considered the extensions of the classical Gaussian ensembles to the non-Gaussian ensembles, and [28] considered the Gaussian fluctuations for the LSS of Wigner matrices with a class of more general test functions. A general CLT for LSS of Wigner matrices was given in [7] where in particular, the limiting mean and covariance functions are specified. Similarly, [6] established the CLT for general sample covariance matrices with explicit limiting parameters. In [19], the authors use characteristic functions rather than the Stieltjes transform to obtain the limiting spectral distributions for Wishart type matrics. In [18], the authors reconsider such CLTs but with a new idea of interpolation that allows the generalisations from Gaussian matrix ensembles to matrix ensembles with general entries satisfying a moment condition (notice however this approach does not provide information on the centring parameters in the CLTs and can only characterise the asymptotic covariance function). Recent improvements are proposed in [22], [23], [24] and [33] that propose a generalisation of the CLT in [6].

The main purpose of this paper is to establish a CLT for LSS of a *rescaled sample precision matrix* of form $\mathbf{S}^{-1}\mathbf{T}$, where $\mathbf{S}$ is a sample covariance matrix with the sample size n from a p-variate population with independent components, and $\mathbf{T}$ is a $p\times p$ non-random non-negative definite scaling matrix. The research is motivated by considering the CLT of LSS for general Fisher matrices that are widely-used in multivariate statistical analysis. The asymptotic distributions of several meaningful test statistics depend on the related Fisher matrices For example, in testing the proportionality of two covariance matrices $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ in [17], $H_0 : \mathbf{\Sigma}_x = c\mathbf{\Sigma}_y$ where c is unknown, the commonly used test statistic is $\mathrm{tr}(\hat{\mathbf{\Sigma}}_x^{-1}\hat{\mathbf{\Sigma}}_y)^2/[\mathrm{tr}(\hat{\mathbf{\Sigma}}_x^{-1}\hat{\mathbf{\Sigma}}_y)]^2$ where $\hat{\mathbf{\Sigma}}_x$ and $\hat{\mathbf{\Sigma}}_y$ are two independent sample covariance matrices. More examples in multivariate analysis involving a Fisher matrix are can be found in [4], [5], [13] and [17]. Moreover, the Fisher matrix $\hat{\mathbf{\Sigma}}_x^{-1}\hat{\mathbf{\Sigma}}_y$ can be rewritten as $\mathbf{F} = \mathbf{S}_y\mathbf{M}\mathbf{S}_x^{-1}\mathbf{M}^*$ where $\mathbf{M} = \mathbf{\Sigma}_y^{1/2}\mathbf{\Sigma}_x^{-1/2}$ is a $p\times p$ deterministic matrix, $*$ denotes the conjugate and transpose, and $\mathbf{S}_x$ and $\mathbf{S}_y$ are two $p \times p$ sample covariance matrices of two independent samples which are assumed to have independent components with zero mean vectors and identity covariance matrices. Notice that here we denote the degrees of freedom of $\mathbf{S}_x$ and $\mathbf{S}_y$ by n_1 and n_2, respectively. In the large-dimensional context, [32] establishes a CLT for the LSS of a standard Fisher matrix where the two population covariance matrices are equal, i.e. the matrix $\mathbf{M} \geq \mathbf{0}$ is an identity matrix and $\mathbf{F} = \mathbf{S}_y \mathbf{S}_x^{-1}$. It is however of significant importance to obtain the CLT for the general Fisher matrices $\mathbf{F}$ with an arbitrary $\mathbf{M}$ matrix, especially when the powers of the tests are concerned with, e. g., for the test of equality of two population covariance matrices, the null distributions of test statistics rely on a standard Fisher matrix with $\mathbf{M} = \mathbf{I}_p$ while those under the alternative hypothesis depend on the general Fisher matrix with an arbitrary $\mathbf{M}$.

In order to establish the CLT of the LSS of the general Fisher matrices $\mathbf{F} = \mathbf{S}_y \mathbf{M} \mathbf{S}_x^{-1} \mathbf{M}^*$ and following the approach of [32], one proceeds in two steps. First, conditional on $\mathbf{S}_x$ and applying [6], one establishes a (conditional) CLT for $\sum_{j=1}^{p} f(\lambda_j)$. In this step, we need to determine the LSD of the conditioning factor $\mathbf{M}\mathbf{S}_x^{-1}\mathbf{M}^*$, or equivalently the scaled precision matrix $\mathbf{S}_x^{-1}\mathbf{M}^*\mathbf{M}$. In this step, it is necessary to find the LSD of $\mathbf{S}_x^{-1}\mathbf{M}^*\mathbf{M}$ as the centring parameter as well as limiting mean and covariance functions for the conditional CLT all depend on complex contours integrals involving this LSD. The second step consists on an un-conditioning calculation that leads to the final CLT for LSS of a general Fisher matrix. In this step, one needs a CLT for LSS of the scaled precision matrix $\mathbf{S}_x^{-1}\mathbf{M}^*\mathbf{M}$. Details on the implementation of this approach is reported elsewhere, see [34]. In this paper, we will only focus on the research of the scaled precision matrix $\mathbf{S}_x^{-1}\mathbf{M}^*\mathbf{M}$.

Despite the close connection between scaled precision matrices and general Fisher matrices described so far, limiting theorems for scaled precision matrix of type $\mathbf{S}^{-1}\mathbf{T}$ have their own interests. Specifically, when $\mathbf{T}$ is invertible, since $\mathbf{S}^{-1}\mathbf{T} = \left[\mathbf{T}^{-1}\mathbf{S}\right]^{-1}$, the CLT for LSS of $\mathbf{S}^{-1}\mathbf{T}$ can be derived from the CLT of [6]. In this case, the LSD of $\mathbf{S}^{-1}\mathbf{T}$ can also be easily derived from the existing literature on general sample covariance matrices. However, the scaling matrix $\mathbf{T}$ is usually not invertible or has eigenvalues close to zero, and in this case both questions on the LSD and CLT for associated LSS cannot be answered using the existing results such as [6].

In summary, this paper treats a general scaled precision matrix where the non-negative definite scaling matrix is allowed to be singular. The main contributions are the establishment of a LSD for the matrix and a CLT for the family of its linear spectral statistics. Particularly and it will be seen, the establishment of the CLT is challenging and needs the implementation of new methods of proof. Compared to the CLT in [6], we show in Section 2.1 that the popular CLT can now be considered as a special case of our new CLT. As a further innovation of the paper, general populations for the sample covariance matrix $\mathbf{S}$ are covered requiring the existence a fourth order moment of *arbitrary* value, that do not necessarily match the values of the the Gaussian case, i.e. the value 3 for real variables and the value 2 for the complex variables. As it is known in the literature on CLTs for LSS of random matrices, allowing arbitrary values for the fourth order moment of the matrix entries is quite delicate and requires supplementary effort.

The organisation of this paper is as follows. Section 2 presents our main results. Theorem 2.1 establishes the limiting spectral distribution (LSD) of the matrix $\mathbf{S}_x^{-1}\mathbf{T}$ under appropriate moment conditions. CLTs on LSS of the matrix are developed in two steps. First, our main Theorem 2.2 provide a CLT for linear spectral statistics while requiring that the independent components in $\mathbf{S}$ have zero mean, unit variance and a fourth moment matching the Gaussian case. Next in Proposition 2.1, this matching condition is removed under some tricky technical conditions. This is not surprising since we know that the fourth moment matching condition cannot be removed "for free", see e.g. discussions in [33]. Such technical conditions can be simplified in the case of *diagonal* scaling matrix $\mathbf{T}$ and this is done in Proposition 2.2. Next in Section 2.1, we explain the fact that when the matrix $\mathbf{T}$ is indeed invertible, Theorem 2.2 is equivalent to the well-known CLT of [6]. It is also in this sense that Theorems 2.2 proposes a valuable extension of [6]'s result.

The next three sections provide proofs for Theorem 2.1, Theorem 2.2 and the two Propositions 2.1 and 2.2, respectively. The last section collects some lemmas use in these proofs.

2. Main results

Following [6], let $\{\mathbf{x}_t\}$, $t = 1, \ldots, n$ be a sequence of independent p-dimensional observations with independent and standardised components, i.e. for $\mathbf{x}_t = (x_{tj})$, $\mathrm{E}x_{tj} = 0$ and $\mathrm{E}|x_{tj}|^2 = 1$. The corresponding sample covariance matrix is

$$\mathbf{S} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t\mathbf{x}_t^* \ . \tag{2.1}$$

The inverse $\mathbf{S}^{-1}$ is the sample precision matrix and we consider its scaled product

$$\mathbf{S}^{-1}\mathbf{T} = \left(\frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t\mathbf{x}_t^*\right)^{-1} \mathbf{T} \ , \tag{2.2}$$

where $\mathbf{T}$ is a $p \times p$ *scaling matrix* which is deterministic and non-negative definite Hermitian matrix. Notice that we do not require $\mathbf{T}$ be invertible. When $\mathbf{T} = \mathbf{T}^*$ and $\mathbf{S} = \mathbf{S}^*$ is invertible, the eigenvalues of $\mathbf{S}^{-1}\mathbf{T}$ are all real.

We first specify the framework for our main results. Recall that the *empirical spectral distribution* (ESD) of a complex-value $p \times p$ matrix $\mathbf{A}$ is the measure $\mu_{\mathbf{A}} = p^{-1}\sum_{j=1}^{p}\delta_{\lambda_j}$ where $\{\lambda_j\}$ are the eigenvalues of $\mathbf{A}$ and δ_a denotes the Dirac mass at a point a. When the sequence of ESD $\{\mu_A\}$ has a limit μ, μ is called the limiting spectral distribution (LSD) of the sequence of matrices $\{\mathbf{A}\}$.

Assumption 1 The $p \times n$ observation matrix $(x_{tj}, t = 1, \cdots, n, j = 1, \cdots, p)$ are made with independent elements satisfying $\mathrm{E}x_{tj} = 0$, $\mathrm{E}|x_{tj}|^2 = 1$.

Moreover, for any $\eta > 0$ and as $p,\ n \to \infty$,

$$\frac{1}{np}\sum_{t=1}^{n}\sum_{j=1}^{p} \mathrm{E}\left[|x_{tj}|^2 I_{\{|x_{tj}|\geq \eta\sqrt{n}\}}\right] \to 0\ , \tag{2.3}$$

where $I_{\{\cdot\}}$ is the indicator function.

The elements are either all real or all complex and we set an index $\kappa = 1$ or $\kappa = 2$, respectively. In the later case, $\mathrm{E}\{x_{tj}^2\} = 0$ for all t, j in the complex case.

Assumption 1* In addition to Assumption 1, the entries $\{x_{tj}\}$ have a *uniform* 4-th moment $\mathrm{E}|x_{tj}|^4 = 1 + \kappa$. Moreover, for any $\eta > 0$ and as $p,\ n \to \infty$,

$$\frac{1}{np}\sum_{t=1}^{n}\sum_{j=1}^{p} \mathrm{E}\left[|x_{tj}|^4 I_{\{|x_{tj}|\geq \eta\sqrt{n}\}}\right] \to 0. \tag{2.4}$$

Assumption 1** In addition to Assumption 1, the entries $\{x_{tj}\}$ have a finite 4-th moment *which are not necessarily the same*. Moreover, for any $\eta > 0$ and as $p,\ n \to \infty$,

$$\frac{1}{np}\sum_{t=1}^{n}\sum_{j=1}^{p} \mathrm{E}\left[|x_{tj}|^4 I_{\{|x_{tj}|\geq \eta\sqrt{n}\}}\right] \to 0. \tag{2.5}$$

Assumption 2 As $p \to \infty$, the ESD H_n of $\{\mathbf{T}\}$ tends to a limit H, which is a deterministic probability measure and $H \neq \delta_0$ (Dirac mass at 0).

Assumption 2* In addition to Assumption 2, the sequence $\{\mathbf{T}\}$ is bounded in spectral norm.

Assumption 3 The dimension p and the sample size n both tend to infinity such that $y_n = p/n \to y \in (0, 1)$.

Assumption 1 states that the entries are independent, not necessarily identically distributed, but with homogeneous moments of first and second order, together with a Lindeberg type condition of order 2. Assumption 1* reinforce Assumption 1 with similar conditions using a homogeneous fourth order moment that matches the Gaussian case. Assumption 1** generalises the previous one by allowing arbitrary values for the fourth moment of the entries.

First we identify the LSD of $\mathbf{S}^{-1}\mathbf{T}$.

Theorem 2.1. *Under Assumptions 1, 2 and 3, with probability 1, the ESD F_n of $\mathbf{S}^{-1}\mathbf{T}$ tends to a non-random distribution $F^{y,H}$ whose Stieltjes transform $s(z) = \int (x-z)^{-1} dF^{y,H}(x)$ is the unique solution to the equation*

$$zs(z) = -1 + \int \frac{t dH(t)}{1 - yz^2 s(z) + t} \tag{2.6}$$

subject to the condition that $s(z)$ has the same sign of imaginary part of z, where z belongs to the set of complex numbers outside the real axis. The distribution $F^{y,H}$ is then the LSD of $\mathbf{S}^{-1}\mathbf{T}$.

Notice that from (2.6), it is easy to obtain $zs(z) \neq 1$. Next, we consider the linear spectral statistics of $\mathbf{S}^{-1}\mathbf{T}$ of form

$$F_n(f) = \int f(x) dF_n(x) = \frac{1}{p} \sum_{j=1}^{p} f(\lambda_j) \ ,$$

where the $\{\lambda_j\}$'s are the eigenvalues of the matrix $\mathbf{S}^{-1}\mathbf{T}$ and f a given test function. Similarly to [6], a special feature here is that fluctuations of $F_n(f)$ will not be considered around the LSD limit $F^{y,H}(f)$, but around $F^{y_n,H_n}(f)$, a finite-sample proxy of $F^{y,H}$ obtained by substituting the parameters (y_n, H_n) for (y, H) in the LSD. Therefore, we consider the random variable

$$Z_n(f) = p\left[F_n(f) - F^{y_n,H_n}(f)\right] = p \int f(x) d[F_n - F^{y_n,H_n}](x) \ .$$

Throughout the paper, we use the following notations

$$\begin{aligned} b(z) &= 1 + yzs(z) \ , \\ g(z) &= 1 - y \int \frac{(1+yzs(z))^2 dH(t)}{\left(t/z - b(z)\right)^2} = 1 - yb^2(z) \int \frac{dH(t)}{\left(t/z - b(z)\right)^2} \ . \end{aligned}$$

We even write g, b for $g(z)$ and $b(z)$ if no ambiguity can arise. Notice that we have

$$\frac{1}{2}\frac{d}{dz} \log g(z) = \frac{1}{z^2} \frac{y \int \frac{tb^3 dH(t)}{(t/z-b)^3}}{\left(1 - y\int \frac{b^2 dH(t)}{(t/z-b)^2}\right)^2} = \frac{yb^3}{z^2 g^2} \int \frac{tdH(t)}{(t/z-b)^3} \ .$$

Define also the interval

$$I := \left[\frac{\liminf_p \lambda_{\min}(\mathbf{T})}{(1+\sqrt{y})^2}, \ \frac{\limsup_p \lambda_{\max}(\mathbf{T})}{(1-\sqrt{y})^2}\right] \ , \tag{2.7}$$

where $\lambda_{\min}(\mathbf{T})$ and $\lambda_{\max}(\mathbf{T})$ are respectively the smallest and the largest eigenvalue of $\mathbf{T}$.

The main result of the paper is the following CLT.

Theorem 2.2. *Assume that Assumptions 1*, 2* and 3 hold. Let $f_1, \cdots, f_k$ be functions analytic on an open domain $\mathcal{D}$ of the complex plane enclosing the interval I in* (2.7) *Then, the random vector $[Z_n(f_1), \cdots, Z_n(f_k)]$ weakly converges to a Gaussian vector $[Z_{f_1}, \cdots, Z_{f_k}]$ with mean function*

$$\mathrm{E}Z_{f_j} = -\frac{\kappa - 1}{2\pi i} \oint_{\mathcal{C}} f_j(z) \frac{yb^3(z)}{z^2 g^2(z)} \int \frac{tdH(t)}{(t/z - b(z))^3} dz \ , \tag{2.8}$$

and covariance function

$$\mathrm{Cov}(Z_{f_i}, Z_{f_j}) = -\frac{\kappa}{4\pi^2} \oint_{\mathcal{C}_2} \oint_{\mathcal{C}_1} \frac{f_i(z_1) f_j(z_2) \frac{\partial(z_1 b(z_1))}{\partial z_1} \frac{\partial(z_2 b(z_2))}{\partial z_2}}{[z_1 b(z_1) - z_2 b(z_2)]^2} dz_1 dz_2 \ . \tag{2.9}$$

The contours $\mathcal{C}$, $\mathcal{C}_1$ and $\mathcal{C}_2$ are closed, positively oriented and enclosing the interval I. Moreover, the contours $\mathcal{C}_1$ and $\mathcal{C}_2$ are non-overlapping.

As said in Introduction, this theorem essentially parallel the CLT in [6] and requires also the fourth moments of the independent variables in $\mathbf{S}$ match the value $\kappa+1$ as in the Gaussian case. We next introduce two variations of this main theorem. First in Proposition 2.1, this moment matching requirement is removed under some tricky technical conditions.

Proposition 2.1. *Assume that Assumptions 1**, 2* and 3 hold. Let $f_1, \cdots, f_k$ be functions analytic on an open domain $\mathcal{D}$ of the complex plane enclosing the interval I in (2.7) Moreover, assume that the following non-random limits exist:*

(1) $\frac{1}{p}\sum_{j=1}^{p}(\mathrm{E}|x_{ij}|^4 - 1 - \kappa)\left[(\mathbf{T}/z_1 - \mathbf{S}_i)^{-1}\right]_{jj}\left[(\mathbf{T}/z_2 - \mathbf{S}_i)^{-1}\right]_{jj}$ *converges to* $h(z_1, z_2)$ *uniformly in i;*

(2) $\frac{1}{p}\sum_{j=1}^{p}(\mathrm{E}|x_{1j}|^4-1-\kappa)\left[(\mathbf{T}/z - \mathbf{S}_1)^{-1}\right]_{jj}\left[(\mathbf{T}/z - \mathbf{S}_1)^{-1}\mathbf{T}\left(\mathbf{T}/z - \mathrm{E}\beta_1(z)\mathbf{I}\right)^{-1}\right]_{jj}$ *converges to* $h_M(z)$.

Then the random vector $[X_n(f_1), \cdots, X_n(f_k)]$ weakly converges to a Gaussian vector $[X_{f_1}, \cdots, X_{f_k}]$ with mean function

$$\mathrm{E}X_{f_j} = -\frac{\kappa-1}{2\pi i}\oint_{\mathcal{C}} f_j(z)\frac{yb^3(z)}{z^2g^2(z)}\int\frac{tdH(t)}{(t/z - b(z))^3}dz - \frac{1}{2\pi i}\oint_{\mathcal{C}} f_j(z)\frac{yb^3(z)}{z^2g(z)}h_M(z)dz,$$

and covariance function

$$\mathrm{Cov}(X_{f_i}, X_{f_j}) = -\frac{\kappa}{4\pi^2}\oint_{\mathcal{C}_2}\oint_{\mathcal{C}_1}\frac{f_i(z_1)f_j(z_2)\frac{\partial(z_1b(z_1))}{\partial z_1}\frac{\partial(z_2b(z_2))}{\partial z_2}}{[z_1b(z_1) - z_2b(z_2)]^2}dz_1dz_2 - \frac{1}{4\pi^2}\oint_{\mathcal{C}_2}\oint_{\mathcal{C}_1} f_i(z_1)f_j(z_2)\frac{\partial^2[yb(z_1)b(z_2)h(z_1,z_2)]}{\partial z_1\partial z_2}dz_1dz_2.$$

The contours $\mathcal{C}$, $\mathcal{C}_1$ and $\mathcal{C}_2$ are closed, positively oriented and enclosing the interval I. Moreover, the contours $\mathcal{C}_1$ and $\mathcal{C}_2$ are non-overlapping.

The appearance of the two tricky conditions (1)-(2) in Proposition 2.1 are surprising since it is known in the literature of CLTs for linear spectral statistics that the fourth moment matching condition cannot be removed without additional hypotheses, see e.g. discussions in [33]. Similar conditions have been also introduced in [23] in the case of a general sample covariance matrix. In Proposition 2.2 below, we consider a special case where the fourth moments are asymptotically identical to a value not necessarily matching the Gaussian case and the scaling matrix is diagonal, and the above tricky conditions can be worked out properly.

Proposition 2.2. *In addition to the assumptions in Proposition 2.1, assume that* $\mathrm{E}|x_{ij}|^4-1-\kappa=\beta_x+o(1)$ *uniformly in* i,j *and* $\mathbf{T}$ *is a diagonal matrix with positive eigenvalues. Then the random vector* $[X_n(f_1),\cdots,X_n(f_k)]$ *weakly converges to a Gaussian vector* $[X_{f_1},\cdots,X_{f_k}]$ *with mean function*

$$\begin{aligned}\mathrm{E}X_{f_j} &= -\frac{\kappa-1}{2\pi i}\oint_{\mathcal{C}} f_j(z)\frac{yb^3(z)}{z^2g^2(z)}\int\frac{tdH(t)}{(t/z-b(z))^3}dz,\\ &\quad -\frac{\beta_x}{2\pi i}\oint\left[f_j(z)\frac{yb^3(z)}{z^2g(z)}\int\frac{tdH(t)}{(t/z-b(z))^3}\right]dz\end{aligned}$$

and covariance function

$$\begin{aligned}&\mathrm{Cov}(X_{f_i},X_{f_j})\\ &= -\frac{\kappa}{4\pi^2}\oint_{\mathcal{C}_2}\oint_{\mathcal{C}_1}\frac{f_i(z_1)f_j(z_2)\frac{\partial(z_1b(z_1))}{\partial z_1}\frac{\partial(z_2b(z_2))}{\partial z_2}}{[z_1b(z_1)-z_2b(z_2)]^2}dz_1dz_2\\ &\quad -\frac{\beta_x}{4\pi^2}\oint\oint\left\{f_i(z_1)f_j(z_2)\frac{\partial^2}{\partial z_1\partial z_2}\left[y\int\frac{z_1b(z_1)z_2b(z_2)dH(t)}{[t-z_1b(z_1)][t-z_2b(z_2)]}\right]\right\}dz_1dz_2.\end{aligned}$$

The contours $\mathcal{C}$, $\mathcal{C}_1$ *and* $\mathcal{C}_2$ *are closed, positively oriented and enclosing the interval* I *(given in Proposition 2.1). Moreover, the contours* $\mathcal{C}_1$ *and* $\mathcal{C}_2$ *do not cross.*

2.1. *Theorem 2.2 is equivalent to the CLT in [6] for invertible* $\mathbf{T}$

As said in Introduction, Theorem 2.2 can also be viewed as a complement to the CLT in [6] while moving from the sample covariance matrix $\mathbf{S}$ to its inverse $\mathbf{S}^{-1}$. When the scaling matrix $\mathbf{T}$ is not invertible, these CLTs are not directly comparable. If $\mathbf{T}$ is indeed invertible, these CLT's become comparable; we now prove that they are indeed the same in this case. More precisely we prove that the mean and covariance functions given in Theorem 2.2 are the same as those given in Theorem 1.1 of [6].

Actually, when $\mathbf{T}$ is invertible, we have

$$\begin{aligned}s_n(z) &= \frac{1}{p}\mathrm{tr}(\mathbf{S}^{-1}\mathbf{T}-z\mathbf{I})^{-1} = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{\lambda_i((\mathbf{S}\mathbf{T}^{-1})^{-1})-z}\\ &= \frac{-1}{pz}\sum_{i=1}^{p}\frac{\lambda_i(\mathbf{S}\mathbf{T}^{-1})}{\lambda_i(\mathbf{S}\mathbf{T}^{-1})-z^{-1}}\\ &= -z^{-1}-\frac{1}{z^2}s_n^{\mathbf{S}\mathbf{T}^{-1}}(z^{-1})\\ &= -\frac{1}{yz}-\frac{1}{yz^2}\underline{m}_n(z^{-1})\ ,\end{aligned}$$

where $\lambda_i(\mathbf{ST}^{-1})$ is the ith eigenvalue of random matrix $\mathbf{ST}^{-1}$ and $\underline{m}_n$ is the Stieltjes transform of $\mathbf{X}_n\mathbf{T}^{-1}\mathbf{X}_n^*$ with $\mathbf{X}_n^* = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ is $p \times n$. That is,

$$s(z) = -\frac{1}{yz} - \frac{1}{yz^2}\underline{m}(z^{-1}), \quad b(z) = 1 - 1 - z^{-1}\underline{m}(z^{-1}) = -z^{-1}\underline{m}(z^{-1}) \ , \tag{2.10}$$

where $s(z)$ is the limit of $s_n(z)$ and $\underline{m}(z)$ is the limit of $\underline{m}_n(z)$. So the CLT of $p(s_n(z) - s(z))$ is the same as $-z^{-2}n(\underline{m}_n(z^{-1}) - \underline{m}(z^{-1}))$. By Lemma 1.1 of [6], we know that the process

$$\left\{-\frac{1}{z^2}n(\underline{m}_n(z^{-1}) - \underline{m}(z^{-1}))\right\}$$

converges weakly to a Gaussian process with the mean function

$$-\frac{1}{z^2}\frac{y\int\frac{t(\underline{m}(1/z))^3 d(H(t))}{(t+\underline{m}(1/z))^3}}{[1 - y\int\frac{(\underline{m}(1/z))^2 d(H(t))}{(t+\underline{m}(1/z))^2}]^2} \ ,$$

and covariance function

$$\frac{1}{z_1^2 z_2^2}\frac{\left.\frac{\partial\underline{m}(z)}{\partial z}\right|_{z=\frac{1}{z_1}}\left.\frac{\partial\underline{m}(z)}{\partial z}\right|_{z=\frac{1}{z_2}}}{(\underline{m}(\frac{1}{z_2}) - \underline{m}(\frac{1}{z_1}))^2} - \frac{1}{(z_1 - z_2)^2}$$

as $p \to \infty$. It is easily to verify that

$$\begin{aligned} &-\frac{1}{z^2}\frac{y\int\frac{t(\underline{m}(1/z))^3 d(H(t))}{(t+\underline{m}(1/z))^3}}{[1 - y\int\frac{(\underline{m}(1/z))^2 d(H(t))}{(t+\underline{m}(1/z))^2}]^2} \\ &= \frac{1}{z^2}\frac{y\int\frac{tz^3(1+ys(z))^3 dH(t)}{(t-z(b(z)))^3}}{[1 - y\int\frac{z^2(b(z))^2 dH(t)}{(t-z(b(z)))^2}]^2} = \frac{1}{z^2}\frac{y\int\frac{t(1+ys(z))^3 dH(t)}{(t/z-b(z))^3}}{[1 - y\int\frac{(b(z))^2 dH(t)}{(t/z-b(z))^2}]^2} \end{aligned}$$

and

$$\frac{1}{z_1^2 z_2^2}\frac{\left.\frac{\partial\underline{m}(z)}{\partial z}\right|_{z=\frac{1}{z_1}}\left.\frac{\partial\underline{m}(z)}{\partial z}\right|_{z=\frac{1}{z_2}}}{(\underline{m}(\frac{1}{z_2}) - \underline{m}(\frac{1}{z_1}))^2} - \frac{1}{(z_1 - z_2)^2} = \frac{\frac{\partial(z_1 b(z_1))}{\partial z_1}\frac{\partial(z_2 b(z_2))}{\partial z_2}}{[z_1 b(z_1) - z_2 b(z_2)]^2} - \frac{1}{(z_1 - z_2)^2} \ ,$$

which are the same as given in Theorem 2.2. This establishes the equivalence between Theorem 1.1 of [6] and Theorem 2.2 in this paper when the scaling matrix $\mathbf{T}$ is invertible.

3. Proof of Theorem 2.1

Using exactly the same approach employed in Section 4.3 of [2], we may truncate the extreme eigenvalues of $\mathbf{T}$ and tails of the random variables x_{ij} and then normalise them without altering the LSD of $\mathbf{S}^{-1}\mathbf{T}$. So we may assume that Assumption 2* is true and $|x_{ij}| \le \eta_n\sqrt{n}$ where $\eta_n \to 0$.

Now, we proceed with the proof of Theorem 2.1. To start with, we assume that $\mathbf{T}$ is invertible and there is a positive constant $\omega > 0$ such that $H([0, \omega]) = 0$, that is, the norm of $\mathbf{T}^{-1}$ is bounded. By Theorem 4.1 of [2] we know that the LSD of $\mathbf{ST}^{-1}$ exists and its Stieltjes transform $m(z)$ satisfies

$$m(z) = \int \frac{1}{t(1-y-yzm(z))-z} dH(1/t) = \int \frac{tdH(t)}{1-y-yzm(z)-tz}. \tag{3.1}$$

Note that $m(z)$ is the unique solution to the equation (3.1) that has the same sign of imaginary part as z.

If we denote the Stieltjes transforms of the ESD of $\mathbf{S}^{-1}\mathbf{T}$ and $\mathbf{ST}^{-1}$ by $s_n(z)$ and $m_n(z)$, respectively. By the relation

$$m_n(z) = -\frac{1}{z} - \frac{1}{z^2} s_n(1/z),$$

and $m_n(z) \to m(z)$ a.s., we know that with probability 1, $s_n(z)$ converges to a limit $s(z)$ that satisfies

$$-\frac{1}{z} - \frac{1}{z^2} s(1/z) = \int \frac{tdH(t)}{1-y-yz(-\frac{1}{z}-\frac{1}{z^2}s(1/z))-tz}. \tag{3.2}$$

Changing z as $1/z$ and simplifying it, we obtain (2.6).

Now, we consider possibly singular $\mathbf{T}$ and will show that for any fixed $z = u+iv$ with $v>0$, $s_n(z)$ still converges to a limit $s(z)$ that satisfies (2.6).

For any fixed $\varepsilon > 0$, define $\mathbf{T}_\varepsilon = \mathbf{T} + \varepsilon\mathbf{I}$ and define $\mathbf{S}_+$ from $\mathbf{S}$ by replacing its eigenvalues less than $\frac{1}{2}a$ as $\frac{1}{2}a$, where $a = (1-\sqrt{y})^2$. By the rank inequality (see Theorem A.43 in [2]), we have

$$\left\| F^{\mathbf{S}^{-1}\mathbf{T}} - F^{\mathbf{S}_+^{-1}\mathbf{T}} \right\| \le \frac{1}{p} \# \left\{ \lambda_i(\mathbf{S}) \le \frac{1}{2}a \right\} \to 0, a.s. \tag{3.3}$$

By Theorem A.45 of [2],

$$L(F^{\mathbf{S}_+^{-1}\mathbf{T}}, F^{\mathbf{S}_+^{-1}\mathbf{T}_\varepsilon}) \le \|\mathbf{S}_+^{-1}(\mathbf{T}-\mathbf{T}_\varepsilon)\| \le 2a^{-1}\varepsilon \tag{3.4}$$

where L is the Levy distance between two distribution functions $F^{\mathbf{S}_+^{-1}\mathbf{T}}$ and $F^{\mathbf{S}_+^{-1}\mathbf{T}_\varepsilon}$. Using again the rank inequality, we have

$$\left\| F^{\mathbf{S}^{-1}\mathbf{T}_\varepsilon} - F^{\mathbf{S}_+^{-1}\mathbf{T}_\varepsilon} \right\| \le \frac{1}{p} \# \left\{ \lambda_i(\mathbf{S}) \le \frac{1}{2}a \right\} \to 0, a.s. \tag{3.5}$$

By what has been proved for invertible $\mathbf{T}$, with probability 1, $s_{n,\varepsilon}(z) = \frac{1}{p}\mathrm{tr}(\mathbf{S}^{-1}\mathbf{T}_\varepsilon) \to s_\varepsilon(z)$ which is a solution to the equation

$$zs_\varepsilon(z) = -1 + \int \frac{tdH_\varepsilon(t)}{1-yz^2s_\varepsilon(z)+t}. \tag{3.6}$$

where $H_\varepsilon(t) = H(t-\varepsilon)$.

To complete the proof of the theorem, we only need to verify that the equation (3.6) has a unique solution that is the Stieltjes transform of a probability measure, and the solution $s_\varepsilon(z)$ is right-continuous at $\varepsilon = 0$. Let $w_\varepsilon(z) = \sqrt{z}(1+zs_\varepsilon(z))$, where $\sqrt{z}$ is the square root of z satisfying $\Im(z)\Im(\sqrt{z}) > 0$, then the equation (3.6) becomes

$$w_\varepsilon(z) = \int \frac{tdH_\varepsilon(t)}{\frac{1+t}{\sqrt{z}} - (1-y)\sqrt{z} - yw_\varepsilon(z)}, \tag{3.7}$$

where $w_\varepsilon(z)$ has the same sign of imaginary part as z.

We only need to consider the case $\Im(z) > 0$. Let $w_2 = \Im(w_\varepsilon(z)) > 0$, comparing the imaginary parts of (3.7), we have

$$\begin{aligned} w_2 &= \int \frac{\frac{(1+t)\Im(\sqrt{z})}{|z|} + (1-y)\Im(\sqrt{z}) + yw_2}{\left|\frac{1+t}{\sqrt{z}} - (1-y)\sqrt{z} - yw_\varepsilon(z)\right|^2} tdH_\varepsilon(t) \\ &> \int \frac{yw_2}{\left|\frac{1+t}{\sqrt{z}} - (1-y)\sqrt{z} - yw_\varepsilon(z)\right|^2} tdH_\varepsilon(t), \end{aligned}$$

which implies that

$$\int \frac{ytdH_\varepsilon(t)}{\left|\frac{1+t}{\sqrt{z}} - (1-y)\sqrt{z} - yw_\varepsilon(z)\right|^2} < 1. \tag{3.8}$$

Suppose (3.7) had two solution $w^{(j)}$ with $w_2^{(j)} = \Im(w^{(j)}) > 0$, $j = 1, 2$. Then making difference of both sides and cancelling $w_1 - w_2$ from both sides, we obtain

$$1 = y \int \frac{tdH_\varepsilon(t)}{(\frac{1+t}{\sqrt{z}} - (1-y)\sqrt{z} - yw^{(1)})(\frac{1+t}{\sqrt{z}} - (1-y)\sqrt{z} - yw^{(2)})},$$

which implies by Cauchy-Schwarz that

$$1 \le \left(\int \frac{ytdH_\varepsilon(t)}{\left|\frac{1+t}{\sqrt{z}} - (1-y)\sqrt{z} - yw^{(1)}\right|^2} \int \frac{ytdH_\varepsilon(t)}{\left|\frac{1+t}{\sqrt{z}} - (1-y)\sqrt{z} - yw^{(2)}\right|^2} \right)^{1/2} < 1,$$

where the last inequality follows by applying (3.8) for both $w^{(1)}$ and $w^{(2)}$. The contradiction proves the uniqueness of a solution to (3.7).

Finally, we show that the solution w_ε is right-continuous at $\varepsilon = 0$. We have,

$$\begin{aligned} \left|\frac{1+t}{\sqrt{z}} - (1-y)\sqrt{z} - yw_\varepsilon(z)\right| &\ge \left| -\frac{1+t}{|z|}\Im(\sqrt{z}) - (1-y)\Im(\sqrt{z}) - y\Im(w_\varepsilon(z)) \right| \\ &> (1-y)\Im(\sqrt{z}), \end{aligned}$$

so that by (3.7), $\{w_\varepsilon(z)\}$ is bounded for small enough ε, say $\varepsilon \le c$. Let $\{w_{\varepsilon_k}(z)\}_{k\ge1}$ be a sequence converging to some limit w_* with $\varepsilon_k \to 0$. By continuity and passing to the limit in (3.7), we see that

$$w_* = \int \frac{tdH(t)}{\frac{1+t}{\sqrt{z}} - (1-y)\sqrt{z} - yw_*}.$$

By the just proved uniqueness property of solution of this equation, $w_* = w_0(z)$ and this is the only sequential limit of $\{w_\varepsilon(z)\}$ when $\varepsilon \to 0$. Since this set is bounded, any sequence $\{w_\varepsilon(z)\}$ with $\varepsilon \to 0$ must converge to $w_0(z)$. This implies that $s_\varepsilon(z) - s(z) \to 0$.

The proof of Theorem 2.1 is complete.

For the understanding of the proofs that will follow in next sections, we give some notes on various identities. First notice that

$$\tilde{m}(z) = \int \frac{1}{\frac{\lambda}{z} - \frac{1}{1-y\tilde{m}(z)}} dH(\lambda), \tag{3.9}$$

where $\tilde{m}(z)$ is the limit of $\frac{1}{p}\mathrm{tr}(z^{-1}\mathbf{T} - \mathbf{S})^{-1}$ and $H(t)$ is the LSD of $\mathbf{T}$. We have

$$\underline{\tilde{s}}_n^0(z) \to \underline{\tilde{s}}(z) = \frac{-z}{1 - y\tilde{m}(z)}, \quad b(z) = \frac{1}{1 - y\tilde{m}(z)} = \frac{1}{1 - y\int \frac{dH(t)}{t/z - b(z)}}, \tag{3.10}$$

where $s(z)$ is the Stieltjes transform of the LSD of $\mathbf{S}^{-1}\mathbf{T}$ and $\underline{\tilde{s}}_n^0(z) = \frac{1}{-z^{-1} + y\int \frac{1}{t + \underline{\tilde{s}}_n^0(z)} dH_n(t)}$ with the ESD $H_n(t)$ of $\mathbf{T}$. It also holds

$$\mathrm{E}\beta_1(z) \to \frac{1}{1 - y\tilde{m}(z)} = b(z), \quad b_i(z) \to \frac{1}{1 - y\tilde{m}(z)} = b(z), \tag{3.11}$$

where $\beta_1(z) = \frac{1}{1 - \boldsymbol{\alpha}_1^*(\frac{\mathbf{T}}{z} - \mathbf{S}_1)^{-1}\boldsymbol{\alpha}_1}$ and $b_1(z) = \frac{1}{1 - n^{-1}\mathrm{E}\mathrm{tr}(z^{-1}\mathbf{T} - \mathbf{S}_1)^{-1}}$. Furthermore,

$$-z(b(z)) = \frac{-z}{1 - y\int \frac{dH(t)}{t/z - (b(z))}} = \frac{-z}{1 - y\int \frac{z dH(t)}{t - z(b(z))}}. \tag{3.12}$$

$$-z^{-1} + y\int \frac{dH(t)}{t - z(b(z))} = \frac{1}{-zb(z)}, \tag{3.13}$$

$$\frac{1}{z^2} - y\int \frac{(-z(b(z)))' dH(t)}{(t - z(b(z)))^2} = \frac{-(-z(b(z))'}{(-z(b(z))^2}, \tag{3.14}$$

$$\frac{(z(b(z))^2}{z^2} - y\int \frac{(z(b(z)))^2 dH(t)}{(t - z(b(z)))^2}(-z(b(z))' = -(-z(b(z))', \tag{3.15}$$

$$(-z(b(z))' = \frac{-1}{z^2} \frac{(-z(b(z))^2}{1 - y\int \frac{(-z(b(z)))^2 dH(t)}{(t - z(b(z)))^2}}. \tag{3.16}$$

Especially, when $\mathbf{T} = \mathbf{I}_p$, by (2.10), (3.10) and the definition of $\tilde{m}(z)$, we have

$$b(z) = \frac{1}{1 + ym(z^{-1})} = -z^{-1}\underline{m}(z^{-1}) = (1 - y) - \frac{y}{z}m(z^{-1}),$$

and

$$\tilde{m}(z) = -m(z^{-1}) = \frac{1}{z^{-1} - \frac{1}{1 + ym(z^{-1}))}}.$$

4. Some useful lemmas and identities

This section collects two lemmas that are used in the proofs developed in theorems.

Lemma 4.1. *Under Assumptions 1-2, we obtain that as $p/n \to y \in (0,1)$,*

$$\frac{1}{p}\mathrm{tr}(z^{-1}\mathbf{T} - \mathbf{S})^{-1} \to \tilde{m}(z), \ a.s. \tag{4.1}$$

where $\tilde{m}(z)$ is the unique solution to the equation $\tilde{m}(z) = \int \frac{dH(t)}{-\frac{t}{z} + \frac{1}{1-y\tilde{m}(z)}}$ satisfying

$$\Im(z)\Im(\tilde{m}(z)) \geq 0.$$

Proof. For any real $z < 0$ and complex w with $\Im(w) > 0$, by (4.1.2) of Page 61 of [2], we have

$$\begin{aligned}\frac{1}{p}\mathrm{tr}(z^{-1}\mathbf{T} - \mathbf{S} + w\mathbf{I})^{-1} &= -\frac{1}{p}\mathrm{tr}(\frac{1}{-z}\mathbf{T} + \mathbf{S} - w\mathbf{I})^{-1} \\ \to -\tilde{m}(z,w) &= -m_{H_z}\left(w - \frac{1}{y}\int \frac{\tau dH_0(\tau)}{1+\tau\tilde{m}(z,w)}\right), \ a.s. \end{aligned} \tag{4.2}$$

where $\tilde{m}(z,w)$ is limit of the Stieltjes transform of the matrix $\frac{1}{-z}\mathbf{T} + \mathbf{S}$, m_{H_z} is the Stieltjes transform of H_z, the LSD of $\frac{1}{-z}\mathbf{T}$, and $H_0(\tau) = I_{(\tau > y)}$. By Theorem 5.11 and Lemma 2.14 (Vitali Lemma) of [2], the convergence of (4.2) is also true for $w = 0$. That is,

$$\begin{aligned}\frac{1}{p}\mathrm{tr}(z^{-1}\mathbf{T} - \mathbf{S})^{-1} &\to -\tilde{m}(z,0) = -m_{H_z}\left(-\frac{1}{1+y\tilde{m}(z,0)}\right), \ a.s. \\ = -\int \frac{1}{\lambda - \frac{1}{1-y\tilde{m}(z,0)}} dH_z(\lambda) &= -\int \frac{1}{\frac{\lambda}{-z} + \frac{1}{1+y\tilde{m}(z,0)}} dH(\lambda), \ a.s. \end{aligned} \tag{4.3}$$

Denoting $\tilde{m}(z) = -\tilde{m}(z,0)$, then the convergence of (4.1) is proved for all real non-positive z. Noting that both sides of (4.1) are analytic functions of z on the region $D^- = \{z \in \mathbb{C} : z \text{ is not non-positive real number}\}$, applying Vitali Lemma again, we conclude that (4.1) is true for all $z \in D^-$ and $\tilde{m}(z)$ satisfies

$$\tilde{m}(z) = \int \frac{1}{\frac{\lambda}{z} - \frac{1}{1-y\tilde{m}(z)}} dH(\lambda) \tag{4.4}$$

Because the imaginary part of LHS of (4.4) has the same sign as z, we conclude that $\Im(\tilde{m}(z))$ should have the same sign as $\Im(z)$.

Our next goal is to show that for every non-real z, the equation (4.4) has a unique solution $\tilde{m}(z)$ whose imaginary part has the same sign as $\Im(z)$. By symmetry, we only need to consider the case where $\Im(z) > 0$. Suppose that there are two different solutions $m_1(z) \neq m_2(z)$. Making difference of both sides of (4.4), we obtain

$$1 = \int \frac{\frac{y}{(1-ym_1)(1-ym_2)}}{(\frac{\lambda}{z} - \frac{1}{1-ym_1})(\frac{\lambda}{z} - \frac{1}{1-ym_2})} dH(\lambda)$$

$$\leq \left(\int \frac{\frac{y}{|1-ym_1|^2}}{\left|\frac{\lambda}{z} - \frac{1}{1-ym_1}\right|^2} dH(\lambda) \int \frac{\frac{y}{|1-ym_2|^2}}{\left|\frac{\lambda}{z} - \frac{1}{1-ym_2}\right|^2} dH(\lambda) \right)^{1/2}. \tag{4.5}$$

Comparing the imaginary parts of both sides of (4.4), we have

$$\Im(m_j) = \int \frac{\frac{\Im(z)\lambda}{|z|^2} + \frac{y\Im(m_j)}{|1-ym_j|^2}}{\left|\frac{\lambda}{z} - \frac{1}{1-ym_j}\right|^2} dH(\lambda),\ j = 1, 2.$$

Since $\Im(m_j) > 0$ implies that

$$\int \frac{\frac{y}{|1-ym_j|^2}}{\left|\frac{\lambda}{z} - \frac{1}{1-ym_j}\right|^2} dH(\lambda) < 1,$$

which contradicts to (4.5). The proof of the lemma is complete. ■

Lemma 4.2. *Under Assumptions 1, 2, 3, we have*

$$s(z) = -z^{-1} - \frac{1}{z^2}\tilde{s}(z)$$

where

$$\underline{\tilde{s}}(z) = \frac{-z}{1 - y\tilde{m}(z)} = \frac{-z}{1 - y\int \frac{1}{t/z-(1-y\tilde{m}(z))^{-1}} dH(t)} = \frac{1}{-z^{-1} + y\int \frac{1}{t+\underline{\tilde{s}}(z)} dH(t)},$$

$\underline{\tilde{s}}(z) = -(1-y)z + y\tilde{s}(z)$, *$s(z)$ is the Stieltjes transform of the LSD of $\mathbf{S}^{-1}\mathbf{T}$, and $\tilde{s}(z)$ is the limit of* $\frac{1}{p}\mathrm{tr}\left(\mathbf{S} - \frac{\mathbf{T}}{z}\right)^{-1}\mathbf{T}$.

Proof. We have

$$\frac{1}{p}\mathrm{tr}(\mathbf{S}^{-1}\mathbf{T} - z\mathbf{I})^{-1} = -z^{-1} - \frac{1}{z^2}\frac{1}{p}\mathrm{tr}\left(\mathbf{S} - \frac{\mathbf{T}}{z}\right)^{-1}\mathbf{T}$$

and

$$s_n(z) = \frac{1}{pz}\sum_{i=1}^{n} \frac{\boldsymbol{\alpha}_i^*(z^{-1}\mathbf{T} - \mathbf{S}_i)^{-1}\boldsymbol{\alpha}_i}{1 - \boldsymbol{\alpha}_i^*(z^{-1}\mathbf{T} - \mathbf{S}_i)^{-1}\boldsymbol{\alpha}_i} = -\frac{1}{yz} + \frac{1}{yz}\frac{1}{n}\sum_{i=1}^{n} \frac{1}{1 - \boldsymbol{\alpha}_i^*(z^{-1}\mathbf{T} - \mathbf{S}_i)^{-1}\boldsymbol{\alpha}_i}$$

where $\boldsymbol{\alpha}_i = \frac{1}{\sqrt{n}}\mathbf{X}_i$, $i = 1\cdots, n$. Let the limit of $\frac{1}{p}\mathrm{tr}\left(\mathbf{S} - \frac{\mathbf{T}}{z}\right)^{-1}\mathbf{T}$ be $\tilde{s}(z)$ and $\tilde{s}_n(z) = \frac{1}{p}\mathrm{tr}\left(\mathbf{S} - \frac{\mathbf{T}}{z}\right)^{-1}\mathbf{T}$. Let

$$\underline{\tilde{s}}(z) = -(1-y)z + y\tilde{s}(z).$$

In fact, we have

$$s_n(z) = -z^{-1} - \frac{1}{z^2}\tilde{s}_n(z) \tag{4.6}$$

$$s(z) = -z^{-1} - \frac{1}{z^2}\tilde{s}(z) = -\frac{1}{yz} + \frac{1}{yz}\frac{1}{1 - y\tilde{m}(z)}$$

$$\underline{\tilde{s}}(z) = -(1-y)z + y\tilde{s}(z) = \frac{-z}{1-y\tilde{m}(z)}, \quad \mathrm{E}\underline{\tilde{s}}_n(z) = -zE\beta_i(z) \tag{4.7}$$

where $\beta_i(z) = \frac{1}{1-\boldsymbol{\alpha}_i^*(\frac{\mathbf{T}}{z}-\mathbf{S}_i)^{-1}\boldsymbol{\alpha}_i}$. Therefore, we have

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

That is,

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

The proof of the lemma is complete. ■

5. Proof of Theorem 2.2

The strategy of the proof follows the one devised in [6] and later improved in [2]. However, because we are dealing with a scaled precision matrix and the scaling matrix $\mathbf{T}$ can be non invertible, most of the steps are different from [2] and need to be worked out properly.

First, due to Assumption 1*, we may truncate the random variables x_{ij} at $\eta_n\sqrt{n}$ and normalise them without altering the CLT of $Z_n(f)$, where $\eta_n \downarrow 0$ with some slow rate. Therefore, we may make the following additional assumptions:

(1) $|x_{ij}| \le \eta_n\sqrt{n}$;
(2) $\mathrm{E}x_{ij}^2 = \kappa - 1 + o(n^{-1})$;
(3) $\mathrm{E}|x_{ij}^4| = 1 + \kappa + o(1)$.

Define $\mathcal{C}$ as a positively oriented rectangle

$$\mathcal{C} = \mathcal{C}_l \cup \mathcal{C}_u \cup \mathcal{C}_b \cup \mathcal{C}_r$$

where

$$\begin{aligned}
&\mathcal{C}_u = \{x+i\nu_0 : x\in[x_l,x_r]\}, \quad \mathcal{C}_l = \{x_l+i\nu : |\nu|\le\nu_0\},\\
&\mathcal{C}_b = \{x-i\nu_0 : x\in[x_l,x_r]\}, \quad \mathcal{C}_r = \{x_r+i\nu : |\nu|\le\nu_0\},\\
&(x_l,x_r) \supset [\liminf\lambda_{\min}(\mathbf{T})/(1+\sqrt{y})^2,\ \limsup\lambda_{\max}(\mathbf{T})/(1-\sqrt{y})^2],
\end{aligned}$$

so that it is enclosed in the analytic region $\mathcal{D}$ of the $f_j(x)$'s. By the selection of (x_l, x_r), there exists a positive constant ε such that

$$x_l \le \frac{\lambda_{\min}(\mathbf{T})}{(1+\sqrt{y})^2+2\varepsilon} < \frac{\lambda_{\min}(\mathbf{T})}{(1-\sqrt{y})^2-2\varepsilon} \le x_r,$$

for all large n. Define

$$\mathcal{B}_n = \{\eta_l = (1-\sqrt{y})^2-\varepsilon \le \lambda_{\min}(\mathbf{S}) < \lambda_{\max}(\mathbf{S}) \le \eta_r = (1+\sqrt{y})^2+\varepsilon\}.$$

It is known from [6] that for any given $t>0$, we have

$$1-\mathrm{P}(\mathcal{B}_n) = o(n^{-t}) \tag{5.1}$$

as $n \to \infty$. It follows that for the proof of Theorem 2.2, we can restrict our attention by conditioning on the event $\mathcal{B}_n$ without altering the central limit results. This will be assumed in what follows in the remaining of the proof.

In particular, conditional on $\mathcal{B}_n$, all the eigenvalues of $\mathbf{S}$ fall inside the interval $((1-\sqrt{y})^2-\varepsilon,(1+\sqrt{y})^2+\varepsilon)$ and all the eigenvalues of $\mathbf{S}^{-1}T$ fall inside the interval (x_l, x_r).

Furthermore, let

$$\mathcal{C}_n = \mathcal{C}\bigcap\{z : |\Im z| > n^{-2}\}.$$

Since f is analytic, by Cauchy's theorem, we have

$$\begin{aligned} X_n(f) &= -\frac{p}{2\pi i}\oint_{\mathcal{C}} f(z)\{s_n(z) - s_n^0(z)\}dz \\ &= -\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)M_n(z)dz \end{aligned} \tag{5.2}$$

where $s_n^0(z)$ is the Stieltjes transform of F^{y_n,H_n} and

$$\begin{aligned} M_n(z) &= p\{s_n(z) - s_n^0(z)\} \\ &= p\{s_n(z) - \mathrm{E}s_n(z)\} + p\{\mathrm{E}s_n(z) - s_n^0(z)\} \\ &=: M_n^1(z) + M_n^2(z), \end{aligned}$$

where $M_n^1(z)$ and $M_n^2(z)$ represent the random fluctuation and the deterministic part of $M_n(z)$, respectively. Let

$$\hat{M}_n(z) = \begin{cases} M_n(z), & x \in \mathcal{C}_n, \\ M_n(x_l + in^{-2}), & \Re z = x_l, \Im z \in [0, n^{-2}], \\ M_n(x_r + in^{-2}), & \Re z = x_r, \Im z \in [0, n^{-2}]. \end{cases}$$

Then, we have

$$\begin{aligned} &\int\limits_{\mathcal{C}} f(z)(M_n(z) - \hat{M}_n(z))dz = \int\limits_{\mathcal{C}/\mathcal{C}_n} f(z)(M_n(z) - \hat{M}_n(z))dz \\ &\le Kn^{-1}\Bigg\{ \left[\min\left(\frac{\liminf \lambda_{\min}(\mathbf{T})}{(1+\sqrt{y})^2}, \frac{\liminf \lambda_{\min}(\mathbf{T})}{\lambda_{\max}(\mathbf{S})}\right) - x_l\right]^{-1} \\ &\qquad + \left[\max\left[\frac{\limsup \lambda_{\max}(\mathbf{T})}{(1-\sqrt{y})^2}, \frac{\limsup \lambda_{\max}(\mathbf{T})}{\lambda_{\min}(\mathbf{S})}\right] - x_r\right]^{-1} \Bigg\} \\ &= o_p(1). \end{aligned}$$

Thus our goal will be to prove that

- $\{\hat{M}_n^1(z);\ z \in \mathcal{C}\}$ converges to a centred Gaussian process indexed by $\mathcal{C}$ with an explicit covariance function;
- $\hat{M}_n^2(z)$ converges uniformly to a deterministic mean function defined on $\mathcal{C}$.

Under the finite dimensional convergence with tightness, it is well known that the contour integration of the empirical Stieltjes transform tends to the contour integration of limit process which is Gaussian. Then the conclusion follows because the contour integration of the limit process is a limit of a sum of Gaussian variables (the Darbous sum). Since $X_n(f)$ is an integral of $M_n(z)$, it will converges to a Gaussian process whose limiting covariance and mean functions are integrals of the corresponding limits of the process $M_n(z)$.

The convergence of $\{\hat{M}_n^1(z);\ z \in \mathcal{C}\}$ is done in two steps. Section 5.1 establishes its finite-dimensional convergence while its tightness is proved in Section 5.2. The convergence of $\{\hat{M}_n^2(z)\}$ is established in Section 5.3. It is here noticed that since we can make ν_0 arbitrarily small, the contributions from the segments $\mathcal{C}_l$ and $\mathcal{C}_r$ will be negligible. Therefore, we need to establish these convergence on the segments $\mathcal{C}_u$ and $\mathcal{C}_b$ only. It is noticed that the convergence of $M_n^1(z)$ on $\mathcal{C}_u \cup \mathcal{C}_b$ will be given in Section 5.1 and the tightness of $M_n^1(z)$ will be given in Section 5.2.

5.1. *Finite-dimensional convergence of $M_n^1(z)$ on $\mathcal{C}_u \cup \mathcal{C}_b$*

Lemma 5.1. *Under Assumptions 1*, 2*, and 3, the process $\{M_n^1(z) = p(s_n(z) - Es_n(z))\}$ converges weakly to a complex Gaussian process $M_1(\cdot)$ on the contour $z \in \mathcal{C}_u \cup \mathcal{C}_b$, with the mean function*

$$\mathrm{E}M_1(z) = 0$$

and covariance function

$$\mathrm{Cov}(M_1(z_1), M_1(z_2)) = \kappa \left[\frac{\frac{\partial(z_1 b(z_1))}{\partial z_1} \frac{\partial(z_2 b(z_2))}{\partial z_2}}{[z_1 b(z_1) - z_2 b(z_2)]^2} - \frac{1}{(z_1 - z_2)^2} \right]. \tag{5.3}$$

Proof. Let E_i denote the conditional expectation given $\{\mathbf{x}_1, \cdots, \mathbf{x}_i\}$ and E_0 denote the unconditional expectation. Denote $z = u + iv$ with $v > 0$ fixed,

$$\begin{aligned}
&\boldsymbol{\alpha}_i = \tfrac{1}{\sqrt{n}}\mathbf{x}_i,\ \mathbf{S}_i = \mathbf{S} - \boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*, && \mathbf{D} = \mathbf{T} - z\mathbf{S},\ \mathbf{D}_i = \mathbf{T} - z\mathbf{S}_i,\\
&\beta_i(z) = \tfrac{1}{1-\boldsymbol{\alpha}_i^*(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1}\boldsymbol{\alpha}_i}, && \bar{\beta}_i(z) = \tfrac{1}{1-\frac{1}{n}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1}},\\
&\hat{\gamma}_i(z) = \tfrac{1}{n}\mathrm{tr}(z^{-1}\mathbf{T} - \mathbf{S}_i)^{-1} - \boldsymbol{\alpha}_i^*(z^{-1}\mathbf{T} - \mathbf{S}_i)^{-1}\boldsymbol{\alpha}_i,\\
&R_i(z) = \log\left(1 - \bar{\beta}_i(z)\hat{\gamma}_i(z)\right) + \bar{\beta}_i(z)\hat{\gamma}_i(z).
\end{aligned}$$

Then we have

$$\beta_i(z)^{-1} = \bar{\beta}_i(z)^{-1} - \hat{\gamma}_i(z), \quad \bar{\beta}_i(z)\beta_i^{-1}(z) = 1 - \bar{\beta}_i(z)\hat{\gamma}_i(z).$$

Therefore, by Taylor expansion

$$\begin{aligned}
&(\mathrm{E}_i - \mathrm{E}_{i-1}) \log \beta_i^{-1}(z)\\
&= (\mathrm{E}_i - \mathrm{E}_{i-1}) \left(\log \beta_i^{-1}(z) - \log \bar{\beta}_i^{-1}(z)\right)\\
&= (\mathrm{E}_i - \mathrm{E}_{i-1})[-\bar{\beta}_i(z)\hat{\gamma}_i(z) + R_i(z)].
\end{aligned} \tag{5.4}$$

Here, we have used a formula that $\log \beta_i^{-1} - \log \bar{\beta}_i^{-1} = \log \bar{\beta}_i(z)\beta_i^{-1}$. In fact we should add an additional term $2\pi i k(z)$ where $k(z)$ is a random integer function of z. This term does not make any contribution to the limiting distribution because we only need the derivative of the function $\log \beta_i^{-1}$ in the next step.

Step 1. Show that $M_n^1(z) = p(s_n(z) - \mathrm{E}s_n(z)) = \sum_{i=1}^n Y_i(z) + o_p(1)$ where $\{Y_i(z)\}$ is a sequence of martingale differences with respect to the filtration given by the σ-algebras generated by the random variables $x_1, \ldots, x_n$.

For this purpose, first we split $s_n(z)$ in the following way: for each $i \le n$, we have

$$\begin{aligned}
s_n(z) &= \frac{1}{p}\mathrm{tr}(\mathbf{S}^{-1}\mathbf{T} - z\mathbf{I})^{-1} = \frac{1}{p}\mathrm{tr}\mathbf{S}(\mathbf{T} - z\mathbf{S})^{-1} \\
&= \frac{1}{p}\mathrm{tr}(\mathbf{S}_i + \boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*)(\mathbf{T} - z\mathbf{S}_i - z\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*)^{-1} = -\frac{1}{pz}\mathrm{tr}(\mathbf{D}_i - \mathbf{T} - z\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*)(\mathbf{D}_i - z\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*)^{-1} \\
&= -z^{-1} + \frac{1}{pz}\mathrm{tr}\mathbf{T}\mathbf{D}_i^{-1} + \frac{1}{p}\frac{\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{T}\mathbf{D}_i^{-1}\boldsymbol{\alpha}_i}{1 - z\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\boldsymbol{\alpha}_i} \\
&= -z^{-1} + \frac{1}{pz}\mathrm{tr}\mathbf{T}\mathbf{D}_i^{-1} + \frac{1}{p}\frac{\partial}{\partial z}\log \beta_i(z) \ .
\end{aligned}$$

Therefore, using the martingale decomposition, $s_n(z) - \mathrm{E}s_n(z)$ can be simplified as

$$\begin{aligned}
s_n(z) - \mathrm{E}s_n(z) &= -\frac{1}{p}\frac{\partial}{\partial z}\sum_{i=1}^n(\mathrm{E}_i - \mathrm{E}_{i-1})\log(1 - z\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\boldsymbol{\alpha}_i) \\
&= -\frac{1}{p}\frac{\partial}{\partial z}\sum_{i=1}^n \mathrm{E}_i\bar{\beta}_i(z)\hat{\gamma}_i(z) - \frac{1}{p}\frac{\partial}{\partial z}\sum_{i=1}^n(\mathrm{E}_i - \mathrm{E}_{i-1})R_i(z) \\
&= -\frac{1}{p}\frac{\partial}{\partial z}\sum_{i=1}^n \mathrm{E}_i\bar{\beta}_i(z)\hat{\gamma}_i(z) + o_p(1),
\end{aligned}$$

because, by Cauchy inequality

$$\begin{aligned}
&\mathrm{E}\left|\frac{d}{dz}\sum_{i=1}^n(\mathrm{E}_{i-1} - \mathrm{E}_i)R_i(z)\right|^2 = \mathrm{E}\left|\frac{1}{2i\pi}\oint_{|\zeta - z| = v/2}\sum_{i=1}^n(\mathrm{E}_{i-1} - \mathrm{E}_i)\frac{R_i(\zeta)d\zeta}{(\zeta - z)^2}\right|^2 \\
&\le C\oint_{|\zeta - z| = v/2}\sum_{i=1}^n \mathrm{E}\left|(\mathrm{E}_{i-1} - \mathrm{E}_i)R_i(\zeta)\right|^2 |d\zeta| \\
&\le C\oint_{|\zeta - z| = v/2}\sum_{i=1}^n \mathrm{E}\left|\bar{\beta}_i^2(\zeta)\hat{\gamma}_i^2(\zeta)\right|^2 |d\zeta| \\
&\le C\oint_{|\zeta - z| = v/2}\sum_{i=1}^n \mathrm{E}\left|\hat{\gamma}_i(\zeta)\right|^4 |d\zeta| = nO(n^{-1}\eta_n^4) = o(1) \qquad (5.5)
\end{aligned}$$

where the last equation is by Lemma 9.1 of [2]) and C depends on other constants, scuh as ν. So we have

$$p((s_n(z)-\mathrm{E}s_n(z)) = -\sum_{i=1}^n \mathrm{E}_i \frac{d}{dz}\bar{\beta}_i(z)\hat{\gamma}_i(z) + o_p(1) = \sum_{i=1}^n Y_i(z) + o_p(1),$$

where $Y_i(z) = -\mathrm{E}_i \frac{d}{dz}\bar{\beta}_i(z)\hat{\gamma}_i(z)$. To prove that $\{M_n^1(z) = p(s_n(z)-\mathrm{E}s_n(z))\}$ converges to a Gaussian process $\{M_1(z)\}$, we first consider a finite sum

$$\sum_{k=1}^r a_k \sum_{i=1}^n Y_i(z_k) = \sum_{i=1}^n \sum_{k=1}^r a_k Y_i(z_k),$$

from r points $\{z_k\}$ on the contour with arbitrary weighting numbers $\{a_k\}$. We are going to apply Lyapounov CLT for martingales.

Step 2: Verify the Lyapunov condition, i.e. $\sum_{i=1}^n \mathrm{E}\,|Y_i(z)|^4 = o(1)$ as $n \to \infty$ uniformly for z on $\mathcal{C}_n$.

In fact, if $z \in \mathcal{C}_u$ or $\mathcal{C}_b$, by the same approach of the proof of (5.5) and the fact that $|\bar{\beta}_i(\zeta)| < 2|\zeta|/\nu_0$, we have

$$\sum_{i=1}^n \mathrm{E}\,|Y_i(z)|^4 \le C \oint_{|z-\zeta|=v/2} \sum_{i=1}^n \mathrm{E}\,|\hat{\gamma}_i(\zeta)|^4\, d\zeta \to 0.$$

Step 3: Find the covariance $\mathrm{Cov}(M_1(z_1), M_1(z_2))$, that is, the limits of

$$\sum_{i=1}^n \mathrm{E}_{i-1} Y_i(z_1)Y_i(z_2) = \frac{\partial^2}{\partial z_1 \partial z_2} \sum_{i=1}^n \mathrm{E}_{i-1}[\mathrm{E}_i\bar{\beta}_i(z_1)\hat{\gamma}_i(z_1)\mathrm{E}_i\bar{\beta}_i(z_2)\hat{\gamma}_i(z_2)].$$

Step 3.1: First we will prove

$$\begin{aligned}&\sum_{i=1}^n \mathrm{E}_{i-1}[\mathrm{E}_i\bar{\beta}_i(z_1)\hat{\gamma}_i(z_1)\mathrm{E}_i\bar{\beta}_i(z_2)\hat{\gamma}_i(z_2)]\\ &= \frac{\kappa}{n^2}\sum_{i=1}^n b_i(z_1)b_i(z_2)\mathrm{tr}\left[\mathrm{E}_i(\frac{1}{z_1}\mathbf{T}-\mathbf{S}_i)^{-1}\mathrm{E}_i(\frac{1}{z_2}\mathbf{T}-\mathbf{S}_i)^{-1}\right] + o_p(1).\end{aligned}$$

We have

$$\bar{\beta}_i(z) - b_i(z) = \bar{\beta}_i(z)b_i(z)\left(n^{-1}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1} - n^{-1}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1}\right)$$

where $b_i(z) = 1/\{1-n^{-1}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1}\}$. Then by similar approach of martingale decomposition, we obtain

$$\begin{aligned}\mathrm{E}|\bar{\beta}_i(z)-b_i(z)|^{2l} &\le K\mathrm{E}\left(n^{-1}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1} - n^{-1}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1}\right)^{2l}\\ &\le K\mathrm{E}\left(\frac{1}{n}\sum_{\substack{j=1\\ j\ne i}}^n\left[\mathrm{E}_j\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1} - \mathrm{E}_{j-1}\mathrm{tr}(\frac{\mathbf{T}}{z}-\mathbf{S}_i)^{-1}\right]\right)^{2l}\end{aligned}$$

$$\leq \frac{K}{n^l \nu^{2l}}.$$

where the inequality is by Lemma 2.12 of [2] and Burkholder inequality. Then $\mathrm{E}|\bar{\beta}_i(z) - b_i(z)|^{2l} = O(n^{-l})$ uniformly in i. Consequently, we have

$$\sum_{i=1}^{n} \mathrm{E}_{i-1}[\mathrm{E}_i\bar{\beta}_i(z_1)\hat{\gamma}_i(z_1)\mathrm{E}_i\bar{\beta}_i(z_2)\hat{\gamma}_i(z_2)] - \sum_{i=1}^{n} b_i(z_1)b_i(z_2)\mathrm{E}_{i-1}[\mathrm{E}_i\hat{\gamma}_i(z_1)\mathrm{E}_i\hat{\gamma}_i(z_2)] = o_p(1).$$

Therefore, we need only consider the limit of

$$\begin{aligned}&\sum_{i=1}^{n} b_i(z_1)b_i(z_2)\mathrm{E}_{i-1}[\mathrm{E}_i\hat{\gamma}_i(z_1)\mathrm{E}_i\hat{\gamma}_i(z_2)]\\ &= \frac{\kappa}{n^2}\sum_{i=1}^{n} b_i(z_1)b_i(z_2)\mathrm{tr}\left[\mathrm{E}_i(\frac{\mathbf{T}}{z_1} - \mathbf{S}_i)^{-1}\mathrm{E}_i(\frac{\mathbf{T}}{z_2} - \mathbf{S}_i)^{-1}\right], \end{aligned} \tag{5.6}$$

where the equality follows from the equation in (1.15) of [6]

$$\begin{aligned}&\mathrm{E}(\mathbf{x}_t^*\mathbf{A}\mathbf{x}_t - \mathrm{tr}\mathbf{A})(\mathbf{x}_t^*\mathbf{B}\mathbf{x}_t - \mathrm{tr}\mathbf{B})\\ &= (\mathrm{E}|X_{11}|^4 - |EX_{11}^2|^2 - 2)\sum_{i=1}^{n} a_{ii}b_{ii} + |EX_{11}^2|^2\mathrm{tr}(\mathbf{A}\mathbf{B}^T) + \mathrm{tr}(\mathbf{A}\mathbf{B}).\end{aligned} \tag{5.7}$$

Step 3.2: we will proceed as in [6] to show that

$$\begin{aligned}&\mathrm{tr}\left[\mathrm{E}_i(\frac{1}{z_1}\mathbf{T} - \mathbf{S}_i)^{-1}\mathrm{E}_i(\frac{1}{z_2}\mathbf{T} - \mathbf{S}_i)^{-1}\right]\\ &= \frac{\mathrm{tr}\left[(\frac{n-1}{n}b_i(z_1)\mathbf{I} - \frac{1}{z_1}\mathbf{T})^{-1}(\frac{n-1}{n}b_i(z_2)\mathbf{I} - \frac{1}{z_2}\mathbf{T})^{-1}\right] + O_p(n^{1/2})}{1 - \frac{(i-1)}{n^2}b_i(z_1)b_i(z_2)\mathrm{tr}(\frac{n-1}{n}b_i(z_1)\mathbf{I} - \frac{1}{z_1}\mathbf{T})^{-1}(\frac{n-1}{n}b_i(z_2)\mathbf{I} - \frac{1}{z_2}\mathbf{T})^{-1}}.\end{aligned}$$

By multiplying $(\frac{n-1}{n}b_i(z)\mathbf{I} - z^{-1}\mathbf{T})^{-1}$ from left and $(z^{-1}\mathbf{T} - \mathbf{S}_i)^{-1}$ from right on the identity

$$(z^{-1}\mathbf{T} - \mathbf{S}_i) - z^{-1}\mathbf{T} + \frac{n-1}{n}b_i(z)\mathbf{I} = -\sum_{k\neq i}\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^* + \frac{n-1}{n}b_i(z)\mathbf{I},$$

we have

$$\begin{aligned}(z^{-1}\mathbf{T} - \mathbf{S}_i)^{-1} &= -(\frac{n-1}{n}b_i(z)\mathbf{I} - z^{-1}\mathbf{T})^{-1}\\ &\quad - \sum_{k\neq i}\beta_{k(i)}(z)(\frac{n-1}{n}b_i(z)\mathbf{I} - z^{-1}\mathbf{T})^{-1}\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^*(z^{-1}\mathbf{T} - \mathbf{S}_{ik})^{-1}\\ &\quad + \frac{n-1}{n}b_i(z)(\frac{n-1}{n}b_i(z)\mathbf{I} - z^{-1}\mathbf{T})^{-1}(z^{-1}\mathbf{T} - \mathbf{S}_i)^{-1}\\ &= -(\frac{n-1}{n}b_i(z)\mathbf{I} - z^{-1}\mathbf{T})^{-1} - b_i(z)\mathbf{A}(z) - \mathbf{B}(z) - \mathbf{C}(z),\end{aligned} \tag{5.8}$$

where

$$\mathbf{A}(z) = \sum_{k\neq i}\left(\frac{n-1}{n}b_i(z)\mathbf{I} - z^{-1}\mathbf{T}\right)^{-1}(\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^* - \frac{1}{n}\mathbf{I})(z^{-1}\mathbf{T} - \mathbf{S}_{ik})^{-1},$$

$$\mathbf{B}(z) = \sum_{k\neq i}(\beta_{k(i)}(z) - b_i(z))\left(\frac{n-1}{n}b_j(z)\mathbf{I} - z^{-1}\mathbf{T}\right)^{-1}\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^*(z^{-1}\mathbf{T} - \mathbf{S}_{ik})^{-1},$$

$$\mathbf{C}(z) = \frac{1}{n}b_i(z)\left(\frac{n-1}{n}b_i(z)\mathbf{I} - z^{-1}\mathbf{T}\right)^{-1}\sum_{k\neq i}\left[(z^{-1}\mathbf{T} - \mathbf{S}_{ik})^{-1} - (z^{-1}\mathbf{T} - \mathbf{S}_i)^{-1}\right].$$

So we have

$$\begin{aligned}
&\operatorname{tr}\left[\mathrm{E}_i(\frac{1}{z_1}\mathbf{T} - \mathbf{S}_i)^{-1}\mathrm{E}_i(\frac{1}{z_2}\mathbf{T} - \mathbf{S}_i)^{-1}\right] \\
&= -\operatorname{tr}(\frac{n-1}{n}b_i(z_1)\mathbf{I} - z_1^{-1}\mathbf{T})^{-1}\mathrm{E}_i(\frac{1}{z_2}\mathbf{T} - \mathbf{S}_i)^{-1} - b_i(z_1)\operatorname{tr}\mathrm{E}_i\mathbf{A}(z_1)\mathrm{E}_i(\frac{\mathbf{T}}{z_2} - \mathbf{S}_i)^{-1} \\
&\quad -\operatorname{tr}\mathrm{E}_i\mathbf{B}(z_1)\mathrm{E}_i(\frac{1}{z_2}\mathbf{T} - \mathbf{S}_i)^{-1} - \operatorname{tr}\mathrm{E}_i\mathbf{C}(z_1)\mathrm{E}_i(\frac{1}{z_2}\mathbf{T} - \mathbf{S}_i)^{-1}. \qquad (5.9)
\end{aligned}$$

In the following we will prove

$$\mathrm{E}\left|\operatorname{tr}\mathrm{E}_i\mathbf{B}(z_1)\mathrm{E}_i(\frac{1}{z_2}\mathbf{T} - \mathbf{S}_i)^{-1}\right| = O(n^{1/2}) \qquad (5.10)$$

$$\mathrm{E}\left|\operatorname{tr}\mathrm{E}_i\mathbf{C}(z_1)\mathrm{E}_i(\frac{1}{z_2}\mathbf{T} - \mathbf{S}_i)^{-1}\right| = O(1) \qquad (5.11)$$

and

$$\begin{aligned}
&b_i(z_1)\operatorname{tr}\mathrm{E}_i\mathbf{A}(z_1)\mathrm{E}_i(\frac{1}{z_2}\mathbf{T} - \mathbf{S}_i)^{-1} \\
&= \frac{i-1}{n^2}b_i(z_1)b_i(z_2)\mathrm{E}_i\operatorname{tr}(\frac{1}{z_1}\mathbf{T} - \mathbf{S}_i)^{-1}\mathrm{E}_i(\frac{1}{z_2}\mathbf{T} - \mathbf{S}_i)^{-1} \\
&\quad \operatorname{tr}\mathrm{E}_i(\frac{1}{z_2}\mathbf{T} - \mathbf{S}_i)^{-1}(\frac{n-1}{n}b_i(z_1)\mathbf{I} - \frac{1}{z_1}\mathbf{T})^{-1} + O_p(n^{1/2}) \qquad (5.12)
\end{aligned}$$

Because

$$\begin{aligned}
&\left|\left(\frac{n-1}{n}b_i(z)\mathbf{I} - \frac{\mathbf{T}}{z}\right)^{-1}\right| \\
&= \left|\left(\frac{1}{1 - \frac{1}{n}\mathrm{E}\operatorname{tr}(z^{-1}\mathbf{T} - \mathbf{S}_i)^{-1}}\mathbf{I} - \frac{\mathbf{T}}{z}\right)^{-1}\right| \\
&= \left|\left(\frac{1}{1 - \frac{z}{n}\mathrm{E}\mathbf{S}_i^{-\frac{1}{2}}\operatorname{tr}(\mathbf{S}_i^{-\frac{1}{2}}\mathbf{T}\mathbf{S}_i^{-\frac{1}{2}} - z\mathbf{I})^{-1}\mathbf{S}_i^{-\frac{1}{2}}}\mathbf{I} - \frac{\mathbf{T}}{z}\right)^{-1}\right| \\
&= \left|\sum_{j=1}^{p}\Gamma_j'\frac{z\left(1 - \frac{z}{n}\mathrm{E}\mathbf{S}_i^{-\frac{1}{2}}\operatorname{tr}(\mathbf{S}_i^{-\frac{1}{2}}\mathbf{T}\mathbf{S}_i^{-\frac{1}{2}} - z\mathbf{I})^{-1}\mathbf{S}_i^{-\frac{1}{2}}\right)}{z - \lambda^{\mathbf{T}}\left(1 - \frac{z}{n}\mathrm{E}\mathbf{S}_i^{-\frac{1}{2}}\operatorname{tr}(\mathbf{S}_i^{-\frac{1}{2}}\mathbf{T}\mathbf{S}_i^{-\frac{1}{2}} - z\mathbf{I})^{-1}\mathbf{S}_i^{-\frac{1}{2}}\right)}\Gamma_j\right| \\
&= \left|\sum_{j=1}^{p}\Gamma_j'\frac{|z|^2\left(1 - \frac{z}{n}\mathrm{E}\mathbf{S}_i^{-\frac{1}{2}}\operatorname{tr}(\mathbf{S}_i^{-\frac{1}{2}}\mathbf{T}\mathbf{S}_i^{-\frac{1}{2}} - z\mathbf{I})^{-1}\mathbf{S}_i^{-\frac{1}{2}}\right)}{|z|^2 - \lambda^{\mathbf{T}}\left(\bar{z} - \frac{|z|^2}{n}\mathrm{E}\mathbf{S}_i^{-\frac{1}{2}}\operatorname{tr}(\mathbf{S}_i^{-\frac{1}{2}}\mathbf{T}\mathbf{S}_i^{-\frac{1}{2}} - z\mathbf{I})^{-1}\mathbf{S}_i^{-\frac{1}{2}}\right)}\Gamma_j\right|
\end{aligned}$$

$$\leq \left| \sum_{j=1}^{p} \Gamma_j' \frac{|z|^2 \left(1 + \frac{|z|}{\nu_0}\right)}{\nu_0} \Gamma_j \right| = \frac{|z|^2 \left(1 + \frac{|z|}{\nu_0}\right)}{\nu_0},$$

then we have

$$\left\| \left(\frac{n-1}{n} b_j(z)\mathbf{I} - z^{-1}\mathbf{T} \right)^{-1} \right\| \leq K,$$

where K is a constant. Moreover, by Lemma 9.1 of [2]

$$|b_{ik}(z) - b_i(z)| = \left| b_i(z) b_{ik}(z) \left[\frac{1}{n} \mathrm{Etr}(\frac{\mathbf{T}}{z} - \mathbf{S}_i)^{-1} - \frac{1}{n} \mathrm{Etr}(\frac{\mathbf{T}}{z} - \mathbf{S}_{ik})^{-1} \right] \right|$$
$$= \frac{1}{n} \left| b_{ik}(z) b_i(z) \mathrm{E}\beta_{k(i)}(z) \boldsymbol{\alpha}_k^* (\frac{\mathbf{T}}{z} - \mathbf{S}_{ik})^{-2} \boldsymbol{\alpha}_k \right| = O(n^{-1}), \quad (5.13)$$

$$\mathrm{E}(\beta_{k(i)}(z) - b_{ik}(z))^2 = O(n^{-1}), \quad (5.14)$$

where

$$\beta_{k(i)}(z) = \frac{1}{1 - \boldsymbol{\alpha}_k^* (z^{-1}\mathbf{T} - \mathbf{S}_{ik})^{-1} \boldsymbol{\alpha}_k}$$

$$b_i(z) = \frac{1}{1 - \frac{1}{n}\mathrm{Etr}(z^{-1}\mathbf{T} - \mathbf{S}_i)^{-1}}, \quad b_{ik}(z) = \frac{1}{1 - \frac{1}{n}\mathrm{Etr}(z^{-1}\mathbf{T} - \mathbf{S}_{ik})^{-1}}.$$

Step 3.2.1: The proof of (5.10) follows easily from

$$\mathrm{E}\left| \mathrm{trE}_i \mathbf{B}(z_1) \mathrm{E}_i (\frac{1}{z_2}\mathbf{T} - \mathbf{S}_i)^{-1} \right|$$
$$\leq \sum_{k \neq i} \mathrm{E} \left| \mathrm{E}_i \mathrm{tr}(\beta_{k(i)}(z_1) - b_i(z_1)) (\frac{n-1}{n} b_j(z_1)\mathbf{I} - \frac{\mathbf{T}}{z_1})^{-1} \boldsymbol{\alpha}_k \boldsymbol{\alpha}_k^* (\frac{\mathbf{T}}{z_1} - \mathbf{S}_{ik})^{-1} (\frac{\mathbf{T}}{z_2} - \breve{\mathbf{S}}_i)^{-1} \right|$$
$$\leq \sum_{k \neq i} \mathrm{E} \left| \mathrm{E}_i (\beta_{k(i)}(z) - b_i(z)) \boldsymbol{\alpha}_k^* (z^{-1}\mathbf{T} - \mathbf{S}_{ik})^{-1} (\frac{1}{z_2}\mathbf{T} - \breve{\mathbf{S}}_i)^{-1} \boldsymbol{\alpha}_k \right|$$
$$\leq \sum_{k \neq i} \mathrm{E}^{1/2} |(\beta_{k(i)}(z) - b_i(z))^2| \mathrm{E}^{1/2} \left| \boldsymbol{\alpha}_k^* (z^{-1}\mathbf{T} - \mathbf{S}_{ik})^{-1} (\frac{1}{z_2}\mathbf{T} - \breve{\mathbf{S}}_i)^{-1} \boldsymbol{\alpha}_k \right|^2$$
$$= O(n^{1/2}),$$

and $\breve{\mathbf{S}}_i$ is the analogue for the matrix $\mathbf{S}_i$ with vectors $\mathbf{x}_{i+1}, \cdots, \mathbf{x}_n$ replaced by their iid copies $\breve{\mathbf{x}}_{i+1}, \cdots, \breve{\mathbf{x}}_n$. $(\breve{\mathbf{x}}_{i+1}, \cdots, \breve{\mathbf{x}}_n)$ is independent of $\mathbf{x}_{i+1}, \cdots, \mathbf{x}_n$.

Step 3.2.2: The estimation (5.11) follows from

$$\frac{1}{n} \sum_{k \neq i} \mathrm{E} \left| \mathrm{trE}_i \mathbf{C}(z_1) \mathrm{E}_i (\frac{1}{z_2}\mathbf{T} - \mathbf{S}_i)^{-1} \right|$$
$$\leq \frac{1}{n} \sum_{k \neq i} \mathrm{E} \left| \mathrm{trE}_i b_i(z) \left(\frac{n-1}{n} b_j(z)\mathbf{I} - \frac{\mathbf{T}}{z} \right)^{-1} \left(\left(\frac{\mathbf{T}}{z_1} - \mathbf{S}_{ik} \right)^{-1} - \left(\frac{\mathbf{T}}{z} - \mathbf{S}_i \right)^{-1} \right) \left(\frac{\mathbf{T}}{z_2} - \breve{\mathbf{S}}_i \right)^{-1} \right|$$

$$= \frac{1}{n}\sum_{k\neq i}\mathrm{E}\left|\mathrm{E}_i b_i(z)\left(\frac{(n-1)\mathbf{I}}{nb_j^{-1}(z)}-\frac{\mathbf{T}}{z}\right)^{-1}\beta_{k(i)}\boldsymbol{\alpha}_k^*\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\left(\frac{\mathbf{T}}{z_2}-\breve{\mathbf{S}}_i\right)^{-1}\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\boldsymbol{\alpha}_k\right|$$
$$\le K.$$

Step 3.2.3: As for (5.12), we have

$$\begin{aligned}
&b_i(z_1)\mathrm{tr}\mathrm{E}_i\mathbf{A}(z_1)\mathrm{E}_i(\frac{1}{z_2}\mathbf{T}-\mathbf{S}_i)^{-1}\\
=\;& b_i(z_1)\mathrm{tr}\sum_{k<i}\left(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{\mathbf{T}}{z_1}\right)^{-1}(\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^*-\frac{1}{n}\mathbf{I})\mathrm{E}_i\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\mathrm{E}_i\left(\frac{\mathbf{T}}{z_2}-\mathbf{S}_i\right)^{-1}\\
=\;& b_i(z_1)\sum_{k<i}\boldsymbol{\alpha}_k^*\mathrm{E}_i\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\mathrm{E}_i\left[\left(\frac{\mathbf{T}}{z_2}-\mathbf{S}_i\right)^{-1}-\left(\frac{\mathbf{T}}{z_2}-\mathbf{S}_{ik}\right)^{-1}\right]\left(\frac{(n-1)\mathbf{I}}{nb_i^{-1}(z_1)}-\frac{\mathbf{T}}{z_1}\right)^{-1}\boldsymbol{\alpha}_k\\
&-b_i(z_1)\mathrm{tr}\frac{1}{n}\sum_{k<i}\left(\frac{(n-1)\mathbf{I}}{nb_i^{-1}(z_1)}-\frac{\mathbf{T}}{z_1}\right)^{-1}\mathrm{E}_i\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\mathrm{E}_i\left[\left(\frac{\mathbf{T}}{z_2}-\mathbf{S}_i\right)^{-1}-\left(\frac{\mathbf{T}}{z_2}-\mathbf{S}_{ik}\right)^{-1}\right]\\
&+b_i(z_1)\mathrm{tr}\sum_{k<i}\left(\frac{(n-1)\mathbf{I}}{nb_i^{-1}(z_1)}-\frac{1}{z_1}\mathbf{T}\right)^{-1}(\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^*-\frac{1}{n}\mathbf{I})\mathrm{E}_i\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\mathrm{E}_i\left(\frac{\mathbf{T}}{z_2}-\mathbf{S}_{ik}\right)^{-1}\\
=:\;& \mathbf{C}_1+\mathbf{C}_2+\mathbf{C}_3\ .
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
\mathrm{E}|\mathbf{C}_2| &= \mathrm{E}\left|\mathrm{tr}\frac{b_i(z_1)}{n}\sum_{k<i}\left(\frac{(n-1)\mathbf{I}}{nb_i^{-1}(z_1)}-\frac{\mathbf{T}}{z_1}\right)^{-1}\mathrm{E}_i\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\mathrm{E}_i\left[\left(\frac{\mathbf{T}}{z_2}-\mathbf{S}_i\right)^{-1}-\left(\frac{\mathbf{T}}{z_2}-\mathbf{S}_{ik}\right)^{-1}\right]\right|\\
&= \mathrm{E}\left|\frac{b_i(z_1)}{n}\sum_{k<i}\left(\frac{(n-1)\mathbf{I}}{nb_i^{-1}(z_1)}-\frac{\mathbf{T}}{z_1}\right)^{-1}\mathrm{E}_i\beta_{k(i)}\boldsymbol{\alpha}_k^*(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik})^{-1}(\frac{\mathbf{T}}{z_1}-\breve{\mathbf{S}}_{ik})^{-1}\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\boldsymbol{\alpha}_k\right|\\
&\le K;
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}|\mathbf{C}_3| &= \mathrm{E}\left|b_i(z_1)\mathrm{tr}\sum_{k<i}\left(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{\mathbf{T}}{z_1}\right)^{-1}(\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^*-\frac{\mathbf{I}}{n})\mathrm{E}_i\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\mathrm{E}_i\left(\frac{\mathbf{T}}{z_2}-\mathbf{S}_{ik}\right)^{-1}\right|\\
&\le \sum_{k<i}K\mathrm{E}\left|\boldsymbol{\alpha}_k^*\mathrm{E}_i\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\left(\frac{\mathbf{T}}{z_2}-\breve{\mathbf{S}}_{ik}\right)^{-1}\boldsymbol{\alpha}_k-\frac{1}{n}\mathrm{tr}\mathrm{E}_i\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\left(\frac{\mathbf{T}}{z_2}-\breve{\mathbf{S}}_{ik}\right)^{-1}\right|\\
&\le \sum_{k<i}K\mathrm{E}^{\frac{1}{2}}\left|\boldsymbol{\alpha}_k^*\mathrm{E}_i\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\left(\frac{\mathbf{T}}{z_2}-\breve{\mathbf{S}}_{ik}\right)^{-1}\boldsymbol{\alpha}_k-\frac{1}{n}\mathrm{tr}\mathrm{E}_i\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\left(\frac{\mathbf{T}}{z_2}-\breve{\mathbf{S}}_{ik}\right)^{-1}\right|^2\\
&= O(n^{\frac{1}{2}}).
\end{aligned}$$

where the second inequality uses the Cauchy-Schwarz inequality and the last equality uses (5.7). Moreover, we have

$$\mathbf{C}_1 = b_i(z_1)\sum_{k<i}\boldsymbol{\alpha}_k^*\mathrm{E}_i\left(\frac{\mathbf{T}}{z_1}-\mathbf{S}_{ik}\right)^{-1}\mathrm{E}_i\left[\left(\frac{\mathbf{T}}{z_2}-\mathbf{S}_i\right)^{-1}-\left(\frac{\mathbf{T}}{z_2}-\mathbf{S}_{ik}\right)^{-1}\right]\left(\frac{(n-1)\mathbf{I}}{nb_i^{-1}(z_1)}-\frac{\mathbf{T}}{z_1}\right)^{-1}\boldsymbol{\alpha}_k$$

$$
\begin{aligned}
&= b_i(z_1)\sum_{k<i}\mathrm{E}_i\breve{\beta}_{k(i)}(z_2)\boldsymbol{\alpha}_k^*(\frac{1}{z_1}\mathbf{T}-\mathbf{S}_{ik})^{-1}\left(\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_{ik}\right)^{-1}\boldsymbol{\alpha}_k \\
&\quad \boldsymbol{\alpha}_k^*\left(\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_{ik}\right)^{-1}\left(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T}\right)^{-1}\boldsymbol{\alpha}_k \\
&= b_i(z_1)b_i(z_2)\sum_{k<i}\mathrm{E}_i\boldsymbol{\alpha}_k^*\left(\frac{1}{z_1}\mathbf{T}-\mathbf{S}_{ik}\right)^{-1}\left(\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_{ik}\right)^{-1}\boldsymbol{\alpha}_k \\
&\quad \boldsymbol{\alpha}_k^*\left(\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_{ik}\right)^{-1}\left(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T}\right)^{-1}\boldsymbol{\alpha}_k + O_p(n^{1/2}) \\
&= \frac{1}{n^2}b_i(z_1)b_i(z_2)\sum_{k<i}\mathrm{E}_i\mathrm{tr}\left(\frac{1}{z_1}\mathbf{T}-\mathbf{S}_{ik}\right)^{-1}\left(\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_{ik}\right)^{-1} \\
&\quad \mathrm{tr}\left(\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_{ik}\right)^{-1}\left(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T}\right)^{-1} + O_p(n^{1/2}) \\
&= \frac{i-1}{n^2}b_i(z_1)b_i(z_2)\mathrm{E}_i\mathrm{tr}\left(\frac{1}{z_1}\mathbf{T}-\mathbf{S}_{i}\right)^{-1}\left(\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_{i}\right)^{-1} \\
&\quad \mathrm{tr}\left(\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_{i}\right)^{-1}\left(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T}\right)^{-1} + O_p(n^{1/2}),
\end{aligned}
$$

where the third equality holds by (5.13) and (5.14). The above estimates imply (5.12). By (5.9)-(5.12), we obtain

$$
\begin{aligned}
&\mathrm{tr}\mathrm{E}_i(\frac{1}{z_1}\mathbf{T}-\mathbf{S}_i)^{-1}(\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_i)^{-1} \\
&\left[1-\frac{(i-1)}{n^2}b_i(z_1)b_i(z_2)\mathrm{tr}(\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_i)^{-1}(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T})^{-1}\right] \\
&= -\mathrm{tr}\left[(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T})^{-1}\mathrm{E}_i(\frac{1}{z_2}\mathbf{T}-\mathbf{S}_i)^{-1}\right] + O_p(n^{1/2}).
\end{aligned}
$$

Furthermore we have

$$
\begin{aligned}
&\mathrm{tr}[\mathrm{E}_i(\frac{1}{z_1}\mathbf{T}-\mathbf{S}_i)^{-1}](\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_i)^{-1} \\
&\left[1-\frac{(i-1)}{n^2}b_i(z_1)b_i(z_2)\mathrm{tr}(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T})^{-1}(\frac{n-1}{n}b_i(z_2)\mathbf{I}-\frac{1}{z_2}\mathbf{T})^{-1}\right] \\
&= \mathrm{tr}\left[(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T})^{-1}(\frac{n-1}{n}b_i(z_2)\mathbf{I}-\frac{1}{z_2}\mathbf{T})^{-1}\right] + O_p(n^{1/2})
\end{aligned}
$$

because

$$
\begin{aligned}
\mathrm{tr}\mathbf{A}(z_2)(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T})^{-1} &= O_p(n^{1/2}), \\
\mathrm{tr}\mathbf{B}(z_2)(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T})^{-1} &= O_p(n^{1/2}), \\
\mathrm{tr}\mathbf{C}(z_2)(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T})^{-1} &= O_p(n^{1/2}).
\end{aligned}
$$

That is,

$$\begin{aligned}
&\mathrm{tr}\left[\mathrm{E}_i(\frac{1}{z_1}\mathbf{T}-\mathbf{S}_i)^{-1}\mathrm{E}_i(\frac{1}{z_2}\mathbf{T}-\mathbf{S}_i)^{-1}\right]\\
&=\frac{\mathrm{tr}\left[(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T})^{-1}(\frac{n-1}{n}b_i(z_2)\mathbf{I}-\frac{1}{z_2}\mathbf{T})^{-1}\right]+O_p(n^{1/2})}{1-\frac{(i-1)}{n^2}b_i(z_1)b_i(z_2)\mathrm{tr}(\frac{n-1}{n}b_i(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T})^{-1}(\frac{n-1}{n}b_i(z_2)\mathbf{I}-\frac{1}{z_2}\mathbf{T})^{-1}}.
\end{aligned}$$

Step 3.3: Obtaining $\mathrm{Cov}(M_1(z_1), M_1(z_2))$, the limit of

$$\begin{aligned}
&\sum_{i=1}^{n}\mathrm{E}_{i-1}Y_i(z_1)Y_i(z_2)\\
&=\frac{\partial^2}{\partial z_1\partial z_2}\sum_{i=1}^{n}\mathrm{E}_{i-1}[\mathrm{E}_i\bar{\beta}_i(z_1)\hat{\gamma}_i(z_1)\mathrm{E}_i\bar{\beta}_i(z_2)\hat{\gamma}_i(z_2)]\\
&=\frac{\partial^2}{\partial z_1\partial z_2}\frac{\kappa}{n^2}\sum_{i=1}^{n}b_i(z_1)b_i(z_2)\mathrm{tr}\left[\mathrm{E}_i(\frac{1}{z_1}\mathbf{T}-\mathbf{S}_i)^{-1}\mathrm{E}_i(\frac{1}{z_2}\mathbf{T}-\mathbf{S}_i)^{-1}\right]+o_p(1).
\end{aligned}$$

By Lemma 9.1 of [2] and (3.10), we have

$$\begin{aligned}
&\max_{i\le n}|b_i(z)-b(z)|=o(1),\quad |b_i(z)-\mathrm{E}\beta_i(z)|\le Kn^{-1/2},\\
&\frac{1}{pz}\sum_{i=1}^{n}\mathrm{E}(-1+\beta_i(z))=\mathrm{E}s_n(z),\quad |\mathrm{E}s_n(z)-s_n^0(z)|\le Kn^{-1},\\
&\mathrm{E}\beta_i(z)=y_nz\mathrm{E}s_n(z)+1.
\end{aligned}$$

So we have

$$\begin{aligned}
&\mathrm{tr}[\mathrm{E}_i(\frac{1}{z_1}\mathbf{T}-\mathbf{S}_i)^{-1}](\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_i)^{-1}\\
&=\frac{\mathrm{tr}\left[b(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T})^{-1}(b(z_2)\mathbf{I}-\frac{1}{z_2}\mathbf{T})^{-1}\right]+o_p(1)}{1-\frac{(i-1)}{n^2}b(z_1)b(z_2)\mathrm{tr}(b(z_1)\mathbf{I}-\frac{1}{z_1}\mathbf{T})^{-1}(b(z_2)\mathbf{I}-\frac{1}{z_2}\mathbf{T})^{-1}}.
\end{aligned}$$

Thus we have

$$\begin{aligned}
&b(z_1)b(z_2)\mathrm{tr}[\mathrm{E}_i(\frac{1}{z_1}\mathbf{T}-\mathbf{S}_i)^{-1}](\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_i)^{-1}\\
&=\frac{p\int\frac{b(z_1)b(z_2)}{(b(z_1)-\frac{1}{z_1}t)(b(z_2)-\frac{1}{z_2}t)}dH_n(t)+O_p(n^{1/2})}{1-\frac{(i-1)}{n}y_n\int\frac{b(z_1)b(z_2)}{(b(z_1)-\frac{1}{z_1}t)(b(z_2)-\frac{1}{z_2}t)}dH_n(t)}.
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
&\frac{b(z_1)b(z_2)}{n^2}\sum_{i=1}^{n}\mathrm{tr}[\mathrm{E}_i(\frac{1}{z_1}\mathbf{T}-\mathbf{S}_i)^{-1}](\frac{1}{z_2}\mathbf{T}-\breve{\mathbf{S}}_i)^{-1}\\
&\to\int_0^1\frac{y\int\frac{b(z_1)b(z_2)}{(b(z_1)-\frac{1}{z_1}t)(b(z_2)-\frac{1}{z_2}t)}dH(t)}{1-xy\int\frac{b(z_1)b(z_2)}{(b(z_1)-\frac{1}{z_1}t)(b(z_2)-\frac{1}{z_2}t)}dH(t)}dx
\end{aligned}$$

$$= a(z_1,z_2)\int_0^1 \frac{1}{1-xa(z_1,z_2)}dx = \int_0^{a(z_1,z_2)} \frac{1}{1-z}dz$$

where

$$\begin{aligned}
a(z_1,z_2) &= y\int \frac{b(z_1)b(z_2)}{(\frac{t}{z_1}-b(z_1))(\frac{t}{z_2}-b(z_2))}dH(t) \\
&= \frac{y}{z_1z_2}\int \frac{\underline{\tilde{m}}(\frac{1}{z_1})\underline{\tilde{m}}(\frac{1}{z_2})}{(\frac{t}{z_1}+\frac{\underline{\tilde{m}}(\frac{1}{z_1})}{z_1})(\frac{t}{z_2}+\frac{\underline{\tilde{m}}(\frac{1}{z_2})}{z_2})}dH(t) \quad (\text{define } \underline{\tilde{m}}(z^{-1}) \overset{\triangle}{=} -zb(z)) \\
&= y\int \frac{\underline{\tilde{m}}(\frac{1}{z_1})\underline{\tilde{m}}(\frac{1}{z_2})}{(t+\underline{\tilde{m}}(\frac{1}{z_1}))(t+\underline{\tilde{m}}(\frac{1}{z_2}))}dH(t) \\
&= \frac{\underline{\tilde{m}}(\frac{1}{z_1})\underline{\tilde{m}}(\frac{1}{z_2})}{\underline{\tilde{m}}(\frac{1}{z_2})-\underline{\tilde{m}}(\frac{1}{z_1})}y\left(\int \frac{1}{t+\underline{\tilde{m}}(\frac{1}{z_1})}dH(t) - \int \frac{1}{t+\underline{\tilde{m}}(\frac{1}{z_2})}dH(t)\right) \\
&= \frac{\underline{\tilde{m}}(\frac{1}{z_1})\underline{\tilde{m}}(\frac{1}{z_2})}{\underline{\tilde{m}}(\frac{1}{z_2})-\underline{\tilde{m}}(\frac{1}{z_1})}\left(\frac{1}{z_1}-\frac{1}{z_2}+\frac{\underline{\tilde{m}}(\frac{1}{z_2})-\underline{\tilde{m}}(\frac{1}{z_1})}{\underline{\tilde{m}}(\frac{1}{z_1})\underline{\tilde{m}}(\frac{1}{z_2})}\right) \\
&= 1+\frac{\underline{\tilde{m}}(\frac{1}{z_1})\underline{\tilde{m}}(\frac{1}{z_2})}{\underline{\tilde{m}}(\frac{1}{z_2})-\underline{\tilde{m}}(\frac{1}{z_1})}\left(\frac{1}{z_1}-\frac{1}{z_2}\right),
\end{aligned} \tag{5.15}$$

and

$$\begin{aligned}
\int \frac{1}{t+\underline{\tilde{m}}(z^{-1})}dH(t) &= z^{-1}\int \frac{1}{\frac{t}{z}+z^{-1}\underline{\tilde{m}}(z^{-1})}dH(t) \\
&= z^{-1}\int \frac{1}{\frac{t}{z}-b(z)}dH(t) = \frac{\tilde{m}(z)}{z} \\
&= \frac{1}{y}\left(z^{-1}+\frac{1}{\underline{\tilde{m}}(z^{-1})}\right)
\end{aligned} \tag{5.16}$$

where the second equality uses (4.4) and (3.10) and the last equality uses (3.10). Notice that by Lemma 7.1, $a(z_1,z_2)<1$. We have

$$\frac{\partial^2}{\partial z_1\partial z_2}\left(\int_0^{a(z_1,z_2)} \frac{1}{1-z}dz\right) = \frac{\partial}{\partial z_2}\left(\frac{\partial a(z_1,z_2)/\partial z_1}{1-a(z_1,z_2)}\right)$$

and

$$\begin{aligned}
\partial a(z_1,z_2)/\partial z_1 &= \frac{\partial}{\partial z_1}\left(\frac{\underline{\tilde{m}}(\frac{1}{z_1})\underline{\tilde{m}}(\frac{1}{z_2})}{\underline{\tilde{m}}(\frac{1}{z_2})-\underline{\tilde{m}}(\frac{1}{z_1})}\left(\frac{1}{z_1}-\frac{1}{z_2}\right)\right) \\
&= -\frac{(\underline{\tilde{m}}(\frac{1}{z_1}))'\underline{\tilde{m}}(\frac{1}{z_2})(\underline{\tilde{m}}(\frac{1}{z_2})-\underline{\tilde{m}}(\frac{1}{z_1}))+\underline{\tilde{m}}(\frac{1}{z_1}))'\underline{\tilde{m}}(\frac{1}{z_1}))\underline{\tilde{m}}(\frac{1}{z_2})}{(\underline{\tilde{m}}(\frac{1}{z_2})-\underline{\tilde{m}}(\frac{1}{z_1}))^2}\left(\frac{1}{z_1}-\frac{1}{z_2}\right) \\
&\quad +\frac{\underline{\tilde{m}}(\frac{1}{z_1})\underline{\tilde{m}}(\frac{1}{z_2})}{\underline{\tilde{m}}(\frac{1}{z_2})-\underline{\tilde{m}}(\frac{1}{z_1})}\left(\frac{-1}{z_1^2}\right)
\end{aligned}$$

$$= -\frac{(\underline{\tilde{m}}(\frac{1}{z_1}))'\underline{\tilde{m}}^2(\frac{1}{z_2})}{(\underline{\tilde{m}}(\frac{1}{z_2}) - \underline{\tilde{m}}(\frac{1}{z_1}))^2}\left(\frac{1}{z_1} - \frac{1}{z_2}\right) + \frac{\underline{\tilde{m}}(\frac{1}{z_1})\underline{\tilde{m}}(\frac{1}{z_2})}{\underline{\tilde{m}}(\frac{1}{z_2}) - \underline{\tilde{m}}(\frac{1}{z_1})}\left(\frac{-1}{z_1^2}\right).$$

So we obtain

$$\frac{\partial a(z_1, z_2)/\partial z_1}{1 - a(z_1, z_2)}$$
$$= \left[-\frac{(\underline{\tilde{m}}(\frac{1}{z_1}))'\underline{\tilde{m}}^2(\frac{1}{z_2})}{(\underline{\tilde{m}}(\frac{1}{z_2}) - \underline{\tilde{m}}(\frac{1}{z_1}))^2}\left(\frac{1}{z_1} - \frac{1}{z_2}\right) + \frac{\underline{\tilde{m}}(\frac{1}{z_1})\underline{\tilde{m}}(\frac{1}{z_2})}{\underline{\tilde{m}}(\frac{1}{z_2}) - \underline{\tilde{m}}(\frac{1}{z_1})}\left(\frac{-1}{z_1^2}\right)\right]\frac{\underline{\tilde{m}}(\frac{1}{z_2}) - \underline{\tilde{m}}(\frac{1}{z_1})}{\underline{\tilde{m}}(\frac{1}{z_1})\underline{\tilde{m}}(\frac{1}{z_2})\left(\frac{1}{z_1} - \frac{1}{z_2}\right)}$$
$$= -\frac{(\underline{\tilde{m}}(\frac{1}{z_1}))'\underline{m}(\frac{1}{z_2})}{m(\frac{1}{z_1})(\underline{\tilde{m}}(\frac{1}{z_2}) - \underline{\tilde{m}}(\frac{1}{z_1}))} - \frac{1/z_1^2}{1/z_1 - 1/z_2},$$

and

$$\begin{aligned}\frac{\partial}{\partial z_2}\left(\frac{\partial a(z_1, z_2)/\partial z_1}{1 - a(z_1, z_2)}\right) &= \frac{(\underline{\tilde{m}}(\frac{1}{z_1}))'(\underline{\tilde{m}}(\frac{1}{z_2}))'}{(\underline{\tilde{m}}(\frac{1}{z_2}) - \underline{\tilde{m}}(\frac{1}{z_1}))^2} - \frac{1}{z_1^2 z_2^2}\frac{1}{(1/z_1 - 1/z_2)^2}\\
&= \frac{(\underline{\tilde{m}}(\frac{1}{z_1}))'(\underline{\tilde{m}}(\frac{1}{z_2}))'}{(\underline{\tilde{m}}(\frac{1}{z_2}) - \underline{\tilde{m}}(\frac{1}{z_1}))^2} - \frac{1}{(z_1 - z_2)^2}\\
&= \frac{[z_1 b(z_1)]'[z_2 b(z_2)]'}{[z_1 b(z_1) - z_2 b(z_2)]^2} - \frac{1}{(z_1 - z_2)^2}.\end{aligned}$$

That is, the limit of

$$\sum_{i=1}^{n} \mathrm{E}_{i-1} Y_i(z_1) Y_i(z_2) = \frac{\partial^2}{\partial z_1 \partial z_2}\frac{\kappa}{n^2}\sum_{i=1}^{n} b_i(z_1) b_i(z_2) \mathrm{tr}\left[\mathrm{E}_i(\frac{1}{z_1}\mathbf{T} - \mathbf{S}_i)^{-1}\mathrm{E}_i(\frac{1}{z_2}\mathbf{T} - \mathbf{S}_i)^{-1}\right]$$

is

$$\kappa\frac{\partial^2}{\partial z_1 \partial z_2}\left(\int_0^{a(z_1,z_2)} \frac{1}{1 - z}dz\right) = \kappa\frac{[z_1 b(z_1)]'[z_2 b(z_2)]'}{[z_1 b(z_1) - z_2 b(z_2)]^2} - \frac{\kappa}{(z_1 - z_2)^2}.$$

So we have

$$\mathrm{Cov}(M_1(z_1), M_1(z_2)) = \kappa\left(\frac{[z_1 b(z_1)]'[z_2 b(z_2)]'}{[z_1 b(z_1) - z_2 b(z_2)]^2} - \frac{1}{(z_1 - z_2)^2}\right). \tag{5.17}$$

Then the proof of Lemma 5.1 is completed. ■

5.2. *Tightness of $M_n^1(z)$*

Lemma 5.2. *Under Assumptions 1*, 2* and 3, the sequence of random functions $\{M_n^1(z)\}$ is tight for $z \in \mathcal{C}_n := \mathcal{C} \cap \{|\Im(z)| > n^{-2}\}$.*

Proof. We proceed to prove the tightness of $\{M_n^1(z)\}$ for $z \in \mathcal{C}$. We will use Theorem 12.3 in [9]. Since $\{Y_j(z_i)\}$ is the martingale difference sequence and $E|Y_j(z_i)|^2 = O(n^{-1})$, $\mathrm{E}|\sum_{i=1}^{r}\sum_{j=1}^{n} a_i Y_j(z_i)|^2$ is bounded by $r\sum_{i=1}^{r}\sum_{j=1}^{n}|a_i|^2 E|Y_j(z_i)|^2$ whenever $\{\alpha_1, \ldots, \alpha_r\}$ are constants: so Condition (i) in this Theorem 12.3 in [9] is satisfied. We will verify Condition (ii) of the same theorem by proving the moment condition (12.51) in [9], that is

$$\sup_{n;z_1,z_2\in\mathcal{C}_n}\frac{\mathrm{E}|M_n^1(z_1)-M_n^1(z_2))|^2}{|z_1-z_2|^2}<\infty.$$

Step 1: We simplify $\frac{M_n^1(z_1)-M_n^1(z_2)}{z_1-z_2}$ as

$$\begin{aligned}\frac{M_n^1(z_1)-M_n^1(z_2)}{z_1-z_2}&=\sum_{i=1}^n(\mathrm{E}_i-\mathrm{E}_{i-1})(\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2)\boldsymbol{\alpha}_i)^2\beta_i(z_1)\beta_i(z_2)\\&+\sum_{i=1}^n(\mathrm{E}_i-\mathrm{E}_{i-1})\beta_i(z_1)\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{F}_i^{-1}(z_2)\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_1)\boldsymbol{\alpha}_i\\&+\sum_{i=1}^n(\mathrm{E}_i-\mathrm{E}_{i-1})\beta_i(z_2)\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{F}_i^{-1}(z_1)\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2)\boldsymbol{\alpha}_i\end{aligned}$$

where $\mathbf{F}_i^{-1}(z)=(\mathbf{S}_i^{-1}\mathbf{T}-z\mathbf{I})^{-1}$. First we have by the martingale decomposition

$$\begin{aligned}M_n^1(z_1)-M_n^1(z_2)&=p(s_n(z_1)-s_n(z_2))-p\mathrm{E}(s_n(z_1)-s_n(z_2))\\&=\sum_{i=1}^n(\mathrm{E}_i-\mathrm{E}_{i-1})s_n(z_1)-s_n(z_2),\\s_n(z_1)-s_n(z_2)&=\frac{1}{p}\mathrm{tr}\left[(\mathbf{S}^{-1}\mathbf{T}-z_1\mathbf{I})^{-1}-(\mathbf{S}^{-1}\mathbf{T}-z_2\mathbf{I})^{-1}\right]\\&=\frac{z_1-z_2}{p}\mathrm{tr}(\mathbf{S}(\mathbf{T}-z_1\mathbf{S})^{-1}\mathbf{S}(\mathbf{T}-z_2\mathbf{S})^{-1}).\end{aligned}$$

Thus we have

$$\begin{aligned}\frac{M_n^1(z_1)-M_n^1(z_2)}{z_1-z_2}&=p\frac{s_n(z_1)-s_n(z_2)-\mathrm{E}(s_n(z_1)-s_n(z_2))}{z_1-z_2}\\&=\sum_{i=1}^n(\mathrm{E}_i-\mathrm{E}_{i-1})\mathrm{tr}(\mathbf{S}(\mathbf{T}-z_1\mathbf{S})^{-1}\mathbf{S}(\mathbf{T}-z_2\mathbf{S})^{-1}).\end{aligned}$$

Furthermore $(\mathbf{S}^{-1}\mathbf{T}-z\mathbf{I})^{-1}=\mathbf{S}(\mathbf{T}-z\mathbf{S})^{-1}$ can be decomposed as

$$\begin{aligned}\mathbf{S}(\mathbf{T}-z\mathbf{S})^{-1}&=(\mathbf{S}_i+\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*)(\mathbf{D}_i-z\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*)^{-1}\\&=\mathbf{S}_i\mathbf{D}_i^{-1}+\frac{z\mathbf{S}_i\mathbf{D}_i^{-1}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}}{1-z\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\boldsymbol{\alpha}_i}+\frac{\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}}{1-z\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\boldsymbol{\alpha}_i}\\&=\mathbf{S}_i\mathbf{D}_i^{-1}+\mathbf{T}\mathbf{D}_i^{-1}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\beta_i(z),\end{aligned}$$

where $\beta_i(z)=\frac{1}{1-z\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\boldsymbol{\alpha}_i}$. That is,

$$(\mathbf{S}^{-1}\mathbf{T}-z\mathbf{I})^{-1}-(\mathbf{S}_i^{-1}\mathbf{T}-z\mathbf{I})^{-1}=\mathbf{T}\mathbf{D}_i^{-1}(z)\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}(z)\beta_i(z).$$

So we obtain

$$\begin{aligned}&\mathrm{tr}(\mathbf{S}^{-1}\mathbf{T}-z_1\mathbf{I})^{-1}(\mathbf{S}^{-1}\mathbf{T}-z_2\mathbf{I})^{-1}-\mathrm{tr}(\mathbf{S}_i^{-1}\mathbf{T}-z_1\mathbf{I})^{-1}(\mathbf{S}_i^{-1}\mathbf{T}-z_2\mathbf{I})^{-1}\\&=|\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}(z_1)\mathbf{T}\mathbf{D}_i^{-1}(z_2)\boldsymbol{\alpha}_i|^2\beta_i(z_1)\beta_i(z_2)\end{aligned}$$

$$+\beta_i(z_1)\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}(z_1)\mathbf{S}_i\mathbf{D}_i^{-1}(z_2)\mathbf{T}\mathbf{D}_i^{-1}(z_1)\boldsymbol{\alpha}_i$$
$$+\beta_i(z_2)\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}(z_2)\mathbf{S}_i\mathbf{D}_i^{-1}(z_1)\mathbf{T}\mathbf{D}_i^{-1}(z_2)\boldsymbol{\alpha}_i.$$

Thus we have

$$\frac{M_n^1(z_1)-M_n^1(z_2)}{z_1-z_2}$$
$$=\sum_{i=1}^n(\mathrm{E}_i-\mathrm{E}_{i-1})\mathrm{tr}(\mathbf{S}(\mathbf{T}-z_1\mathbf{S})^{-1}\mathbf{S}(\mathbf{T}-z_2\mathbf{S})^{-1})$$
$$=\sum_{i=1}^n(\mathrm{E}_i-\mathrm{E}_{i-1})|\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}(z_1)\mathbf{T}\mathbf{D}_i^{-1}(z_2)\boldsymbol{\alpha}_i|^2\beta_i(z_1)\beta_i(z_2) \tag{5.18}$$
$$+\sum_{i=1}^n(\mathrm{E}_i-\mathrm{E}_{i-1})\beta_i(z_1)\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}(z_1)\mathbf{S}_i\mathbf{D}_i^{-1}(z_2)\mathbf{T}\mathbf{D}_i^{-1}(z_1)\boldsymbol{\alpha}_i \tag{5.19}$$
$$+\sum_{i=1}^n(\mathrm{E}_i-\mathrm{E}_{i-1})\beta_i(z_2)\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}(z_2)\mathbf{S}_i\mathbf{D}_i^{-1}(z_1)\mathbf{T}\mathbf{D}_i^{-1}(z_2)\boldsymbol{\alpha}_i\ . \tag{5.20}$$

Step 2: Our goal is to show that the absolute second moment $\mathrm{E}\left|\frac{M_n^1(z_1)-M_n^1(z_2)}{z_1-z_2}\right|^2$ is bounded.

Step 2.1: We want to prove that the absolute second moment of (5.19) is uniformly bounded, i.e., $\mathrm{E}\left|\sum_{i=1}^n(\mathrm{E}_i-\mathrm{E}_{i-1})\beta_i(z_1)\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}(z_1)\mathbf{S}_i\mathbf{D}_i^{-1}(z_2)\mathbf{T}\mathbf{D}_i^{-1}(z_1)\boldsymbol{\alpha}_i\right|^2\le C.$

First (5.19) can be decomposed as

$$\sum_{i=1}^n(\mathrm{E}_i-\mathrm{E}_{i-1})\beta_i(z_1)\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}(z_1)\mathbf{S}_i\mathbf{D}^{-1}(z_2)\mathbf{T}\mathbf{D}_i^{-1}(z_1)\boldsymbol{\alpha}_i$$
$$=\sum_{i=1}^n(\mathrm{E}_i-\mathrm{E}_{i-1})\beta_i(z_1)(\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}(z_1)\mathbf{S}_i\mathbf{D}^{-1}(z_2)\mathbf{T}\mathbf{D}_i^{-1}(z_1)\boldsymbol{\alpha}_i-\frac{1}{n}\mathrm{tr}(\mathbf{D}_i^{-1}(z_1)\mathbf{S}_i\mathbf{D}^{-1}(z_2)\mathbf{T}\mathbf{D}_i^{-1}(z_1)))$$
$$-\frac{1}{n}\sum_{i=1}^n(\mathrm{E}_i-\mathrm{E}_{i-1})\beta_i(z_1)b_i(z_1)\varepsilon_i(z_1)\mathrm{tr}(\mathbf{D}_i^{-1}(z_1)\mathbf{S}_i\mathbf{D}^{-1}(z_2)\mathbf{T}\mathbf{D}_i^{-1}(z_1))$$
$$=W_1-W_2,$$

where $\varepsilon_i(z)=\boldsymbol{\alpha}_i^*(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1}\boldsymbol{\alpha}_i-\frac{1}{n}\mathrm{E}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1}$. Then by Lemma (9.9.6) of [2], we have

$$\mathrm{E}|W_1|^2$$
$$=\sum_{i=1}^n\mathrm{E}|\beta_i(z_1)|^2\left|\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}(z_1)\mathbf{S}_i\mathbf{D}^{-1}(z_2)\mathbf{T}\mathbf{D}_i^{-1}(z_1)\boldsymbol{\alpha}_i-\frac{1}{n}\mathrm{tr}(\mathbf{D}_i^{-1}(z_1)\mathbf{S}_i\mathbf{D}^{-1}(z_2)\mathbf{T}\mathbf{D}_i^{-1}(z_1))\right|^2$$
$$\le K.$$

Moreover we have

$$\mathrm{E}|W_2|^2=\sum_{i=1}^n\mathrm{E}|(\mathrm{E}_i-\mathrm{E}_{i-1})\beta_i(z_1)b_i(z_1)\varepsilon_i(z_1)\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{F}_i^{-1}(z_2)\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_1)\boldsymbol{\alpha}_i|^2$$

$$
\begin{aligned}
&= n\mathrm{E}|(\mathrm{E}_i - \mathrm{E}_{i-1})\beta_i(z_1)b_i(z_1)\varepsilon_i(z_1)\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{F}_i^{-1}(z_2)\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_1)\boldsymbol{\alpha}_i|^2 \\
&\leq 2nb_i^2(z_1)\mathrm{E}\beta_i^2(z_1)\varepsilon_i^2(z_1)\left(\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{F}_i^{-1}(z_2)\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_1)\boldsymbol{\alpha}_i\right)^2 \leq K
\end{aligned}
$$

where

$$\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{F}_i^{-1}(z_2)\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_1)\boldsymbol{\alpha}_i \leq K|\boldsymbol{\alpha}_i|^2 + n^t I(\|\mathbf{S}_i\| \geq \eta_r \ \text{ or } \ \lambda_{\min}^{\mathbf{S}_i} \leq \eta_l).$$

So we obtain

$$\mathrm{E}\left|\sum_{i=1}^n (\mathrm{E}_i - \mathrm{E}_{i-1})\beta_i(z_1)\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{F}_i^{-1}(z_2)\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_1)\boldsymbol{\alpha}_i\right|^2 \leq K$$

where K is a constant. Similarly, it can be proved that the absolute second moment of (5.20) is uniformly bounded.

Step 2.2: Now we prove that the absolute second moment of (5.18) is uniformly bounded, that is,

$$\mathrm{E}\left|\sum_{i=1}^n (\mathrm{E}_i - \mathrm{E}_{i-1})(\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2)\boldsymbol{\alpha}_i)^2\beta_i(z_1)\beta_i(z_2)\right|^2 \leq \text{a constant}$$

where the constant depends on $\mathcal{C}$. First (5.18) can be decomposed as

$$
\begin{aligned}
&\sum_{i=1}^n (\mathrm{E}_i - \mathrm{E}_{i-1})(\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2)\boldsymbol{\alpha}_i)^2\beta_i(z_1)\beta_i(z_2) \\
&= \sum_{i=1}^n (\mathrm{E}_i - \mathrm{E}_{i-1})\left(\left[\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2)\boldsymbol{\alpha}_i\right]^2 - \left[\frac{1}{n}\mathrm{tr}(\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2)\right]^2\right)b_i(z_1)b_i(z_2) \\
&\quad - \sum_{i=1}^n (\mathrm{E}_i - \mathrm{E}_{i-1})(\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2)\boldsymbol{\alpha}_i)^2\beta_i(z_1)\beta_i(z_2)b_i(z_2)\varepsilon_i(z_2) \\
&\quad - \sum_{i=1}^n (\mathrm{E}_i - \mathrm{E}_{i-1})(\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2)\boldsymbol{\alpha}_i)^2 b_i(z_1)b_i(z_2)\beta_i(z_1)\varepsilon_i(z_1) \\
&= Z_1 - Z_2 - Z_3.
\end{aligned}
$$

By a method similar to the one employed for W_2, we prove that the second moments of Z_2 and Z_3 are uniformly bounded. For Z_1, we have

$$
\begin{aligned}
&\mathrm{E}|Z_1|^2 \\
&\leq K\sum_{i=1}^n \mathrm{E}\left|\left[\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2)\boldsymbol{\alpha}_i\right]^2 - \left[\frac{1}{n}\mathrm{tr}(\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2))\right]^2\right|^2 \\
&\leq 2K\sum_{i=1}^n \mathrm{E}\left|\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2)\boldsymbol{\alpha}_i - \frac{1}{n}\mathrm{tr}(\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2))\right|^4 \\
&\quad + \frac{K}{n}\sum_{i=1}^n \mathrm{E}\left|\left(\boldsymbol{\alpha}_i^*\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2)\boldsymbol{\alpha}_i - \frac{1}{n}\mathrm{tr}(\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2))\right)|\mathbf{D}_i^{-1}\mathbf{S}_i^{-1}\mathbf{T}\mathbf{F}_i^{-1}(z_2)|\right|^2
\end{aligned}
$$

$\leq K.$

Because the absolute second moments of (5.18)-(5.20) are uniformly bounded, then we have proved that

$$\sup_{n;z_1,z_2\in\mathcal{C}+}\frac{\mathrm{E}|M_n^1(z_1)-M_n^1(z_2)|^2}{|z_1-z_2|^2}\leq K.$$

Then the proof of Lemma 5.2 is complete. ■

5.3. *Uniform convergence of $M_n^2(z)=p(\mathrm{E}s_n(z)-s_n^0(z))$ for $z\in\mathcal{C}_n$*

Lemma 5.3. *We have*

$$\sup_{z\in\mathcal{C}_n}\left|M_n^2(z)-(\kappa-1)\frac{yb^3(z)}{z^2g^2(z)}\int\frac{tdH(t)}{(t/z-b(z))^3}\right|\to 0\quad as\ n\to\infty.$$

Proof. Before the core of the proof, we first establish the following two facts:

$$\sup_{z\in\mathcal{C}_n}|\mathrm{E}\beta_1(z)-b_1(z)|\to 0\quad \text{as } n\to\infty; \tag{5.21}$$

and there exists a positive number $K>0$ satisfies

$$\sup_{n,z\in\mathcal{C}_n}\|(\mathbf{T}-z\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\|<K. \tag{5.22}$$

First, (5.21) can be obtained by a method similar to the one used on page 287 of [2]. Moreover, (5.22) is equivalent to $\sup\limits_{n,z\in\mathcal{C}_u}\|[\mathbf{T}-z(1+y_nz\mathrm{E}s_n(z))\mathbf{I}]^{-1}\|<K$, a bound we now prove.

Let $x=x_l$ or x_r. Since x is outside the support of the LSD of $\mathbf{S}^{-1}\mathbf{T}$, by (3.13), for any t in the support of H, we have $t-zb(z)\neq 0$ where $b(z)=1+yzs(z)$. Choose any t_0 in the support of H. Since $s(z)$ is continuous on $\mathcal{C}^0=\{x+i\nu:\nu\in[0,\nu_0]\}$, there exist constants δ_1 and μ_0 satisfy

$$\inf_{z\in\mathcal{C}^0}|t_0-zb(z)|>\delta_1,\quad \sup_{z\in\mathcal{C}^0}|zb(z)|<\mu_0.$$

Using $H_n\xrightarrow{\mathcal{D}}H$, for all large n, there exists an eigenvalue λ of $\mathbf{T}$ such that $|\lambda-t_0|<\delta_1/(4\mu_0)$ and $\sup\limits_{z\in\mathcal{C}_l\cup\mathcal{C}_r}|\mathrm{E}s_n(z)-s(z)|<\delta_1/4$. Therefore, we have

$$\int\limits_{z\in\mathcal{C}_l\cup\mathcal{C}_r}|\lambda-z(1+y_nz\mathrm{E}s_n(z))|>\delta_1/2.$$

This leads to the announced bound and establishes the two facts in (5.21) and (5.22).

Now we develop the main parts of the proof in several steps.

Step 1: Show that

$$n(\mathrm{E}\underline{\tilde{s}}_n(z)-\underline{\tilde{s}}_n^0(z))$$
$$=\frac{\mathrm{E}\underline{\tilde{s}}_n(z)\underline{\tilde{s}}_n^0(z)nR_n}{1-\frac{y\int(t+\mathrm{E}\underline{\tilde{s}}_n(z))^{-1}(t+\underline{\tilde{s}}_n^0(z))^{-1}dH_n(t)}{\left(-z^{-1}+y\int(t+\mathrm{E}\underline{\tilde{s}}_n(z))^{-1}dH_n(t)-R_n\right)\left(-z^{-1}+y\int(t+\underline{\tilde{s}}_n^0(z))^{-1}dH_n(t)\right)}}$$

where $s_n^0(z)$ is the Stieltjes transform of F^{y_n,H_n},

$$s_n^0(z)=-z^{-1}-\frac{1}{z^2}\tilde{s}_n^0(z),\quad \underline{\tilde{s}}_n^0(z)=-(1-y)z+y\tilde{s}_n^0(z),$$
$$s_n(z)=-z^{-1}-\frac{1}{z^2}\tilde{s}_n(z),\quad \underline{\tilde{s}}_n(z)=-(1-y)z+y\tilde{s}_n(z),$$
$$R_n=-\frac{1}{\mathrm{E}\underline{\tilde{s}}_n(z)}\frac{y}{z}\left(\mathrm{E}\tilde{s}_n(z)+\int\frac{tdH_n(t)}{t/z+\mathrm{E}\underline{\tilde{s}}_n(z)/z}\right).$$

First we have

$$\frac{1}{p}\mathrm{tr}(\mathbf{S}^{-1}\mathbf{T}-z\mathbf{I})^{-1}=-z^{-1}-\frac{1}{z^2}\frac{1}{p}\mathrm{tr}\left(\mathbf{S}-\frac{\mathbf{T}}{z}\right)^{-1}\mathbf{T}$$

$$s_n(z)=\frac{1}{pz}\sum_{i=1}^n\frac{\boldsymbol{\alpha}_i^*(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1}\boldsymbol{\alpha}_i}{1-\boldsymbol{\alpha}_i^*(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1}\boldsymbol{\alpha}_i}=-\frac{1}{yz}+\frac{1}{yz}\frac{1}{n}\sum_{i=1}^n\frac{1}{1-\boldsymbol{\alpha}_i^*(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1}\boldsymbol{\alpha}_i}.$$

By Lemma 4.2, we have $s(z)=-z^{-1}-\frac{1}{z^2}\tilde{s}(z)$ where

$$\underline{\tilde{s}}(z)=\frac{-z}{1-y\tilde{m}(z)}=\frac{-z}{1-y\int\frac{1}{\frac{t}{z}-\frac{1}{1-y\tilde{m}(z)}}dH(t)}=\frac{1}{-z^{-1}+y\int\frac{1}{t+\underline{\tilde{s}}(z)}dH(t)},$$

$\tilde{m}(z)$ is the limit of $\frac{1}{p}\mathrm{tr}\left(z^{-1}\mathbf{T}-\mathbf{S}\right)^{-1}$ by Lemma 4.1, $\tilde{s}(z)$ is the limit of $\frac{1}{p}\mathrm{tr}\left(\mathbf{S}-\frac{\mathbf{T}}{z}\right)^{-1}\mathbf{T}$ and $\underline{\tilde{s}}(z)=-(1-y)z+y\tilde{s}(z)$. Then we have

$$\mathrm{E}\underline{\tilde{s}}_n(z)=\frac{1}{-z^{-1}+y\int\frac{1}{t+\mathrm{E}\underline{\tilde{s}}_n(z)}dH_n(t)-R_n}.$$

Thus we obtain

$$\mathrm{E}\underline{\tilde{s}}_n(z)-\underline{\tilde{s}}_n^0(z)=\frac{(\mathrm{E}\underline{\tilde{s}}_n(z)-\underline{\tilde{s}}_n^0(z))y\int\frac{1}{(t+\mathrm{E}\underline{\tilde{s}}_n(z))(t+\underline{\tilde{s}}_n^0(z))}dH_n(t)}{\left(-z^{-1}+y\int\frac{1}{t+\mathrm{E}\underline{\tilde{s}}_n(z)}dH_n(t)-R_n\right)\left(-z^{-1}+y\int\frac{1}{t+\underline{\tilde{s}}_n^0(z)}dH_n(t)\right)}$$
$$+\mathrm{E}\underline{\tilde{s}}_n(z)\underline{\tilde{s}}_n^0(z)R_n.$$

That is,

$$n(\mathrm{E}\underline{\tilde{s}}_n(z)-\underline{\tilde{s}}_n^0(z))$$
$$=\frac{\mathrm{E}\underline{\tilde{s}}_n(z)\underline{\tilde{s}}_n^0(z)nR_n}{1-\frac{y\int(t+\mathrm{E}\underline{\tilde{s}}_n(z))^{-1}(t+\underline{\tilde{s}}_n^0(z))^{-1}dH_n(t)}{\left(-z^{-1}+y\int(t+\mathrm{E}\underline{\tilde{s}}_n(z))^{-1}dH_n(t)-R_n\right)\left(-z^{-1}+y\int(t+\underline{\tilde{s}}_n^0(z))^{-1}dH_n(t)\right)}}.$$

Step 2: Show that

$$\frac{-z\mathrm{E}\underline{\tilde{s}}_n(z)}{y}R_n = -\frac{(\mathrm{E}\beta_1(z))^2}{y}\mathrm{E}[\boldsymbol{\alpha}_1^*(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\boldsymbol{\alpha}_1$$
$$-\frac{1}{n}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}]\varepsilon_j$$
$$+\frac{(\mathrm{E}\beta_1(z))^2}{yn^2}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-2}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}+o(\frac{1}{n})$$

where $o(\frac{1}{n})$ is uniform in $z \in \mathcal{C}_n$. First notice that

$$(z^{-1}\mathbf{T}-\mathbf{S}) = (z^{-1}\mathbf{T}+\frac{\mathrm{E}\underline{\tilde{s}}_n(z)}{z}\mathbf{I}) - \frac{\mathrm{E}\underline{\tilde{s}}_n(z)}{z}\mathbf{I} - \sum_{i=1}^n \boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*.$$

Then we have

$$(z^{-1}\mathbf{T}-\mathbf{S})^{-1}\mathbf{T}$$
$$= (z^{-1}\mathbf{T}-\mathrm{E}\beta_i(z)\mathbf{I})^{-1}\mathbf{T} + (z^{-1}\mathbf{T}-\mathrm{E}\beta_i(z)\mathbf{I})^{-1}(\sum_{i=1}^n \boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^* - \mathrm{E}\beta_i(z)\mathbf{T})(z^{-1}\mathbf{T}-\mathbf{S})^{-1}\mathbf{T}$$
$$= (z^{-1}\mathbf{T}-\mathrm{E}\beta_i(z)\mathbf{I})^{-1}\mathbf{T} - \mathrm{E}\beta_i(z)(z^{-1}\mathbf{T}-\mathrm{E}\beta_i(z)\mathbf{I})^{-1}(z^{-1}\mathbf{T}-\mathbf{S})^{-1}\mathbf{T}$$
$$+\sum_{i=1}^n (z^{-1}\mathbf{T}-\mathrm{E}\beta_i(z)\mathbf{I})^{-1}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1}\mathbf{T}\beta_i(z).$$

Taking trace and expectation on both sides and dividing by p, we get

$$\frac{1}{p}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathbf{S})^{-1}\mathbf{T}$$
$$= \frac{1}{p}\mathrm{tr}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\mathbf{T} - \mathrm{E}\beta_1(z)\frac{1}{p}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}(z^{-1}\mathbf{T}-\mathbf{S})^{-1}\mathbf{T}$$
$$+\frac{1}{p}\mathrm{Etr}\sum_{i=1}^n (z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^*(z^{-1}\mathbf{T}-\mathbf{S}_i)^{-1}\mathbf{T}\beta_i(z)$$
$$= \frac{1}{p}\mathrm{tr}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\mathbf{T} - \mathrm{E}\beta_1(z)\frac{1}{p}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}(z^{-1}\mathbf{T}-\mathbf{S})^{-1}\mathbf{T}$$
$$+\frac{1}{y_n}\mathrm{E}\beta_1(z)\boldsymbol{\alpha}_1^*(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\boldsymbol{\alpha}_1$$
$$= \frac{1}{p}\mathrm{tr}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\mathbf{T} + \frac{1}{y_n}\mathrm{E}\beta_1(z)[\boldsymbol{\alpha}_1^*(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\boldsymbol{\alpha}_1$$
$$-\frac{1}{n}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}]$$
$$+[-\frac{1}{y_n}\mathrm{E}\beta_1(z)\frac{1}{n}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}(z^{-1}\mathbf{T}-\mathbf{S})^{-1}\mathbf{T}$$
$$+\frac{1}{y_n}\mathrm{E}\beta_1(z)\frac{1}{n}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}]$$
$$= \frac{1}{p}\mathrm{tr}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\mathbf{T} + \frac{1}{y_n}\mathrm{E}\beta_1(z)[\boldsymbol{\alpha}_1^*(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\boldsymbol{\alpha}_1$$

$$-\frac{1}{n}\mathrm{E}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}]$$
$$-\frac{(\mathrm{E}\beta_1(z))^2}{y}\frac{1}{n^2}\mathrm{E}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}+o(\frac{1}{n}).$$

So we obtain

$$\begin{aligned}
\frac{-z\mathrm{E}\underline{\tilde{s}}_n(z)}{y}R_n &= \int\frac{tdH_n(t)}{\frac{t}{z}-\mathrm{E}\beta_1(z)}-\mathrm{E}\tilde{s}_n(z)\\
&=\frac{1}{p}\mathrm{tr}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\mathbf{T}-\frac{1}{p}\mathrm{E}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S})^{-1}\mathbf{T}\\
&=-\frac{1}{y_n}\mathrm{E}\beta_1(z)[\boldsymbol{\alpha}_1^*(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\boldsymbol{\alpha}_1\\
&\quad-\frac{1}{n}\mathrm{E}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}]\\
&\quad+\frac{(\mathrm{E}\beta_1(z))^2}{y}\frac{1}{n^2}\mathrm{E}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-2}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}+o(\frac{1}{n})\\
&=-\frac{1}{y_n}\mathrm{E}\bar{\beta}_1^2(z)[\boldsymbol{\alpha}_1^*(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\boldsymbol{\alpha}_1\\
&\quad-\frac{1}{n}\mathrm{E}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}]\varepsilon_j\\
&\quad+\frac{(\mathrm{E}\beta_1(z))^2}{y}\frac{1}{n^2}\mathrm{E}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-2}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}+o(\frac{1}{n})\\
&=-\frac{(\mathrm{E}\beta_1(z))^2}{y}\mathrm{E}[\boldsymbol{\alpha}_1^*(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\boldsymbol{\alpha}_1\\
&\quad-\frac{1}{n}\mathrm{E}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}]\varepsilon_j\\
&\quad+\frac{(\mathrm{E}\beta_1(z))^2}{y}\frac{1}{n^2}\mathrm{E}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-2}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}+o(\frac{1}{n})\\
&=-\frac{(\mathrm{E}\beta_1(z))^2}{y}\mathrm{E}[\boldsymbol{\alpha}_1^*(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\boldsymbol{\alpha}_1\\
&\quad-\frac{1}{n}\mathrm{E}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}]\varepsilon_j\\
&\quad+\frac{(\mathrm{E}\beta_1(z))^2}{yn^2}\mathrm{E}\mathrm{tr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-2}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}+o(\frac{1}{n}),
\end{aligned}$$

where $\beta_j=\bar{\beta}_j+\bar{\beta}_j^2\varepsilon_j+\bar{\beta}_j^2\beta_j\varepsilon_j^2$, $\varepsilon_j=\boldsymbol{\alpha}_j'(\frac{\mathbf{T}}{z}-\mathbf{S}_j)^{-1}\boldsymbol{\alpha}_j-\frac{1}{n}\mathrm{E}\mathrm{tr}(\frac{\mathbf{T}}{z}-\mathbf{S}_j)^{-1}$ and $\beta_j(z)=\frac{1}{1-ym(z)}+O\left(\frac{1}{n}\right)$.

Step 3: Show that $\dfrac{-z\mathrm{E}\underline{\tilde{s}}_n(z)}{y}pR_n=-\dfrac{y\int\frac{t(b(z))^2dH(t)}{(t/z-b(z))^3}}{1-y\int\frac{(b(z))^2dH(t)}{(t/z-b(z))^2}}+o(1).$

By (1.15) of [6] and (5.23), when all x_{tj} are complex $R_n=0$. For the real case,

$$\frac{\mathrm{E}\underline{\tilde{s}}_n(z)z}{y}R_n$$

$$
\begin{aligned}
=&\frac{-(\mathrm{E}\beta_1(z))^2}{y}\frac{1}{n^2}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-2}\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}+o(\frac{1}{n}) \qquad (5.23)\\
=&\frac{-(\mathrm{E}\beta_1(z))^2}{y}\frac{1}{n^2}\mathrm{Etr}\Big\{\\
&-\mathbf{T}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}(\mathrm{E}\beta_1(z)\mathbf{I}-z^{-1}\mathbf{T})^{-1}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\\
&-\mathrm{E}\beta_1(z)\sum_{k\neq 1}\mathbf{E}_2^{-1}(z)\frac{1}{n}(\frac{\mathbf{T}}{z}-\mathbf{S}_{1k})^{-1}\mathbf{T}\mathrm{E}_1^{-1}(z)\left((\frac{\mathbf{T}}{z}-\mathbf{S}_1)^{-1}-(\frac{\mathbf{T}}{z}-\mathbf{S}_{1k})^{-1}\right)\\
&+\mathrm{E}\beta_1(z)\sum_{k\neq 1}\mathbf{E}_2^{-1}(z)\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^*(\frac{\mathbf{T}}{z}-\mathbf{S}_{1k})^{-1}\mathbf{T}\mathbf{E}_1^{-1}(z)\left((\frac{\mathbf{T}}{z}-\mathbf{S}_1)^{-1}-(\frac{\mathbf{T}}{z}-\mathbf{S}_{1k})^{-1}\right)\\
&+\mathrm{E}\beta_1(z)\sum_{k\neq 1}\mathbf{E}_2^{-1}(z)\left(\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^*-\frac{\mathbf{I}}{n}\right)(\frac{\mathbf{T}}{z}-\mathbf{S}_{1k})^{-1}\mathbf{T}\mathbf{E}_1^{-1}(z)(\frac{\mathbf{T}}{z}-\mathbf{S}_{1k})^{-1}\Big\}+O(\frac{1}{n^{3/2}})\\
=&\frac{(\mathrm{E}\beta_1(z))^2}{yn^2}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\mathbf{T}(\mathrm{E}\beta_1(z)\mathbf{I}-z^{-1}\mathbf{T})^{-1}(z^{-1}\mathbf{T}-\mathbf{S}_1)^{-1}\\
&+y^{-1}n^{-2}(\mathrm{E}\beta_1(z))^3\sum_{k\neq 1}\mathrm{E}\boldsymbol{\alpha}_k^*(z^{-1}\mathbf{T}-\mathbf{S}_{1k})^{-1}\mathbf{T}\mathbf{E}_2^{-1}(z)\boldsymbol{\alpha}_k\\
&\qquad\frac{\boldsymbol{\alpha}_k^*(z^{-1}\mathbf{T}-\mathbf{S}_{1k})^{-2}\mathbf{T}\mathbf{E}_1^{-1}(z)\boldsymbol{\alpha}_k}{1-\boldsymbol{\alpha}_k^*(z^{-1}\mathbf{T}-\mathbf{S}_{1k})^{-1}\boldsymbol{\alpha}_k}+O(\frac{1}{n^{3/2}})\\
=&-\frac{(\mathrm{E}\beta_1(z))^2}{y}\frac{1}{n^2}\mathrm{Etr}(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\mathbf{T}(\mathrm{E}\beta_1(z)\mathbf{I}-z^{-1}\mathbf{T})^{-2}\\
&+\frac{(\mathrm{E}\beta_1(z))^4}{yn^2}\sum_{k\neq 1}\mathrm{E}\boldsymbol{\alpha}_k^*(z^{-1}\mathbf{T}-\mathbf{S}_{1k})^{-1}\mathbf{T}\mathbf{E}_2^{-1}(z)\boldsymbol{\alpha}_k\\
&\qquad\boldsymbol{\alpha}_k^*(z^{-1}\mathbf{T}-\mathbf{S}_{1k})^{-2}\mathbf{T}\mathbf{E}_1^{-1}(z)\boldsymbol{\alpha}_k+O(\frac{1}{n^{3/2}}),
\end{aligned}
$$

where

$$
\begin{aligned}
&\mathbf{E}_1^{-1}(z)=(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1},\\
&\mathbf{E}_2^{-1}(z)=(\mathrm{E}\beta_1(z)\mathbf{I}-z^{-1}\mathbf{T})^{-1},\\
&\mathbf{A}(z)=\sum_{k\neq i}(\frac{n-1}{n}b_i(z)\mathbf{I}-z^{-1}\mathbf{T})^{-1}(\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^*-\frac{1}{n}\mathbf{I})(z^{-1}\mathbf{T}-\mathbf{S}_{ik})^{-1},
\end{aligned}
$$

because

$$
\begin{aligned}
&-\frac{-1}{y(1-y\tilde{m}(z))^2}\frac{1}{n^2}\mathrm{Etr}(\frac{\mathbf{T}}{z}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\sum_{k\neq 1}(\mathrm{E}\beta_1(z)\mathbf{I}-z^{-1}\mathbf{T})^{-1}\frac{1}{n}(\frac{\mathbf{T}}{z}-\mathbf{S}_{1k})^{-1}\\
&((\frac{\mathbf{T}}{z}-\mathbf{S}_1)^{-1}-(\frac{\mathbf{T}}{z}-\mathbf{S}_{1k})^{-1})=O(\frac{1}{n^2}),
\end{aligned}
$$

and

$$
\frac{-1}{y(1-y\tilde{m}(z))^2}\frac{1}{n^2}\mathrm{Etr}(\frac{\mathbf{T}}{z}-\mathrm{E}\beta_1(z)\mathbf{I})^{-1}\sum_{k\neq 1}(\mathrm{E}\beta_1(z)\mathbf{I}-\frac{\mathbf{T}}{z})^{-1}
$$

$$\left(\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^* - \frac{1}{n}\mathbf{I}\right)(\frac{\mathbf{T}}{z} - \mathbf{S}_{1k})^{-2} = O(\frac{1}{n^{3/2}}).$$

Furthermore, we have

$$\begin{aligned}
\frac{-z\mathrm{E}\underline{\tilde{s}}_n(z)}{y}pR_n &= -z\mathrm{E}\underline{\tilde{s}}_n(z)nR_n \\
&= -\frac{\frac{(\mathrm{E}\beta_1(z))^2}{y}\frac{y}{n}\mathrm{Etr}(\mathrm{E}\beta_1(z)\mathbf{I} - z^{-1}\mathbf{T})^{-2}\mathbf{T}(z^{-1}\mathbf{T} - \mathrm{E}\beta_1(z)\mathbf{I})^{-1}}{1 - \frac{1}{(1-y\tilde{m}(z))^2}\frac{1}{n}\mathrm{tr}(\mathrm{E}\beta_1(z)\mathbf{I} - z^{-1}\mathbf{T})^{-2}} \\
&= -\frac{\frac{1}{(1-y\tilde{m}(z))^2}\int\frac{ytdH(t)}{(\mathrm{E}\beta_1(z)-\frac{t}{z})^2(\frac{t}{z}-\mathrm{E}\beta_1(z))}}{1 - \frac{y}{(1-y\tilde{m}(z))^2}\int\frac{dH(t)}{(\mathrm{E}\beta_1(z)-\frac{t}{z})^2}} + o(1) \\
&= -\frac{\frac{1}{(1-y\tilde{m}(z))^2}\int\frac{ytdH(t)}{(t/z-1/(1-y\tilde{m}(z)))^3}}{1 - \frac{y}{(1-y\tilde{m}(z))^2}\int\frac{dH(t)}{(t/z-1/(1-y\tilde{m}(z)))^2}} + o(1) \\
&= -\frac{y\int\frac{t(b(z))^2dH(t)}{(t/z-b(z))^3}}{1 - y\int\frac{(b(z))^2dH(t)}{(t/z-b(z))^2}} + o(1), \qquad (5.24)
\end{aligned}$$

where $\mathrm{E}\beta_1(z) \to \frac{1}{1-y\tilde{m}(z)}$ by (3.11).

Conclusion. By (5.24), (4.6) and (3.10), we have

$$\begin{aligned}
&p(\mathrm{E}s_n(z) - s_n^0(z)) \\
&= \frac{-1}{z^2}p(\mathrm{E}\tilde{s}_n(z) - \tilde{s}_n^0(z)) = \frac{-1}{z^2}n(\mathrm{E}\underline{\tilde{s}}_n(z) - \underline{\tilde{s}}_n^0(z)) \\
&= \frac{-1}{z^2}\frac{\mathrm{E}\underline{\tilde{s}}_n(z)\underline{\tilde{s}}_n^0(z)}{1 - \frac{y\int(t+\mathrm{E}\underline{\tilde{s}}_n(z))^{-1}(t+\underline{\tilde{s}}_n^0(z))^{-1}dH_n(t)}{\left(-z^{-1}+y\int(t+\mathrm{E}\underline{\tilde{s}}_n(z))^{-1}dH_n(t) - R_n\right)\left(-z^{-1}+y\int(t+\underline{\tilde{s}}_n^0(z))^{-1}dH_n(t)\right)}}nR_n \\
&= \frac{\kappa-1}{z^2}\frac{y\int\frac{t(b(z))^3dH(t)}{(t/z-b(z))^3}}{\left(1 - y\int\frac{(b(z))^2dH(t)}{(t/z-b(z))^2}\right)^2} + o(1) \qquad (5.25)
\end{aligned}$$

So we conclude that in the real case

$$\sup_{z\in\mathcal{C}_n}\left|M_n^2(z) - \frac{\kappa-1}{z^2}\frac{y\int\frac{t(b(z))^3dH(t)}{(t/z-b(z))^3}}{\left(1 - y\int\frac{(b(z))^2dH(t)}{(t/z-b(z))^2}\right)^2}\right| \to 0 \quad \text{as } n \to +\infty.$$

Hence the proof of Lemma 5.3 is complete. ■

6. Additional proofs

6.1. *Proof of Proposition 2.1*

In this case, the fourth moments of the entries are different from the Gaussian matching value $\kappa+1$ (3 or 2), the expression (5.6) has an additional term

$$\frac{1}{n^2}\sum_{i=1}^{n} b_i(z_1)b_i(z_2)\sum_{j=1}^{p}(\mathrm{E}|X_{ij}|^4-1-\kappa)\left[\mathrm{E}_{i-1}\left(z_1^{-1}\mathbf{T}-\mathbf{S}_i\right)^{-1}\right]_{jj}\left[\mathrm{E}_{i-1}\left({z_2}^{-1}\mathbf{T}-\mathbf{S}_i\right)^{-1}\right]_{jj}$$

and the expression (5.23) has an additional term

$$-\tfrac{\mathrm{E}(\beta_1(z))^2}{n^2 y}\sum_{j=1}^{p}(\mathrm{E}|X_{1j}|^4-1-\kappa)\left[\left(z^{-1}\mathbf{T}-\mathbf{S}_1\right)^{-1}\right]_{jj} \left[\left(z^{-1}\mathbf{T}-\mathbf{S}_1\right)^{-1}\mathbf{T}\left(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I}\right)^{-1}\right]_{jj}$$

where z_1, z_2 belong to the complex plane $\mathcal{C}$ with $z_1\neq z_2$, $\mathbf{S}_i=\mathbf{S}-n^{-1}\mathbf{x}_i\mathbf{x}_i^*$ and $\beta_i(z)=1/(1-n^{-1}\mathbf{x}_i^*(\mathbf{T}/z-\mathbf{S}_i)^{-1}\mathbf{x}_i)$. Then the covariance (5.17) and mean (5.25) will have additional terms, the limits of almost surely convergence of

$$\frac{\partial^2}{\partial z_1\partial z_2}\left\{\frac{1}{n^2}\sum_{i=1}^{n} b_i(z_1)b_i(z_2)\sum_{j=1}^{p}(\mathrm{E}|X_{ij}|^4-1-\kappa)\left[\mathrm{E}_{i-1}\left(z_1^{-1}\mathbf{T}-\mathbf{S}_i\right)^{-1}\right]_{jj} \left[\mathrm{E}_{i-1}\left({z_2}^{-1}\mathbf{T}-\mathbf{S}_i\right)^{-1}\right]_{jj}\right\}$$

and

$$\frac{yb^3(z)}{z^2g(z)}\frac{1}{p}\sum_{j=1}^{p}\left\{(\mathrm{E}|X_{1j}|^4-1-\kappa)\left[\left(z^{-1}\mathbf{T}-\mathbf{S}_1\right)^{-1}\right]_{jj} \left[\left(z^{-1}\mathbf{T}-\mathbf{S}_1\right)^{-1}\mathbf{T}\left(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I}\right)^{-1}\right]_{jj}\right\}.$$

Therefore, if $\frac{1}{p}\sum_{j=1}^{p}(\mathrm{E}|X_{ij}|^4-1-\kappa)\left[\left({z_1}^{-1}\mathbf{T}-\mathbf{S}_i\right)^{-1}\right]_{jj}\left[\left({z_2}^{-1}\mathbf{T}-\mathbf{S}_i\right)^{-1}\right]_{jj}$ converges to $h(z_1,z_2)$ uniformly in i, then the covariance (5.17) will have the additional term

$$\frac{\partial^2}{\partial z_1\partial z_2}[yb(z_1)b(z_2)h(z_1,z_2)]$$

because $\mathrm{E}b_i(z)\to b(z)$ by (3.10). Collecting these limits lead to the given limiting mean and covariance functions.

6.2. *Proof of Proposition 2.2*

Here we have by assumption that $\mathrm{E}|X_{ij}|^4-1-\kappa=\beta_x+o(1)$ uniformly in i,j and $\mathbf{T}$ is a diagonal matrix with positive eigenvalues. Then we have

$$\frac{1}{p}\sum_{j=1}^{p}(\mathrm{E}|X_{1j}|^4-1-\kappa)\left[\left(z^{-1}\mathbf{T}-\mathbf{S}_1\right)^{-1}\right]_{jj}\left[\left(z^{-1}\mathbf{T}-\mathbf{S}_1\right)^{-1}\mathbf{T}\left(z^{-1}\mathbf{T}-\mathrm{E}\beta_1(z)\mathbf{I}\right)^{-1}\right]_{jj}$$

$$\to h_M(z) = \beta_x \int \frac{z^3 t dH(t)}{[t - zb(z)]^3}$$

and

$$\frac{1}{p}\sum_{j=1}^{p}(\mathrm{E}|X_{ij}|^4 - 1 - \kappa)\left[\left(z_1^{-1}\mathbf{T} - \mathbf{S}_i\right)^{-1}\right]_{jj}\left[\left(z_2{}^{-1}\mathbf{T} - \mathbf{S}_i\right)^{-1}\right]_{jj}$$
$$\to h(z_1, z_2) = \beta_x \int \frac{z_1 z_2 dH(t)}{[t - z_1 b(z_1)][t - z_2 b(z_2)]}.$$

Then the mean (5.25) has the additional term

$$\frac{\beta_x y z b^3(z)}{g(z)} \int \frac{t dH(t)}{[t - zb(z)]^3}$$

and the covariance (5.17) has the additional term

$$\beta_x \frac{\partial^2}{\partial z_1 \partial z_2}\left[y \int \frac{z_1 b(z_1) z_2 b(z_2)}{[t - z_1 b(z_1)][t - z_2 b(z_2)]} dH(t)\right].$$

Applying Proposition 2.1 easily lead to the limiting mean and covariance functions given in this proposition.

7. One lemma

Lemma 7.1. *Under the assumptions of Theorem 2.2, it holds that*

(1) the function $a(z_1, z_2)$ defined in (5.15) satisfies $a(z_1, z_2) < 1$; and
(2) $1 + yzs(z) \neq 0$ for $z \neq 0$ where $s(z)$ denotes the Stieltjes transforms of the LSD given in Theorem 2.1.

Proof. For the first assertion, we have by the formulas given after (5.15),

$$\int \frac{1}{t + \underline{\tilde{m}}(z^{-1})} dH(t) = \frac{\tilde{m}(z)}{z} = \frac{1}{y}(z^{-1} + (\underline{\tilde{m}}(z^{-1}))^{-1}).$$

Then we have

$$-\Im(\underline{\tilde{m}}(z^{-1})) \int \frac{y|\underline{\tilde{m}}(z^{-1})|^2}{|t + \underline{\tilde{m}}(z^{-1})|^2} dH(t) = \Im(\tilde{m}(z)/z) = -\Im(z)\frac{|\underline{\tilde{m}}(z^{-1})|^2}{|z|^2} - \Im(\underline{\tilde{m}}(z^{-1})). \tag{7.1}$$

When $\Im(z) > 0$, then we have

$$-\Im(\underline{\tilde{m}}(z^{-1})) \int \frac{y|\underline{\tilde{m}}(z^{-1})|^2}{|t + \underline{\tilde{m}}(z^{-1})|^2} dH(t) < -\Im(\underline{\tilde{m}}(z^{-1})) \tag{7.2}$$

because $\Im(\underline{\tilde{m}}(z^{-1})) \neq 0$. By (4.1), we have $\tilde{m}(z)/z$ is the limit of

$$p^{-1}\mathrm{tr}(\mathbf{T} - z\mathbf{S})^{-1} = p^{-1}\mathrm{tr}\mathbf{S}^{-1}\mathbf{\Gamma}^T \mathrm{diag}(\frac{\lambda_1 - \Re(z)}{|\lambda_1 - z|^2}, \ldots, \frac{\lambda_p - \Re(z)}{|\lambda_p - z|^2})\mathbf{\Gamma}$$
$$+\mathbf{i}p^{-1}\mathrm{tr}\mathbf{S}^{-1}\mathbf{\Gamma}^T \mathrm{diag}(\frac{\Im(z)}{|\lambda_1 - z|^2}, \ldots, \frac{\Im(z)}{|\lambda_p - z|^2})\mathbf{\Gamma}$$

where $\{\lambda_j\}_{j=1}^p$ are the eigenvalues of $\mathbf{S}^{-1}\mathbf{T}$. By $\Im(z) > 0$, we have $\Im(\tilde{m}(z)/z) > 0$. By (7.1), we know that $-\Im(\underline{\tilde{m}}(z^{-1})) > 0$. Then by (7.2), we have

$$\int \frac{y|\underline{\tilde{m}}(z^{-1})|^2}{|t+\underline{\tilde{m}}(z^{-1})|^2} dH(t) < 1. \tag{7.3}$$

By (5.16), we have

$$|a(z_1,z_2)| \le \sqrt{\int \frac{y|\underline{\tilde{m}}(z_1^{-1})|^2}{|t+\underline{\tilde{m}}(z_1^{-1})|^2} dH(t) \int \frac{y|\underline{\tilde{m}}(z_2^{-1})|^2}{|t+\underline{\tilde{m}}(z_2^{-1})|^2} dH(t)} < 1.$$

That is,

$$|a(z_1,z_2)| < 1.$$

As for the second assertion, we have by (7.2), $\Im(\underline{\tilde{m}}(z^{-1})) \ne 0$ by $z \ne 0$. Also by Line 3 in the equation (5.15), we have $\underline{\tilde{m}}(z^{-1}) = -zb(z) = -z(1+yzs(z))$. It follows then $1 + yzs(z) \ne 0$ for $z \ne 0$. □

Acknowledgement

We thank the Editor, the Associate Editor and the referees for their helpful comments that have led to substantial improvements of the paper. The authors would also acknowledge the support from the following grants: NSFC 11171058 (Shurong Zheng); NSFC 11171057 and Fundamental Research Funds for the Central Universities (Zhidong Bai) and GRF 17305814 (Jianfeng Yao).

References

[1] Alexeev, N. V., Götze, F. and Tikhomirov, A. N. (2010). On the singular spectrum of powers and products of random matrices. *Doklady Mathematics*, **82**(1), 505-507.
[2] Bai, Z. D. and Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Science Press: Beijing.
[3] Bai, Z. D., Hu, J., Pan, G. M., Zhou, W. (2013). Convergence of the empirical spectral distribution function of Beta matrices. *Bernoulli*, **21**(3), 1538-1574.
[4] Bai, Z. D., Jiang, D. D., Yao, J. F. and Zheng, S. R. (2009). Corrections to LRT on large dimensional covariance matrix by RMT. *Annals of Statistics*, **37(6B)**, 3822-3840.
[5] Bai, Z. D., Jiang, D. D., Yao, J. F. and Zheng, S. R. (2009). Testing linear hypotheses in high-dimensional regressions. *Statistics*, **47(6)**, 1207-1223.
[6] Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large dimensional sample covariance matrices. *Ann. Probab.*, **32(1)**, 553-605.
[7] Bai, Z. D. and Yao, J. F. (2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli*, **11(6)**, 1059-1092.
[8] Bai, Z. D., Yin, Y. Q. and Krishnaiah, P. R. (1987). On the limiting empirical distribution function of the eigenvalues of a multivariate F matrix. *Theory Probba. Appl.*, **32**, 537-548.
[9] Billingsley, P. (1995). *Probability and Measure*, John Wiley & Sons, Inc., New York.
[10] Costin, O. and J. Lebowitz (1995). Gaussian fluctuations in random matrices. *Physical Review Letters*, **75**, 69–72.

[11] Couillet, R. and Mérouane, D. (2013). Signal processing in large systems: A new paradigm. *IEEE Signal Processing Magazine*, **30(1)**, 24-39.
[12] Götze, F. and Tikhomirov, A. (2010). The circular law for random matrices. *Ann. of Probbab.*, **38**(4), 1444-1491.
[13] Jiang, D. D., Bai, Z. D. and Zheng, S. R. (2013). Testing the independence of sets of large dimensional variables. *Scicence China Mathematics*, **56(1)**, 135-147.
[14] Johnstone, I. M. (2007). High dimensional statistical inference and random matrices. *Int. Cong. Mathematicians*, **Vol. I**, 307-333. Zürich, Switzerland: European Mathematical Society.
[15] Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.*, **12**, 1-38.
[16] Johansson, K. (1998). On fluctuations of eigenvalues of random hermitian matrices. *Duke Mathematical J.*, **91**, 151–204.
[17] Liu, B. S., Xu, L., Zheng, S. R. and Tian, G. L. (2014). A new test for the proportionality of two large dimensional covariance matrices. *J. Multivariate Anal.*, **131**, 293-308.
[18] Lytova, A. and Pastur, L. (2009). Central limit theorem for linear eigenvalue statistics of the Wigner and the sample covariance random matrices. *Metrika*, **69**, 153–172.
[19] Lytova, A. and Pastur, L. (2009). Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.*, **37**(5), 1178-1840.
[20] Mingo, J. A., S̈niady, P. and Speicher, R. (2007). Second order freeness and fluctuations of random matrices. II. Unitary random matrices. *Adv. Math.*, **209**(1), 212-240.
[21] Mingo, J. A. and Speicher, R. (2006). Second order freeness and fluctuations of random matrices: I. Gaussian and Wishart matrices and cyclic Fock spaces. *J. Funct. Anal.*, **235**(1), 226-270.
[22] Najim, J. and Yao, J. F.(2013). Gaussian fluctions for linear spectral staitstics of large random covairance matrices. *arXiv: 1309.3728.*
[23] Pan, G. M. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.*, **18**, 1232-1270.
[24] Pan, G. M. (2012). Comparison between two types of large sample covariance matrices. *Annales de l'Institut Henri Poincare-Probabiliteset Statistiques*, **50**(2), 655-677.
[25] Paul, D. and Aue, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Planning and Inference*, **150**, 1-29.
[26] Rao, N. R. and Speicher, R. (2007). Multiplication of free random variables and the S-transform: the case of vanishing mean. *Elect. Comm. in Probab.*, **12**, 248-258.
[27] Silverstein, J. W. (1985). The limiting eigenvalue distribution of a multivariate F matrix. *SIAM J. Math. Anal.*, **16(3)**, 641-646.
[28] Sinaĭ, Y. A. and Soshnikov, A. (1998). Central limit theorems for traces of large random matrices with independent entries. *Bol. Soc. Brasil. Mat.*, **29**, 1-24.
[29] Voiculescu, D. (1987). Multiplication of certain noncommuting random variables. *J. Operator Theory*, **18**, 223-235.
[30] Voiculescu, D., Dykema, K. and Nica, A. (1992). Free random variables. *Vol. 1 of CRm Monograph Series, American Mathematical Society, Providence, RI.*
[31] Wachter K. W. (1980). The limiting empirical measure of multiple discriminant ratios. *Ann. Statist.*, **8**(5), 937-957.
[32] Zheng, S. R. (2012). Central Limit Theorem for Linear Spectral Statistics of Large Dimensional F-Matrix. *Annales de l'Institut Henri Poincare-Probabiliteset Statistiques*, **48**(2), 444-476.
[33] Zheng, S. R., Bai, Z. D. and Yao, J. F. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications

July 17, 2015 9:23 WSPC/INSTRUCTION FILE st-rev-2

to hypothesis testing. *Annals of Statistics*, **43**(2), 546-591.

[34] Zheng, S. R., Bai, Z. D. and Yao, J. F. (2015). CLT for large dimensional general Fisher matrices and its applications in high-dimensional data analysis. *Preprint*, DOI: 10.13140/RG.2.1.3973.7765