Title	Joint CLT for several random sesquilinear forms with applications to large-dimensional spiked population models
Author(s)	Wang, Q; SU, Z; Yao, JJ
Citation	Electronic Journal of Probability, 2014, v. 19 n. 103, p. 1-28
Issued Date	2014
URL	http://hdl.handle.net/10722/217227
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Electronic Journal of Probability

Electron. J. Probab. **19** (2014), no. 103, 1–28.
ISSN: 1083-6489 DOI: 10.1214/EJP.v19-3339

Joint CLT for several random sesquilinear forms with applications to large-dimensional spiked population models*

Wang Qinwen[†] Su Zhonggen [‡] Yao Jianfeng[§]

Abstract

In this paper, we derive a joint central limit theorem for random vector whose components are function of random sesquilinear forms. This result is a natural extension of the existing central limit theory on random quadratic forms. We also provide applications in random matrix theory related to large-dimensional spiked population models. For the first application, we find the joint distribution of grouped extreme sample eigenvalues correspond to the spikes. And for the second application, under the assumption that the population covariance matrix is diagonal with k (fixed) simple spikes, we derive the asymptotic joint distribution of the extreme sample eigenvalue and its corresponding sample eigenvector projection.

Keywords: Central limit theorem; Extreme eigenvalues; Extreme eigenvectors; Joint distribution; Large-dimensional sample covariance matrices; Random quadratic form; Random sesquilinear form; Spiked population model.
AMS MSC 2010: Primary 60F05, Secondary 60B20.
Submitted to EJP on February 23, 2014, final version accepted on October 27, 2014.
Supersedes arXiv:1402.6064.

1 Introduction

The aim of this paper is to derive the joint central limit theorem of a new type of random vector whose components are made with several groups of random sesquilinear forms. To be more specific, we consider a sequence $\left\{(x_i, y_i)_{i\in\mathbb{N}}\right\}$ of iid. complex-valued, zero-mean random vector belonging to $\mathbb{C}^K \times \mathbb{C}^K$ (K fixed) with a finite moment of fourth-order. For positive integer $n \geq 1$, write

$$x_i = (x_{1i}, \cdots, x_{Ki})^T, \quad X(l) = (x_{l1}, \cdots, x_{ln})^T \ (1 \leq l \leq K) \ , \tag{1.1}$$

*Supported by the National Natural Science Foundation of China [11071213, 11371317] and Research Grants Council, Hong Kong SAR, China [781511M, 705413P].
[†]Department of Mathematics, Zhejiang University, China. E-mail: wqw8813@gmail.com
[‡]Department of Mathematics, Zhejiang University, China. E-mail: suzhonggen@zju.edu.cn
[§]Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam, Hong Kong.
E-mail: jeffyao@hku.hk

with a similar definition for the vectors $\{y_i\}$ and $\{Y(l)\}_{1\le l\le K}$. The covariance between x_{l1} and y_{l1} is denoted as $\rho(l) = E[\overline{x}_{l1}y_{l1}]$, $1 \le l \le K$. Let $\big\{A_n = [a_{ij}(n)]\big\}_n$ and $\big\{B_n = [b_{ij}(n)]\big\}_n$ be two sequences of $n \times n$ Hermitian matrices, and define

$$U(l) := \frac{1}{\sqrt{n}}\big[X(l)^* A_n Y(l) - \rho(l) tr A_n\big] \ , \qquad (1.2)$$
$$V(l) := \frac{1}{\sqrt{n}}\big[X(l)^* B_n Y(l) - \rho(l) tr B_n\big] \ .$$

We are studying the joint central limit theorem of the $2K$-dimensional complex-valued random vector:

$$\big(U(1), \cdots, U(K), V(1), \cdots, V(K)\big)^T \ .$$

If we use only one sequence of Hermitian matrix, say $\{A_n\}$ and consider one form ($K = 1$), then the problem reduces to the central limit theorem of a simple random sesquilinear form:

$$U(1) := \frac{1}{\sqrt{n}}\big[X(1)^* A_n Y(1) - \rho(1) tr A_n\big] \ .$$

If we further impose $Y \equiv X$, we obtain a classical random quadratic form

$$U^*(1) := \frac{1}{\sqrt{n}}\big[X(1)^* A_n X(1) - \rho(1) tr A_n\big]$$

with independent random variables.

There exists an extensive literature on the asymptotic distribution of quadratic form $U^*(1)$. The pioneering work in this area dates back to [23], who deals principally with the case when the variables X have normal distribution. This CLT is extended to arbitrary iid. components in X by [24], with additional conditions on the matrix A: in particular, A has a zero diagonal (i.e quadratic form: $\tilde{U}(1) := \frac{1}{\sqrt{n}} X(1)^* A_n X(1)$). Later extensions deal with other types of limiting theorem (functional CLT, law of iterated logarithm) or dependent random variables in X, see: [21], [9], [10], [17] and [12] for reference.

In a different area, [18] and [11] established the asymptotic behavior of quadratic form and bilinear form, where $A = S_n$ is a sample covariance matrix and $A = (M_n - zI)^{-1}$ is the resolvent of some large dimensional random matrix M_n, respectively. Such CLT can be used in the areas of wireless communications and electrical engineering.

In the paper of [2], the authors derived the central limit theorem for $U(l)$ in (1.2) (i.e with one group of sesquilinear forms) in their Appendix as a tool for establishing the central limit theory for the extreme sample eigenvalues when the population has a spiked covariance structure.

In this paper, we follow the lines and strategy that was put forward in [2], and extend this CLT to arbitrary number of groups of random sesquilinear forms, which is presented in Section 2. Indeed, this extension has been motivated by applications in the field of random matrix theory related to the spiked population model. When the population has a spiked covariance structure, we establish the asymptotic joint distribution of any two groups of extreme sample eigenvalues that correspond to the spikes. Besides, when the population covariance matrix is diagonal with k (fixed) simple spikes, we find the joint distribution of the extreme sample eigenvalue and its corresponding sample eigenvector projection using our main result. All these applications are developed in Section 3. Section 4 and the last Section contain proofs and some additional technical lemmas.

2 Main result: central limit theorem for random sesquilinear forms

Theorem 2.1. *Let* $\{A_n = [a_{ij}(n)]\}_n$ *and* $\{B_n = [b_{ij}(n)]\}_n$ *be two sequences of* $n \times n$ *Hermitian matrices and the vector* $\{X(l), Y(l)\}_{1\le l\le K}$ *be defined as in* (1.1). *Assume that the following limits exist:*

$$\begin{aligned} w_1 &= \lim_{n\to\infty} \frac{1}{n}\, tr[A_n \circ A_n], w_2 = \lim_{n\to\infty} \frac{1}{n}\, tr[B_n \circ B_n], w_3 = \lim_{n\to\infty} \frac{1}{n}\, tr[A_n \circ B_n],\\ \theta_1 &= \lim_{n\to\infty} \frac{1}{n}\, tr[A_n A_n^*], \quad \theta_2 = \lim_{n\to\infty} \frac{1}{n}\, tr[B_n B_n^*], \quad \theta_3 = \lim_{n\to\infty} \frac{1}{n}\, tr[A_n B_n^*],\\ \tau_1 &= \lim_{n\to\infty} \frac{1}{n}\, tr[A_n^2], \qquad \tau_2 = \lim_{n\to\infty} \frac{1}{n}\, tr[B_n^2], \qquad \tau_3 = \lim_{n\to\infty} \frac{1}{n}\, tr[A_n B_n], \end{aligned}$$

where $A \circ B$ *denotes the Hadamard product of two matrices* A *and* B*, i.e.* $(A \circ B)_{ij} = A_{ij} \cdot B_{ij}$. *Define two groups of sesquilinear forms:*

$$U(l) = \frac{1}{\sqrt{n}}\left[X(l)^* A_n Y(l) - \rho(l) tr A_n\right], V(l) = \frac{1}{\sqrt{n}}\left[X(l)^* B_n Y(l) - \rho(l) tr B_n\right].$$

Then, the $2K$*-dimensional complex-valued random vector:*

$$\big(U(1), \cdots, U(K), V(1), \cdots, V(K)\big)^T$$

converges weakly to a zero-mean complex-valued vector W *whose real and imaginary parts are Gaussian. Moreover, the Laplace transform of* W *is given by*

$$\mathbb{E}\exp\left(\begin{pmatrix} c\\ d\end{pmatrix}^T W\right) = \exp\left[\frac{1}{2}\begin{pmatrix} c\\ d\end{pmatrix}^T B\begin{pmatrix} c & d\end{pmatrix}\right], \quad c, d \in \mathbb{C}^K,$$

with

$$B = \begin{pmatrix} B_{11} & B_{12}\\ B_{12} & B_{22}\end{pmatrix}_{2K\times 2K}.$$

Each block within B *is a* $K \times K$ *matrix, having the structure (*$l, l^{'} = 1, \cdots, K$*):*

$$\begin{aligned} B_{11}(l,l^{'}) &= \mathrm{Cov}\ (U(l), U(l^{'})) = w_1 A_1 + (\tau_1 - w_1) A_2 + (\theta_1 - w_1) A_3,\\ B_{22}(l,l^{'}) &= \mathrm{Cov}\ (V(l), V(l^{'})) = w_2 A_1 + (\tau_2 - w_2) A_2 + (\theta_2 - w_2) A_3,\\ B_{12}(l,l^{'}) &= \mathrm{Cov}\ (U(l), V(l^{'})) = w_3 A_1 + (\tau_3 - w_3) A_2 + (\theta_3 - w_3) A_3, \end{aligned}$$

where A_1*,* A_2 *and* A_3 *are given by*

$$A_1 = \mathbb{E}(\overline{x}_{l1} y_{l1} \overline{x}_{l^{'}1} y_{l^{'}1}) - \rho(l)\rho(l^{'}), \tag{2.1}$$

$$A_2 = \mathbb{E}(\overline{x}_{l1} \overline{x}_{l^{'}1}) E(y_{l1} y_{l^{'}1}), \tag{2.2}$$

$$A_3 = \mathbb{E}(\overline{x}_{l1} y_{l^{'}1}) E(\overline{x}_{l^{'}1} y_{l1}). \tag{2.3}$$

Proof. (proof of Theorem 2.1) It is sufficient to establish the CLT for the linear combinations of random Hermitian sesquilinear forms:

$$\sum_{l=1}^{K} [c_l X(l)^* A_n Y(l) + d_l X(l)^* B_n Y(l)],$$

where the coefficients $(c_l), (d_l) \in \mathbb{C}^K \times \mathbb{C}^K$ are arbitrary. Also, it holds that

$$\mathbb{E}[X(l)^* A_n Y(l)] = \rho(l) tr A_n, \quad \mathbb{E}[X(l)^* B_n Y(l)] = \rho(l) tr B_n.$$

We use the moment method as in [2]. Consider the linear combination of the two sesquilinear forms

$$\eta_n = \frac{1}{\sqrt{n}} \sum_{l=1}^{K} \left\{ c_l[X(l)^* A_n Y(l) - \rho(l) tr A_n] + d_l[X(l)^* B_n Y(l) - \rho(l) tr B_n] \right\} ,$$

which can be expanded as follows:

$$\begin{aligned} \eta_n &= \frac{1}{\sqrt{n}} \sum_{l=1}^{K} \Big\{ c_l \big[\sum_{u=1}^{n} (X(l)_u^* Y(l)_u - \rho(l)) a_{uu} + \sum_{u \neq v} X(l)_u^* Y(l)_v a_{uv} \big] \\ &\quad + d_l \big[\sum_{u=1}^{n} (X(l)_u^* Y(l)_u - \rho(l)) b_{uu} + \sum_{u \neq v} X(l)_u^* Y(l)_v b_{uv} \big] \Big\} \\ &= \frac{1}{\sqrt{n}} \sum_{e=(u,v)} \Big\{ \sum_{l=1}^{K} \big[(c_l \overline{x}_{lu} y_{lu} - c_l \rho(l)) a_{uu} + c_l \overline{x}_{lu} y_{lv} a_{uv} \big] \\ &\quad + \sum_{l=1}^{K} \big[(d_l \overline{x}_{lu} y_{lu} - d_l \rho(l)) b_{uu} + d_l \overline{x}_{lu} y_{lv} b_{uv} \big] \Big\} \\ &= \frac{1}{\sqrt{n}} \sum_{e} (a_e \psi_e + b_e \varphi_e) , \end{aligned}$$

where e is an edge associated with vertex u and v, i.e. $e = (u, v) \in \{1, \cdots, n\}^2$; and

$$\psi_e \triangleq \begin{cases} \sum_{l=1}^{K} c_l(\overline{x}_{lu} y_{lu} - \rho(l)) , & u = v , \\ \sum_{l=1}^{K} c_l \overline{x}_{lu} y_{lv} , & u \neq v , \end{cases} \tag{2.4}$$

$$\varphi_e \triangleq \begin{cases} \sum_{l=1}^{K} d_l(\overline{x}_{lu} y_{lu} - \rho(l)) , & u = v , \\ \sum_{l=1}^{K} d_l \overline{x}_{lu} y_{lv} , & u \neq v . \end{cases} \tag{2.5}$$

Then

$$\begin{aligned} n^{\frac{K}{2}} \eta_n^K &= \sum_{e_1 \cdots e_K} (a_{e_1} \psi_{e_1} + b_{e_1} \varphi_{e_1}) \cdots (a_{e_K} \psi_{e_K} + b_{e_K} \varphi_{e_K}) \\ &= \sum_{G_1 \bigcup G_2} a_{G_1} \psi_{G_1} b_{G_2} \varphi_{G_2} , \end{aligned} \tag{2.6}$$

where

$$a_{G_1} = \prod_{e \in G_1} a_e , \quad \psi_{G_1} = \prod_{e \in G_1} \psi_e , \quad b_{G_2} = \prod_{e \in G_2} b_e , \quad \varphi_{G_2} = \prod_{e \in G_2} \varphi_e .$$

To each sum in equation (2.6), we associate a directed graph G by drawing an arrow $u \to v$ for each factor $e_j = (u, v)$. We denote G_1 as a subgraph of G corresponding to the coefficients being $a\psi$, and G_2 the remaining: $G_2 = G \backslash G_1$. Besides, to a loop $u \to u$ corresponds the product $a_{uu}\psi_{uu} = a_{uu} \sum_{l=1}^{K} c_l(\overline{x}_{lu} y_{lu} - \rho(l))$ and to an edge $u \to v$ $(u \neq v)$ corresponds the product $a_{uv}\psi_{uv} = a_{uv} \sum_{l=1}^{K} c_l \overline{x}_{lu} y_{lv}$. The same holds for $b_{uu}\varphi_{uu}$ and $b_{uv}\varphi_{uv}$.

In the paper of [2] (proof of Theorem 7.1), they show that only three types of components in the graph G contribute to a non-negligible term (see Figure 1):

Because G_1 and G_2 are subgraphs of G, and by the definition in equation (2.4) and (2.5), ψ_e differs from φ_e only through the coefficient c_l or d_l in front. So the difference between ψ_e and φ_e is at most $O(1)$, which means that for the components in the graph

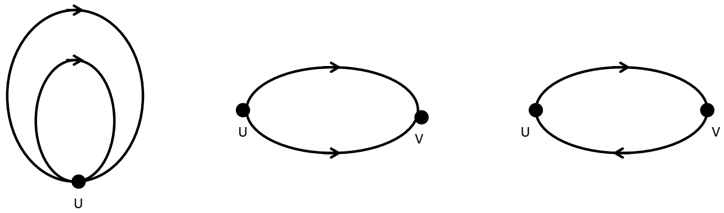

Figure 1: three major components in the graph G

G that have $o(1)$ contribution to $En^{K/2}\xi_n^K$ (see [2] for detail of ξ_n) should still have $o(1)$ contribution to $En^{K/2}\eta_n^K$. Based on this fact, we get this time that only the influence of the following nine components (in Figure 2) counts. The numbers $k_1, \cdots, k_9$ in Figure 2 stand for the multiplicity of each component, so by degree of each vertex, we also have the restriction that $4(k_1 + \cdots + k_9) = 2K$, which means K should be an even number, denoted as $2p$ for convenience.

From the combinatorics, we have this time

$$
\begin{aligned}
\mathbb{E}n^{K/2}\eta_n^K &= \mathbb{E}\sum_{G_1 \bigcup G_2} a_{G_1}\psi_{G_1}b_{G_2}\varphi_{G_2} \\
&= \sum_{2(k_1+\cdots+k_9)=K} \frac{\binom{K}{2}\binom{K-2}{2}\cdots\binom{2}{2}\cdot 2^{k_3+k_6+k_9}}{k_1!\cdots k_9!} \times D_1D_2\cdots D_9 + o(n^{K/2}) \\
&= \sum_{k_1+\cdots+k_9=p} \frac{(2p)!\cdot 2^{k_3+k_6+k_9}}{2^p\cdot k_1!\cdots k_9!} \times D_1D_2\cdots D_9 + o(n^{K/2})\ . \qquad (2.7)
\end{aligned}
$$

The coefficients in front of $D_1D_2\cdots D_9$ is due to the fact that by observing the nine components in Figure 2, we find that each component is made of two edges; first we combine two edges in a group in the total of K edges, that is $\binom{K}{2}\binom{K-2}{2}\cdots\binom{2}{2}$; second, the first k_1 (also the following $k_2, \cdots, k_9$) groups should be the same, we must exclude the $k_1!\cdots k_9!$ possibilities from the total of $\binom{K}{2}\binom{K-2}{2}\cdots\binom{2}{2}$; and last, for the three components in the last column of Figure 2, the two edges in each component belong to different subgraphs (one edge in G_1 and the other in G_2), so there should be an additional perturbation $2^{k_3+k_6+k_9}$ added, and combine all these facts leads to the result.

Then we specify the terms of $D_1, D_2, \cdots, D_9$ in the following:

$$
\begin{aligned}
D_1 &= \prod_{j=1}^{k_1}\mathbb{E}\Big[a_{u_ju_j}^2\big\{\sum_{l=1}^{K}c_l\big(\overline{x}_{lu_j}y_{lu_j}-\rho(l)\big)\big\}^2\Big] \\
&= \prod_{j=1}^{k_1}a_{u_ju_j}^2\sum_{l,l'}c_lc_{l'}\left[E(\overline{x}_{l1}y_{l1}\overline{x}_{l'1}y_{l'1})-\rho(l)\rho(l^{'})\right] \\
&= \prod_{j=1}^{k_1}a_{u_ju_j}^2\sum_{l,l'}c_lc_{l'}A_1 \\
&\triangleq \prod_{j=1}^{k_1}a_{u_ju_j}^2\alpha_1\ .
\end{aligned}
$$

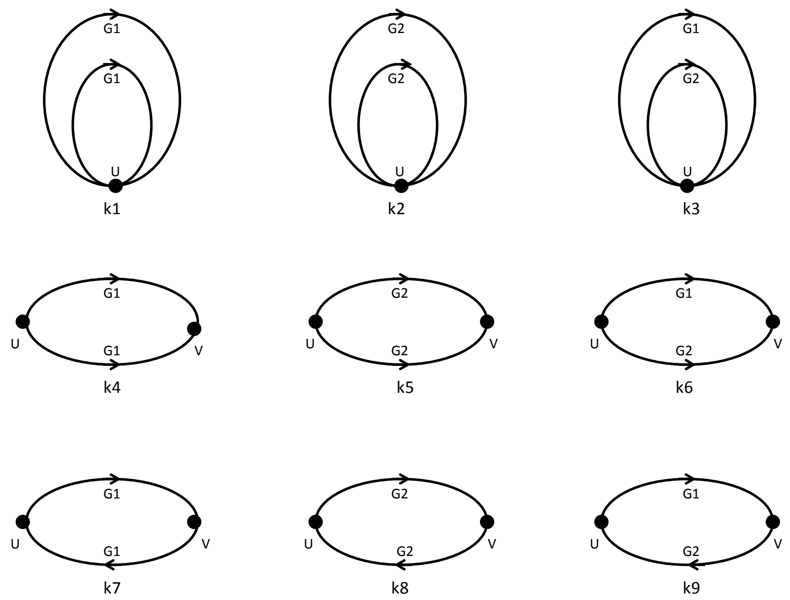

Figure 2: nine major components in the graph $G_1 \bigcup G_2$

Similarly, we have:

$$
\begin{aligned}
D_2 &= \prod_{j=1}^{k_2} \mathbb{E}\Big[b_{u_j u_j}^2 \big\{ \sum_{l=1}^{K} d_l \big(\overline{x}_{lu_j} y_{lu_j} - \rho(l)\big) \big\}^2 \Big] \\
&\triangleq \prod_{j=1}^{k_2} b_{u_j u_j}^2 \beta_1 \ ,
\end{aligned}
$$

$$
\begin{aligned}
D_3 &= \prod_{j=1}^{k_3} \mathbb{E}\Big[a_{u_j u_j} b_{u_j u_j} \sum_{l=1}^{K} c_l \big(\overline{x}_{lu_j} y_{lu_j} - \rho(l)\big) \sum_{l=1}^{K} d_l \big(\overline{x}_{lu_j} y_{lu_j} - \rho(l)\big) \Big] \\
&\triangleq \prod_{j=1}^{k_3} a_{u_j u_j} b_{u_j u_j} \gamma_1 \ ,
\end{aligned}
$$

$$
\begin{aligned}
D_4 &= \prod_{j=1}^{k_4} \mathbb{E}\Big[a_{u_j v_j}^2 \big(\sum_{l=1}^{K} c_l \overline{x}_{lu_j} y_{lv_j} \big)^2 \Big] \\
&= \prod_{j=1}^{k_4} a_{u_j v_j}^2 \sum_{l,l'} c_l c_{l'} E\big(\overline{x}_{l1} \overline{x}_{l'1}\big) E\big(y_{l1} y_{l'1}\big) \\
&= \prod_{j=1}^{k_4} a_{u_j v_j}^2 \sum_{l,l'} c_l c_{l'} A_2 \\
&\triangleq \prod_{j=1}^{k_4} a_{u_j v_j}^2 \alpha_2 \ ,
\end{aligned}
$$

$$\begin{aligned} D_5 &= \prod_{j=1}^{k_5} \mathbb{E}\Big[b_{u_jv_j}^2\big(\sum_{l=1}^{K} d_l\overline{x}_{lu_j}y_{lv_j}\big)^2\Big] \\ &\triangleq \prod_{j=1}^{k_5} b_{u_jv_j}^2\beta_2\ , \end{aligned}$$

$$\begin{aligned} D_6 &= \prod_{j=1}^{k_6} \mathbb{E}\Big[a_{u_jv_j}b_{u_jv_j}\big(\sum_{l=1}^{K} c_l\overline{x}_{lu_j}y_{lv_j}\big)\big(\sum_{l=1}^{K} d_l\overline{x}_{lu_j}y_{lv_j}\big)\Big] \\ &\triangleq \prod_{j=1}^{k_6} a_{u_jv_j}b_{u_jv_j}\gamma_2\ , \end{aligned}$$

$$\begin{aligned} D_7 &= \prod_{j=1}^{k_7} \mathbb{E}\Big[|a_{u_jv_j}|^2\big(\sum_{l=1}^{K} c_l\overline{x}_{lu_j}y_{lv_j}\big)\big(\sum_{l=1}^{K} c_l\overline{x}_{lv_j}y_{lu_j}\big)\Big] \\ &= \prod_{j=1}^{k_7} |a_{u_jv_j}|^2\sum_{l,l'} c_lc_{l'}\,E\big(\overline{x}_{l1}y_{l'1}\big)E\big(\overline{x}_{l'1}y_{l1}\big) \\ &= \prod_{j=1}^{k_7} |a_{u_jv_j}|^2\sum_{l,l'} c_lc_{l'}\,A_3 \\ &\triangleq \prod_{j=1}^{k_7} |a_{u_jv_j}|^2\alpha_3\ , \end{aligned}$$

$$\begin{aligned} D_8 &= \prod_{j=1}^{k_8} \mathbb{E}\Big[|b_{u_jv_j}|^2\big(\sum_{l=1}^{K} d_l\overline{x}_{lu_j}y_{lv_j}\big)\big(\sum_{l=1}^{K} d_l\overline{x}_{lv_j}y_{lu_j}\big)\Big] \\ &\triangleq \prod_{j=1}^{k_8} |b_{u_jv_j}|^2\beta_3\ , \end{aligned}$$

$$\begin{aligned} D_9 &= \prod_{j=1}^{k_9} \mathbb{E}\Big[a_{u_jv_j}b_{v_ju_j}\big(\sum_{l=1}^{K} c_l\overline{x}_{lu_j}y_{lv_j}\big)\big(\sum_{l=1}^{K} d_l\overline{x}_{lv_j}y_{lu_j}\big)\Big] \\ &\triangleq \prod_{j=1}^{k_9} a_{u_jv_j}b_{v_ju_j}\gamma_3\ . \end{aligned}$$

Combine these nine terms with equation (2.7), we have

$$\begin{aligned} \mathbb{E}\eta_n^{2p} &= n^{-p}\sum_{k_1+\cdots+k_9=p}\frac{(2p)!\cdot 2^{k_3+k_6+k_9}}{2^p\cdot k_1!\cdots k_9!}\prod_{(j_1\cdots j_9)=(1\cdots 1)}^{(k_1\cdots k_9)} a_{u_{j_1}u_{j_1}}^2\alpha_1^{k_1}b_{u_{j_2}u_{j_2}}^2 \\ &\quad\times\beta_1^{k_2}a_{u_{j_3}u_{j_3}}b_{u_{j_3}u_{j_3}}\gamma_1^{k_3}a_{u_{j_4}v_{j_4}}^2\alpha_2^{k_4}b_{u_{j_5}v_{j_5}}^2\beta_2^{k_5}a_{u_{j_6}v_{j_6}}b_{u_{j_6}v_{j_6}}\gamma_2^{k_6} \\ &\quad\times|a_{u_{j_7}v_{j_7}}|^2\alpha_3^{k_7}|b_{u_{j_8}v_{j_8}}|^2\beta_3^{k_8}a_{u_{j_9}v_{j_9}}b_{v_{j_9}u_{j_9}}\gamma_3^{k_9}+o(1) \\ &= \frac{(2p-1)!!}{n^p}\big(\alpha_1\sum_{u=1}^{n}a_{uu}^2+\beta_1\sum_{u=1}^{n}b_{uu}^2+2\gamma_1\sum_{u=1}^{n}a_{uu}b_{uu}+\alpha_2\sum_{u\neq v}a_{uv}^2 \\ &\quad+\beta_2\sum_{u\neq v}b_{uv}^2+2\gamma_2\sum_{u\neq v}a_{uv}b_{uv}+\alpha_3\sum_{u\neq v}|a_{uv}|^2+\beta_3\sum_{u\neq v}|b_{uv}|^2 \\ &\quad+2\gamma_3\sum_{u\neq v}a_{uv}b_{vu}\big)^p+o(1)\ , \end{aligned}$$

which means that $\eta_n \Longrightarrow \mathcal{N}(0, \sigma^2)$ by the moment method, with

$$\begin{aligned}
\sigma^2 &= \lim_{n\to\infty} \frac{1}{n}\Big[\alpha_1 \sum_{u=1}^{n} a_{uu}^2 + \beta_1 \sum_{u=1}^{n} b_{uu}^2 + 2\gamma_1 \sum_{u=1}^{n} a_{uu}b_{uu} + \alpha_2 \sum_{u\neq v} a_{uv}^2 + \beta_2 \sum_{u\neq v} b_{uv}^2 \\
&\quad +2\gamma_2 \sum_{u\neq v} a_{uv}b_{uv} + \alpha_3 \sum_{u\neq v} |a_{uv}|^2 + \beta_3 \sum_{u\neq v} |b_{uv}|^2 + 2\gamma_3 \sum_{u\neq v} a_{uv}b_{vu}\Big] \\
&= \alpha_1 w_1 + \beta_1 w_2 + 2\gamma_1 w_3 + \alpha_2(\tau_1 - w_1) + \beta_2(\tau_2 - w_2) + 2\gamma_2(\tau_3 - w_3) \\
&\quad +\alpha_3(\theta_1 - w_1) + \beta_3(\theta_2 - w_2) + 2\gamma_3(\theta_3 - w_3) \\
&= \sum_{l,l'} c_l c_{l'} A_1 w_1 + \sum_{l,l'} d_l d_{l'} A_1 w_2 + 2\sum_{l,l'} c_l d_{l'} A_1 w_3 + \sum_{l,l'} c_l c_{l'} A_2(\tau_1 - w_1) \\
&\quad + \sum_{l,l'} d_l d_{l'} A_2(\tau_2 - w_2) + 2\sum_{l,l'} c_l d_{l'} A_2(\tau_3 - w_3) + \sum_{l,l'} c_l c_{l'} A_3(\theta_1 - w_1) \\
&\quad + \sum_{l,l'} d_l d_{l'} A_3(\theta_2 - w_2) + 2\sum_{l,l'} c_l d_{l'} A_3(\theta_3 - w_3) \\
&= \sum_{l,l'} c_l c_{l'} \big(A_1 w_1 + A_2(\tau_1 - w_1) + A_3(\theta_1 - w_1)\big) \\
&\quad + \sum_{l,l'} d_l d_{l'} \big(A_1 w_2 + A_2(\tau_2 - w_2) + A_3(\theta_2 - w_2)\big) \\
&\quad +2\sum_{l,l'} c_l d_{l'} \big(A_1 w_3 + A_2(\tau_3 - w_3) + A_3(\theta_3 - w_3)\big) .
\end{aligned}$$

The proof of Theorem 2.1 is complete. □

Corollary 2.2. *Under the same conditions as in Theorem 2.1, but with real random vectors* $\{(x_i, y_i)_{i\in N}\}$*, symmetric matrices* $\{A_n = [a_{ij}(n)]\}_n$ *and* $\{B_n = [b_{ij}(n)]\}_n$*, the* $2K$*-dimensional real-valued random vector:*

$$\big(U(1), \cdots, U(K), V(1), \cdots, V(K)\big)^T$$

converges weakly to a zero-mean $2K$*-dimensional Gaussian vector with covariance matrix* B*.*

Theorem 2.1 can be generalized to the joint distribution of several sesquilinear forms. We present this generalization in the following theorem. Recall that in the proof of Theorem 2.1, we use the moment method and find the nine major components presented in Figure 2, which all contain two edges. Therefore, if now we consider the k sesquilinear forms as a whole, there should be $\frac{3}{2}k(1+k)$ major components that will lead to a nonnegligible contribution. And each component still has two edges, from the same subgraph (both from G_i $(i = 1, \cdots, k)$ or from two different subgraphs (one from G_i and the other from G_j $(i \neq j)$). This means that the k sesquilinear forms packed together only has pairwise covariance function. The proof for other steps is similar and omitted.

Theorem 2.3. *Let* $\big\{A_m = [a_{ij}^{(m)}(n)]\big\}_n$ $m = (1, \cdots, k)$ *be* k *sequences of* $n\times n$ *Hermitian matrices and the vector* $\{X(l), Y(l)\}_{1\le l\le K}$ *are defined as* (1.1)*. Assume that the following limits exists* ($m, m' = (1, \cdots, k)$ *and* $m \neq m'$*):*

$$\begin{aligned}
w_m &= \lim_{n\to\infty} \frac{1}{n}\, tr[A_m \circ A_m] , &\quad w_{mm'} &= \lim_{n\to\infty} \frac{1}{n}\, tr[A_m \circ A_{m'}] , \\
\theta_m &= \lim_{n\to\infty} \frac{1}{n}\, tr[A_m A_m^*] , &\quad \theta_{mm'} &= \lim_{n\to\infty} \frac{1}{n}\, tr[A_m A_{m'}^*] , \\
\tau_m &= \lim_{n\to\infty} \frac{1}{n}\, tr[A_m^2] , &\quad \tau_{mm'} &= \lim_{n\to\infty} \frac{1}{n}\, tr[A_m A_{m'}] .
\end{aligned}$$

Denote the sesquilinear forms:

$$U^{(m)}(l) = \frac{1}{\sqrt{n}}\left[X(l)^* A_m Y(l) - \rho(l) tr A_m\right] , m = 1, \cdots , k,$$

then the $(K \cdot k)$*-dimensional complex-valued random vector:*

$$\left(U^{(1)}(1), \cdots , U^{(1)}(K), U^{(2)}(1), \cdots , U^{(2)}(K), U^{(k)}(1), \cdots , U^{(k)}(K)\right)^T$$

converges weakly to a zero-mean complex-valued vector W *whose real and imaginary parts are Gaussian. Moreover, the Laplace transform of* W *is given by*

$$\mathbb{E}\exp\left(\begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix}^T W\right) = \exp\left[\frac{1}{2}\begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix}^T B\begin{pmatrix} c_1 & \cdots & c_k \end{pmatrix}\right] , \quad c_i \in \mathbb{C}^K ,$$

where B *could be written as*

$$B = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1k} \\ B_{21} & B_{22} & \cdots & B_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ B_{k1} & B_{k2} & \cdots & B_{kk} \end{pmatrix}_{(K\cdot k)\times(K\cdot k)} ,$$

each block is a $K \times K$ *matrices with entries (for* $l, l^{'} = 1, \cdots , K$*):*

$$B_{ii}(l, l^{'}) = \textit{Cov}\ (U^i(l), U^i(l^{'})) = w_i A_1 + (\tau_i - w_i)A_2 + (\theta_i - w_i)A_3 , \tag{2.8}$$

$$B_{ij}(l, l^{'}) = \textit{Cov}\ (U^i(l), U^j(l^{'})) = w_{ij} A_1 + (\tau_{ij} - w_{ij})A_2 + (\theta_{ij} - w_{ij})A_3 , \tag{2.9}$$

and A_1 A_2 *and* A_3 *are the same as* (2.3)*.*

Here we give an application related to the existing literature on large-dimensional covariance matrices. In [18], they establish the central limit theorem of the random quadratic forms $s_1^T(SS^T)^i s_1$, where $S = (s_1, \cdots , s_k)$, $s_i = \frac{1}{\sqrt{n}}(v_{i1} \cdots v_{in})^T$, $\{v_{ij}\}$ are i.i.d. with $\mathbb{E}v_{11} = 0$, $\mathbb{E}v_{11}^2 = 1$, $\mathbb{E}v_{11}^4 = \nu_4 < \infty$. This $s_1^T(SS^T)^i s_1$ can be written as a linear combination of a series of random quadratic forms whose random matrices involved are independent of the random vector. Their Lemma 3.2 states such joint distribution of these random quadratic forms, which can be restated and proved using our Theorem 2.3.

Proposition 2.4. *[[18]] Let* $S_1 = (s_2 \cdots s_k)$*, independent of* s_1*. Then the random vector*

$$\sqrt{\frac{n}{2}}\begin{pmatrix} s_1^T(S_1S_1^T)s_1 - y_n\int x dG_{y_n}(x) \\ \vdots \\ s_1^T(S_1S_1^T)^i s_1 - y_n^i \int x^i dG_{y_n}(x) \\ s_1^T s_1 - 1 \end{pmatrix}$$

is asymptotically normal with mean 0 *and covariance matrix*

$$\begin{pmatrix} B_{11} & \cdots & B_{1i} & B_{1\,i+1} \\ \vdots & \ddots & \vdots & \vdots \\ B_{i1} & \cdots & B_{ii} & B_{i\,i+1} \\ B_{i+1\,1} & \cdots & B_{i+1\,i} & B_{i+1\,i+1} \end{pmatrix}_{(i+1)\times(i+1)} ,$$

where

$$B_{mm} = \begin{cases} \frac{\nu_4-1}{2} \cdot f^2(m) + f(2m) - f^2(m) \ , & 1 \le m \le i \ , \\ \frac{\nu_4-1}{2} \ , & m = i+1 \ , \end{cases} \tag{2.10}$$

$$B_{ml} = \begin{cases} \frac{\nu_4-1}{2} \cdot f(m)f(l) + f(m+l) - f(m)f(l) \ , & 1 \le m, l \le i \ , \\ \frac{\nu_4-1}{2} \cdot f(m) \ , & 1 \le m \le i, l = i+1 \ , \end{cases} \tag{2.11}$$

here $y_n = k/n$, $y = \lim y_n$, $f(m) := y^m \int x^m dG_y(x)$, *and* $G_y(x)$ *is the limiting spectral distribution of* $\frac{n}{k} S_1 S_1^T$.

The proof of this Proposition is in Section 4.1.

3 Two applications in spiked population models

It is well known that the empirical spectral distribution of a large-dimensional sample covariance matrix tends to the Marčenko-Pastur distribution $F_y(dx)$:

$$F_y(dx) = \frac{1}{2\pi xy} \sqrt{(x-a_y)(b_y-x)} dx, \qquad a_y \le x \le b_y \ ,$$

where $y = \lim p/n$, $a_y = (1-\sqrt{y})^2$ and $b_y = (1+\sqrt{y})^2$ under fairly general conditions, see [16]. Moreover, under a fourth moment assumption, the smallest and largest sample eigenvalues converge almost surely to the end points a_y and b_y, respectively.

While in recent empirical data analysis, there is often the case that some eigenvalues are well separated from the bulk, in order to explain such phenomenon, [13] proposed a *spiked population model*, where all the population eigenvalues equal to 1 except some fixed number of them (spikes). Clearly, the spiked population model can be considered as a finite-rank perturbation of the *null case* where all the population eigenvalues equal to 1. Then there raises the question that what's the influence of these spikes on the individual sample eigenvalues. [3] first unveiled the phase transition phenomenon in the case of complex Gaussian variables, stating that when the population spikes are above (or under) a certain threshold $1+\sqrt{y}$ (or $1-\sqrt{y}$), the corresponding extreme sample eigenvalues will jump out of the bulk (become outliers). [4] consider more general random variable: complex or real and not necessarily Gaussian and they found the same transition phenomenon. As for the central limit theorem, [3] proposed the result for the largest sample eigenvalue in the Gaussian complex case. [19] found the Gaussian limiting distribution when the population vector is real Gaussian and all the spikes of the population covariance matrix are simple. [2] established the central limit theorem for the largest as well as for the smallest sample eigenvalues under general population variables.

Beyond the sample covariance matrix, there exist many recent and related results concerning the almost sure limit as well as the central limit theorem of the extreme eigenvalue of the Wigner matrix or general Hermitian matrix perturbed by a low rank matrix. Interested reader is referred to [7], [5], [6], [8], [14], [22] and [20], for a selection of such results.

In this section, we establish two new central limit theorems for the extreme sample eigenvalues as well as sample eigenvector projections. First, Section 3.1 gives introductions on the model and some preliminary results. In Section 3.2, a joint central limit theorem is proposed for groups of packed sample eigenvalues corresponding to the spikes (primary CLT in [2] concerns only one such group). Next in Section 3.3, assuming the simple spiked case, we derive a joint CLT for the extreme sample eigenvalue and its corresponding sample eigenvector projection. Such CLT is a new result; indeed, we do not know any CLT related to spike eigenvectors from the literature. Finally, both applications are based on the general CLT for random sesquilinear forms in our Theorem 2.1.

3.1 Some notation and preliminary results

Suppose the zero-mean complex-valued random vector $x = (\xi^T, \eta^T)^T$, where $\xi = (\xi(1), \cdots, \xi(M))^T$, $\eta = (\eta(1), \cdots, \eta(p))^T$ are independent, of dimension M (fixed) and p ($p \to \infty$), respectively. And denote $x_i = (\xi_i^T, \eta_i^T)^T$ $(i = 1, \cdots, n)$ the n i.i.d. copies of x. Moreover, assume that $E||x||^4 < \infty$ and the coordinates of η are independent and identically distributed with unit variance.

The population covariance matrix of the vector x is then

$$V = \text{Cov}(x) = \begin{pmatrix} \Sigma & 0 \\ 0 & I_p \end{pmatrix} . \tag{3.1}$$

Assume Σ has the spectral decomposition:

$$\Sigma = U \,\text{diag}(\underbrace{a_1, \cdots, a_1}_{n_1}, \cdots, \underbrace{a_k, \cdots, a_k}_{n_k})\, U^*, \tag{3.2}$$

where U is an unitary matrix, the a_i's are positive and different from 1, and the n_i's satisfy $n_1 + \cdots + n_k = M$. Besides, let M_a be the number of $j's$ such that $a_j < 1 - \sqrt{y}$ (here, y is the limit of dimension to sample size ratio: $y = \lim p/n \in (0, 1)$), and let M_b be the number of $j's$ such that $a_j > 1 + \sqrt{y}$. More specifically, if we arrange the $a_i's$ in decreasing order, then Σ could be diagonalized as

$$\text{diag}(\underbrace{\underbrace{a_1, \cdots, a_1}_{n_1}, \cdots, \underbrace{a_{M_b}, \cdots, a_{M_b}}_{n_{M_b}}}_{>1+\sqrt{y}}, \cdots, \underbrace{\underbrace{a_{k-M_a+1}, \cdots, a_{k-M_a+1}}_{n_{k-M_a+1}}, \cdots, \underbrace{a_k, \cdots, a_k}_{n_k}}_{<1-\sqrt{y}}) .$$

The sample covariance matrix of x is

$$S_n = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^* ,$$

which can be partitioned as

$$S_n = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \begin{pmatrix} X_1 X_1^* & X_1 X_2^* \\ X_2 X_1^* & X_2 X_2^* \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \sum \xi_i \xi_i^* & \sum \xi_i \eta_i^* \\ \sum \eta_i \xi_i^* & \sum \eta_i \eta_i^* \end{pmatrix} ,$$

with

$$X_1 = \frac{1}{\sqrt{n}} (\xi_1, \cdots, \xi_n)_{M \times n} := \frac{1}{\sqrt{n}} \xi_{1:n} ,$$

$$X_2 = \frac{1}{\sqrt{n}} (\eta_1, \cdots, \eta_n)_{p \times n} := \frac{1}{\sqrt{n}} \eta_{1:n} .$$

Since M is fixed and $p \to \infty$, $n \to \infty$ such that $p/n \to y \in (0, 1)$, the empirical spectral distribution of the eigenvalues of S_n, as well as the one of S_{22}, converges to the Marčenko-Pastur distribution $F_y(dx)$. For real constant $\lambda \notin [a_y, b_y]$, we define the following integrals with respect to $F_y(dx)$:

$$
\begin{aligned}
m_0(\lambda) &:= \int \frac{1}{\lambda - x} F_y(dx) \ , & m_1(\lambda) &:= \int \frac{x}{\lambda - x} F_y(dx) \ , \\
m_2(\lambda) &:= \int \frac{x^2}{(\lambda - x)^2} F_y(dx) \ , & m_3(\lambda) &:= \int \frac{x}{(\lambda - x)^2} F_y(dx) \ , \\
m_4(\lambda) &:= \int \frac{1}{(\lambda - x)^2} F_y(dx) \ , & m_5(\lambda) &:= \int \frac{x}{(\lambda - x)^3} F_y(dx) \ , \\
m_6(\lambda) &:= \int \frac{x^2}{(\lambda - x)^4} F_y(dx) \ , & m_7(\lambda) &:= \int \frac{x^2}{(\lambda - x)^3} F_y(dx) \ .
\end{aligned} \tag{3.3}
$$

Let $l_1 \geq l_2 \geq \cdots \geq l_p$ be the eigenvalues of S_n. Let $s_j = n_1 + \cdots + n_j$ for $1 \leq j \leq M_b$ or $k - M_a + 1 \leq j \leq k$. [4] derive the almost sure limit of those extreme sample eigenvalues. They have proven that for each $m \in \{1, \cdots, M_b\}$ or $m \in \{k - M_a + 1, \cdots, k\}$ and $s_{m-1} < i \leq s_m$,

$$
l_i \to \lambda_m = \phi(a_m) := a_m + \frac{y a_m}{a_m - 1}
$$

almost surely. In other words, if a spike eigenvalue a_m lies outside the interval $[1 - \sqrt{y}, 1 + \sqrt{y}]$, then the n_m-packed sample eigenvalues $\{l_i, i \in J_m\}$ (associated to a_m) converge to the limit λ_m, which is outside the support of the M-P distribution $[a_y, b_y]$ (here, we denote $J_m = (s_{m-1}, s_m]$ when $m \in \{1, \cdots, M_b\}$ or $m \in \{k - M_a + 1, \cdots, k\}$).

Recently [2] derives the CLT for those extreme sample eigenvalues. More specifically, let $\delta_{n,i} := \sqrt{n}(l_i - \lambda_m)$, where $m \in \{1, \cdots, M_b\}$ or $m \in \{k - M_a + 1, \cdots, k\}$, $i \in J_m$, and $\lambda_m = \phi(a_m) \notin [a_y, b_y]$ as defined before. They have proven that $\delta_{n,i}$ tends to the solution v of the following equation:

$$
\Big| - \big[U^* R_n(\lambda_m) U\big]_{mm} + v\big(1 + y m_3(\lambda_m) a_m\big) I_{n_m} + o_n(1) \Big| = 0 \ , \tag{3.4}
$$

here $|\cdot|$ stands for determinant, $\big[U^* R_n(\lambda_m) U\big]_{mm}$ is the m-th diagonal block of $U^* R_n(\lambda_m) U$ corresponding to the index $\{u, v \in J_m\}$, and

$$
\begin{aligned}
R_n(\lambda) &= \frac{1}{\sqrt{n}} \Big\{ \xi_{1:n}\big(I + A_n(\lambda)\big)\xi_{1:n}^* - \Sigma tr\big(I + A_n(\lambda)\big) \Big\} \ , \\
A_n(\lambda) &= X_2^* (\lambda I - X_2 X_2^*)^{-1} X_2 \ .
\end{aligned}
$$

Let $R(\lambda)$ denote the $M \times M$ matrix limit of $R_n(\lambda)$, and $\widetilde{R}(\lambda) := U^* R(\lambda) U$. According to (3.4), it says that $\delta_{n,i}$ tends to an eigenvalue of the matrix $(1 + y m_3(\lambda_m) a_m)^{-1} [\widetilde{R}(\lambda_m)]_{mm}$. Besides, since the index i is arbitrary over J_m, all the J_m random variables $\sqrt{n}\{l_i - \lambda_m, \ i \in J_m\}$ converge almost surely to the set of eigenvalues of this matrix. The following theorem in [2] identifies the covariance of the elements within the limit matrix $R(\lambda)$. For simplicity, we only consider the real case in all the following unless otherwise noted.

Proposition 3.1. *[[2]] Assume that the variables ξ and η are real, then the random matrix $R = R_{ij}$ is symmetric, with zero-mean Gaussian entries, having the following covariance function: for $1 \leq i \leq j \leq M$ and $1 \leq i^{'} \leq j^{'} \leq M$*

$$
\begin{aligned}
&\textit{Cov} \ (R(i,j), R(i^{'}, j^{'})) \\
&= w \Big\{ E[\xi(i)\xi(j)\xi(i^{'})\xi(j^{'})] - \Sigma_{ij}\Sigma_{i^{'}j^{'}} \Big\} + (\theta - w)\Sigma_{ij^{'}}\Sigma_{i^{'}j} \\
&\quad + (\theta - w)\Sigma_{ii^{'}}\Sigma_{jj^{'}} \ ,
\end{aligned}
$$

where the constants θ and w are defined as follows:

$$\theta = 1 + 2ym_1(\lambda) + ym_2(\lambda) ,$$
$$w = 1 + 2ym_1(\lambda) + \left(\frac{y(1+m_1(\lambda))}{\lambda - y(1+m_1(\lambda))}\right)^2 .$$

3.2 Application 1: Asymptotic joint distribution of two groups of extreme sample eigenvalues in the spiked population model

In this subsection, we consider the asymptotic joint distribution of two groups of extreme sample eigenvalues, say, $\{l_i,\ i \in J_m\}$ and $\{l_{i'},\ i^{'} \in J_{m'}\}$ ($m \neq m^{'}$) when Σ has the structure (3.2), namely the random vector

$$\begin{pmatrix} \{\sqrt{n}(l_i - \lambda_m),\ i \in J_m\} \\ \{\sqrt{n}(l_{i'} - \lambda_{m'}),\ i^{'} \in J_{m'}\} \end{pmatrix} .$$

Following the work of [2], we know that this $n_m + n_{m'}$ dimensional random vector converges to the eigenvalues of the symmetric $(n_m + n_{m'}) \times (n_m + n_{m'})$ random matrix

$$\begin{pmatrix} \frac{[\widetilde{R}(\lambda_m)]_{mm}}{1+ym_3(\lambda_m)a_m} & 0 \\ 0 & \frac{[\widetilde{R}(\lambda_{m'})]_{m'm'}}{1+ym_3(\lambda_{m'})a_{m'}} \end{pmatrix} . \tag{3.5}$$

Here, this random matrix (3.5) has two diagonal blocks with dimension n_m and $n_{m'}$, respectively. The covariance function of the elements within each block has been fully identified by [2], see Proposition 3.1. But if we consider them as a whole, there's still need to explore the covariance between the elements from the different two blocks $[\widetilde{R}(\lambda_m)]_{mm}$ and $[\widetilde{R}(\lambda_{m'})]_{m'm'}$.

We establish such a covariance function in Theorem 3.2 when the observation vector x is real with the help of our Corollary 2.2. However, it can also be generalized to the complex case by considering the real and imaginary parts as two independent real random variables with the help of our Theorem 2.1, readers who are interested in this can refer to [2] (see the proof of their Proposition 3.2).

3.2.1 Main result

Theorem 3.2. *Assume that the variables ξ and η are real, then the two diagonal blocks of the $2M \times 2M$ random matrix*

$$\begin{pmatrix} R(\lambda_m) & 0 \\ 0 & R(\lambda_{m'}) \end{pmatrix} \tag{3.6}$$

are symmetric, having zero-mean Gaussian entries, with the following covariance function between each other: for $1 \le i \le j \le M$ and $1 \le i^{'} \le j^{'} \le M$, we have

$$\begin{aligned} &\textit{Cov}\ (R(\lambda_m)(i,j), R(\lambda_{m'})(i^{'},j^{'})) \\ &= w(m,m^{'})\Big\{\mathbb{E}[\xi(i)\xi(j)\xi(i^{'})\xi(j^{'})] - \Sigma_{ij}\Sigma_{i'j'}\Big\} \\ &\quad + \big(\theta(m,m^{'}) - w(m,m^{'})\big)\Sigma_{ij'}\Sigma_{i'j} \\ &\quad + \big(\theta(m,m^{'}) - w(m,m^{'})\big)\Sigma_{ii'}\Sigma_{jj'} , \end{aligned} \tag{3.7}$$

where

$$\theta(m,m^{'}) = 1 + ym_1(\lambda_m) + ym_1(\lambda_{m'}) + y\left(\frac{\lambda_{m'}}{\lambda_m - \lambda_{m'}}m_1(\lambda_{m'}) + \frac{\lambda_m}{\lambda_{m'} - \lambda_m}m_1(\lambda_m)\right) ,$$

$$w(m,m^{'}) = 1 + ym_1(\lambda_m) + ym_1(\lambda_{m'}) + \frac{y^2(1+m_1(\lambda_m))(1+m_1(\lambda_{m'}))}{(\lambda_m - y(1+m_1(\lambda_m)))(\lambda_{m'} - y(1+m_1(\lambda_{m'})))} \ .$$

Remark 3.3. *If we restrict the index* (i,j) *to the region* $J_m \times J_m$ *and* $(i^{'}, j^{'})$ *to* $J_{m'} \times J_{m'}$*, we can get the covariance function between the two blocks of* (3.5)*. And it should be noticed that the two regions* $J_m \times J_m$ *and* $J_{m'} \times J_{m'}$ *do not intersect with each other.*

Remark 3.4. *In general, the covariance of the elements from two blocks are not independent asymptotically, that is Cov* $(R(\lambda_m)(i,j), R(\lambda_{m'})(i^{'}, j^{'})) \neq 0$*. Notice that same phenomenon also exists in the Wigner case, for example, see Theorem 2.11 in [14].*

Remark 3.5. *If the coordinates* $\{\xi(i)\}$ *of* ξ *are independent (thus,* Σ *is diagonal and* $U = I_M$*), [2] has already proved that the covariance matrix within each diagonal block in* (3.6) *is diagonal; in other words, the Gaussian matrix* $R(\lambda_m)$ *and* $R(\lambda_{m'})$ *are both made with independent entries. And by noting that the regions* $J_m \times J_m$ *and* $J_{m'} \times J_{m'}$ *are disjoint, the only covariance function that may exist between the two blocks is Cov*$(R(\lambda_m)(i,i), R(\lambda_{m'})(i^{'}, i^{'}))$ *(*$i \in J_m,\ i^{'} \in J_{m'}$*). Using* (3.7) *and the fact that* $\{\xi(i)\}$ *are independent, we have*

$$\begin{aligned}
&\text{Cov } (R(\lambda_m)(i,i), R(\lambda_{m'})(i^{'}, i^{'})) \\
&= w(m,m^{'})\Big\{\mathbb{E}[\xi(i)^2\xi(i^{'})^2] - \Sigma_{ii}\Sigma_{i^{'}i^{'}}\Big\} + 2\big(\theta(m,m^{'}) - w(m,m^{'})\big)\big(\Sigma_{ii'}\big)^2 \\
&= w(m,m^{'})\Big\{\Sigma_{ii}\Sigma_{i^{'}i^{'}} - \Sigma_{ii}\Sigma_{i^{'}i^{'}}\Big\} + 2\big(\theta(m,m^{'}) - w(m,m^{'})\big)\big(\Sigma_{ii'}\big)^2 \\
&= 0 \ ,
\end{aligned}$$

which means that the two diagonal blocks in (3.5) *are independent. Besides, [2] have already pointed out the variances within each block:*

$$Var(R(i,j)) = \theta\Sigma_{ii}\Sigma_{jj}, \quad i < j \tag{3.8}$$

$$Var(R(i,i)) = w(E\xi(i)^4 - 3\Sigma_{ii}^2) + 2\theta\Sigma_{ii}^2 \ . \tag{3.9}$$

Therefore, if $\{\xi(i)\}$ *are independent, then any two groups of packed extreme sample eigenvalues* $\{\sqrt{n}(l_i - \lambda_m), i \in J_m\}$ *and* $\{\sqrt{n}(l_{i'} - \lambda_{m'}), i^{'} \in J_{m'}\}$ *are asymptotically independent, converging to the eigenvalues of the Gaussian random matrices* $\frac{1}{1+ym_3(\lambda_m)a_m}[R(\lambda_m)]_{mm}$ *and* $\frac{1}{1+ym_3(\lambda_{m'})a_{m'}}[R(\lambda_{m'})]_{m'm'}$*, respectively. And both the Gaussian random matrices are made with independent entries, with a fully identified variance function given by* (3.8) *and* (3.9)*. Moveover, if the observations are Gaussian,* (3.9) *reduces to* $Var(R(i,i)) = 2\theta\Sigma_{ii}^2$*.*

3.2.2 Conditions that two groups of packed extreme sample eigenvalues are pairwise independent

An interesting question in the asymptotical analysis of spiked eigenvalues is to know whether two groups of packed extreme sample eigenvalues are asymptotically pairwise independent. In Remark 3.5, we have seen that when $\{\xi(i)\}$ are independent, $\{\sqrt{n}(l_i - \lambda_m), i \in J_m\}$ and $\{\sqrt{n}(l_{i'} - \lambda_{m'}), i^{'} \in J_{m'}\}$ are asymptotically independent.

We aim to relax the independent restriction of $\{\xi(i)\}$ under the condition that all the eigenvalues of Σ are simple, that is, Σ has the spectral decomposition:

$$\Sigma = U \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_M \end{pmatrix} U^* \ ,$$

where the $a_i's$ are arranged in decreasing order. We discuss the condition that when the extreme sample eigenvalues are pairwise independent, asymptotically.

Let l_i, l_j denote the extreme sample eigenvalues correspond to two different spikes a_i and a_j, where $a_i, a_j \notin [1-\sqrt{y}, 1+\sqrt{y}]$. Then, the two-dimensional random vector

$$\begin{pmatrix} \delta_{n,i} \\ \delta_{n,j} \end{pmatrix} = \begin{pmatrix} \sqrt{n}\big(l_i - \lambda_i\big) \\ \sqrt{n}\big(l_j - \lambda_j\big) \end{pmatrix}$$

converges to the eigenvalues of the following random matrix:

$$\begin{pmatrix} \frac{1}{1+ym_3(\lambda_i)a_i}[\widetilde{R}(\lambda_i)]_{ii} & 0 \\ 0 & \frac{1}{1+ym_3(\lambda_j)a_j}[\widetilde{R}(\lambda_j)]_{jj} \end{pmatrix} .$$

Since all the eigenvalues of Σ are simple, the multiplicity numbers n_i and n_j both equal to 1. Therefore, $[\widetilde{R}(\lambda_i)]_{ii}$ and $[\widetilde{R}(\lambda_j)]_{jj}$ are now two Gaussian random variables (actually, they are the (i,i)-th and (j,j)-th elements of the $M \times M$ Gaussian random matrices $\widetilde{R}(\lambda_i)$ and $\widetilde{R}(\lambda_j)$, respectively, denoted as $\widetilde{R}(\lambda_i)(i,i)$ and $\widetilde{R}(\lambda_j)(j,j)$). As a result,

$$\begin{pmatrix} \delta_{n,i} & \delta_{n,j} \end{pmatrix}^T$$

actually converges to the Gaussian random vector

$$\begin{pmatrix} \frac{1}{1+ym_3(\lambda_i)a_i}\widetilde{R}(\lambda_i)(i,i) \\ \frac{1}{1+ym_3(\lambda_j)a_j}\widetilde{R}(\lambda_j)(j,j) \end{pmatrix}$$

with

$$Var\ \big(R(\lambda_i)(i,i)\big) = w(i)\big\{\mathbb{E}\big[\xi(i)^4\big] - \Sigma_{ii}^2\big\} + 2\big(\theta(i) - w(i)\big)\Sigma_{ii}^2 \ , \tag{3.10}$$

$$Var\ \big(R(\lambda_j)(j,j)\big) = w(j)\big\{\mathbb{E}\big[\xi(j)^4\big] - \Sigma_{jj}^2\big\} + 2\big(\theta(j) - w(j)\big)\Sigma_{jj}^2 \ , \tag{3.11}$$

$$\begin{aligned} \text{Cov}\ \big(R(\lambda_i)(i,i), R(\lambda_j)(j,j)\big) &= w(i,j)\big\{\mathbb{E}\big[\xi(i)^2\xi(j)^2\big] - \Sigma_{ii}\Sigma_{jj}\big\} \\ &\quad + 2\big(\theta(i,j) - w(i,j)\big)\Sigma_{ij}^2 \ , \end{aligned} \tag{3.12}$$

where

$$\theta(i) = 1 + 2ym_1(\lambda_i) + ym_2(\lambda_i) \ ,$$

$$w(i) = 1 + 2ym_1(\lambda_i) + \left(\frac{y(1+m_1(\lambda_i))}{\lambda_i - y(1+m_1(\lambda_i))}\right)^2$$

are given in [2]. From the definitions of $w(i,j)$ and $\theta(i,j)$ in Theorem 3.2, taking the fact that $m_1(\lambda_i) = 1/(a_i - 1)$ (see Lemma 5.1) into consideration, we have,

$$\begin{aligned} w(i,j) &= 1 + ym_1(\lambda_i) + ym_1(\lambda_j) \\ &\quad + \frac{y^2\big(1+m_1(\lambda_i)\big)\big(1+m_1(\lambda_j)\big)}{\big(\lambda_i - y(1+m_1(\lambda_i))\big)\big(\lambda_j - y(1+m_1(\lambda_j))\big)} \\ &= 1 + \frac{y}{a_i - 1} + \frac{y}{a_j - 1} + \frac{y^2}{(a_i-1)(a_j-1)} \\ &= \frac{(y + a_i - 1)(y + a_j - 1)}{(a_i - 1)(a_j - 1)} \ , \end{aligned}$$

$$\begin{aligned}\theta(i,j) - w(i,j) &= y\left(\frac{\lambda_j}{\lambda_i - \lambda_j} m_1(\lambda_j) + \frac{\lambda_i}{\lambda_j - \lambda_i} m_1(\lambda_i)\right) \\ &\quad - \frac{y^2}{(a_i - 1)(a_j - 1)} \\ &= y \cdot \frac{(y + a_i - 1)(y + a_j - 1)}{(a_i - 1)(a_j - 1)\big[(a_i - 1)(a_j - 1) - y\big]} \ .\end{aligned}$$

The values of $w(i,j)$ will always be positive whenever $a_i, a_j \notin [1 - \sqrt{y}, 1 + \sqrt{y}]$, while $\theta(i,j) - w(i,j)$ will be negative if $a_i > 1 + \sqrt{y}$ and $0 < a_j < 1 - \sqrt{y}$ (corresponding to one extreme large and one extreme small sample eigenvalues), and positive if $a_i, a_j > 1 + \sqrt{y}$ or $0 < a_i, a_j < 1 - \sqrt{y}$ (corresponding to two extreme large or two extreme small sample eigenvalues).

Therefore, if any two extreme large (or small) sample eigenvalues are mutually independent (equivalent to the condition that $\mathrm{Cov}(R(\lambda_i)(i,i), R(\lambda_j)(j,j)) = 0$), a sufficient and necessary condition is

$$\mathbb{E}\big[\xi(i)^2 \xi(j)^2\big] - \Sigma_{ii}\Sigma_{jj} = 0 \ ,$$

and

$$\mathbb{E}[\xi(i)\xi(j)] = 0 \ (= \mathbb{E}\xi(i)\mathbb{E}\xi(j)) \ ;$$

another way of saying this is

$$(*) \quad \begin{cases} \mathrm{Cov}\big(\xi(i), \xi(j)\big) = 0 \quad (\Sigma \textit{ is diagonal or } U = I_M)\ , \textit{and} \\ \mathrm{Cov}\big(\xi(i)^2, \xi(j)^2\big) = 0 \end{cases} .$$

Obviously, when $\{\xi(i)\}$ are independent, the condition $(*)$ is satisfied.

We consider a special case that the observations are Gaussian, with a diagonal population covariance matrix. This model satisfies condition $(*)$. It is due to the fact that when the observations are Gaussian, uncorrelation between $\xi(i)$ and $\xi(j)$ implies independence, which further implies $\xi(i)^2$ and $\xi(j)^2$ are uncorrelated. Therefore, if the observations are Gaussian and the population covariance matrix is diagonal, then any two extreme large (or small) sample eigenvalues are mutually independent. Furthermore, we can derive explicitly the joint distribution of $\delta_{n,i}$ and $\delta_{n,j}$. According to (3.10), (3.11) and (3.12), we have a much more simplified form due to the Gaussian assumption:

$$\begin{aligned}&Var\ \big(R(\lambda_i)(i,i)\big) = 2\theta(i)a_i^2\ , \\ &Var\ \big(R(\lambda_j)(j,j)\big) = 2\theta(j)a_j^2\ , \\ &\mathrm{Cov}\ \big(R(\lambda_i)(i,i), R(\lambda_j)(j,j)\big) = 0\ ,\end{aligned} \tag{3.13}$$

where $\theta(i) = \frac{(a_i - 1 + y)^2}{(a_i - 1)^2 - y}$ and $\theta(j) = \frac{(a_j - 1 + y)^2}{(a_j - 1)^2 - y}$ by definition. And using the expression $m_3(\lambda) = \frac{1}{(a-1)^2 - y}$ (see Lemma 5.1), we finally derive the asymptotic joint distribution:

$$\begin{pmatrix} \sqrt{n}\big(l_i - \lambda_i\big) \\ \sqrt{n}\big(l_j - \lambda_j\big) \end{pmatrix} \Longrightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{2a_i^2[(a_i-1)^2 - y]}{(a_i - 1)^2} & 0 \\ 0 & \frac{2a_j^2[(a_j-1)^2 - y]}{(a_j - 1)^2} \end{pmatrix}\right) .$$

But, if we only assume Σ is diagonal, and no Gaussian assumptions are made, things are different. One such example is that $\xi(i)$ and $\xi(j)$ come from the uniform distribution inside the ellipse:

$$\frac{\xi(i)^2}{16} + \frac{\xi(j)^2}{36} \le 1,$$

one can check that $\mathbb{E}\xi(i)\xi(j) = \mathbb{E}\xi(i) \cdot \mathbb{E}\xi(j) = 0$, but $\mathbb{E}\xi(i)^2 = 4$, $\mathbb{E}\xi(j)^2 = 9$ and $\mathbb{E}\xi(i)^2\xi(j)^2 = 24$, that is $\mathbb{E}\xi(i)^2\xi(j)^2 \neq \mathbb{E}\xi(i)^2 \cdot \mathbb{E}\xi(j)^2$, therefore, condition $(*)$ is not satisfied. From this example, we see there could happen that although $\xi(i)$ and $\xi(j)$ are uncorrelated, $\xi(i)^2$ and $\xi(j)^2$ are correlated. And in such a case, even though the population covariance matrix is diagonal, the two extreme large (or small) eigenvalues of the sample covariance matrix may actually have correlation between each other.

A small simulation is conducted below to check this covariance formula according to the two cases mentioned above. The dimension p is fixed to be 200 and the sample size n is fixed to be 300. We choose two spikes $a_1 = 9$ and $a_2 = 4$, which are both larger than the critical value $1 + \sqrt{y}$ $(= 1 + \sqrt{2/3})$. We repeat 10000 times to calculate the empirical covariance value between the largest (l_1) and the second largest (l_2) sample eigenvalues. The first case is the two-dimensional multivariate Gaussian vector $(\xi(1), \xi(2))^T$, which has a joint distribution

$$\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}9 & 0\\0 & 4\end{pmatrix}\right) .$$

According to (3.13), the theoretical covariance value between l_1 and l_2 should be 0, and the empirical covariance value from the 10000 sample simulated turns out to be 0.0019. The second case is the aforementioned uniform distribution inside the ellipse: $\xi(1)^2/36 + \xi(2)^2/16 \leq 1$. This time, the theoretical covariance value between l_1 and l_2 could be calculated as -0.0366 according to (3.12), and the empirical covariance value from the 10000 sample simulated turns out to be -0.0371. The two errors are both smaller than the order $O(1/\sqrt{10000})$ under both cases.

3.3 Application 2: Asymptotic joint distribution of the largest sample eigenvalue and its corresponding sample eigenvector projection

In this subsection, we consider the joint central limit theorem of extreme sample eigenvalue and its corresponding sample eigenvector projection, which may find applications in principal component scores, where both the eigenvalue and its eigenvector are involved, see [15].

Let the population covariance matrix be diagonal with k simple spikes:

$$V = \mathrm{diag}(\underbrace{a_1, \cdots, a_k}_{k}, \underbrace{1, \cdots, 1}_{p}) ,$$

where now the Σ in (3.1) reduces to a diagonal matrix $\mathrm{diag}(a_1, \cdots, a_k)$ with all the diagonal elements a_i larger than the critical value $1 + \sqrt{y}$. The sample covariance matrix S_n is also partitioned as before:

$$S_n = \begin{pmatrix}S_{11} & S_{12}\\S_{21} & S_{22}\end{pmatrix} = \begin{pmatrix}X_1X_1^* & X_1X_2^*\\X_2X_1^* & X_2X_2^*\end{pmatrix} = \frac{1}{n}\begin{pmatrix}\sum \xi_i\xi_i^* & \sum \xi_i\eta_i^*\\\sum \eta_i\xi_i^* & \sum \eta_i\eta_i^*\end{pmatrix} ,$$

with

$$X_1 = \frac{1}{\sqrt{n}}(\xi_1, \cdots, \xi_n)_{k\times n} := \frac{1}{\sqrt{n}}\xi_{1:n} ,$$

$$X_2 = \frac{1}{\sqrt{n}}(\eta_1, \cdots, \eta_n)_{p\times n} := \frac{1}{\sqrt{n}}\eta_{1:n} ,$$

which are mutually independent. And we denote $\nu_4(i) = \mathbb{E}\xi(i)^4/a_i^2 - 3$ for $i = 1, \cdots, k$ as the kurtosis coefficient of the i-th coordinate of ξ.

Now suppose l_i is an extreme eigenvalue of S_n, converging to the value $\lambda_i = \phi(a_i) = a_i + ya_i/(a_i - 1)$ and let $(u_i, v_i)^T$ be the corresponding sample eigenvector with u_i its first k components and v_i the remaining p components. We derive the following central limit theorem that establishes the asymptotic joint distribution of the extreme sample eigenvalue l_i and its corresponding sample eigenvector projection $u_i(i)^2$ (here $u_i(i)$ stands for the i-th element of the $k \times 1$ vector u_i). Notice that the population eigenvector corresponding to the spike a_i is simply $e_i = (0, \cdots, 1, \cdots, 0)^T$, the i-th standard canonical basis vector. Therefore, $u_i(i)$ represents the inner product between the sample eigenvector $(u_i, v_i)^T$ and the population one e_i.

Theorem 3.6.

$$\begin{pmatrix} \sqrt{n}\left(u_i(i)^2 - \frac{(a_i-1)^2-y}{(a_i-1)(a_i-1+y)}\right) \\ \sqrt{n}(l_i - \lambda_i) \end{pmatrix} \Longrightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix} \right) ,$$

where

$$\begin{aligned} v_{11} &= \frac{a_i^2 y^2 (a_i^2 + y - 1)^2}{(a_i-1)^4 (a_i - 1 + y)^4} \nu_4(i) + \frac{2a_i^2 y((a_i + y - 1)^2 + y a_i^2)}{((a_i-1)^2 - y)(a_i - 1 + y)^4} , \\ v_{12} &= \frac{y a_i^2 (a_i^2 - 1 + y)((a_i-1)^2 - y)}{(a_i-1)^6 (a_i - 1 + y)^4} \nu_4(i) + \frac{2a_i^3 y}{(a_i-1)(a_i-1+y)^2} , \\ v_{22} &= \frac{a_i^2((a_i-1)^2 - y)^2}{(a_i-1)^4} \nu_4(i) + \frac{2a_i^2((a_i-1)^2 - y)}{(a_i-1)^2} . \end{aligned}$$

Remark 3.7. *If the observations are Gaussian ($\nu_4(i) = 0$ for $i = 1, \cdots, k$), then the three values above are simplified to be:*

$$\begin{aligned} v_{11} &= \frac{2a_i^2 y((a_i + y - 1)^2 + y a_i^2)}{((a_i-1)^2 - y)(a_i - 1 + y)^4} , \\ v_{12} &= \frac{2a_i^3 y}{(a_i-1)(a_i-1+y)^2} , \\ v_{22} &= \frac{2a_i^2((a_i-1)^2 - y)}{(a_i-1)^2} . \end{aligned}$$

Remark 3.8. *Trivially, the following central limit theorem of the eigenvector projection holds*

$$\sqrt{n}\left(u_i(i)^2 - \frac{(a_i-1)^2 - y}{(a_i-1)(a_i-1+y)}\right) \longrightarrow \mathcal{N}(0, v_{11}) .$$

In particular,

$$u_i(i)^2 \xrightarrow{p} \frac{(a_i-1)^2 - y}{(a_i-1)(a_i-1+y)} .$$

Observe that this limit $\in (0,1)$. In particular, the sample eigenvector does not converge to the population eigenvector; only their angle tends to a limit. Notice that the limit of the angle has already been established by [19] for the Gaussian case and [6] on somewhat different but closely related random matrix models with a finite-rank perturbation.

4 Proof of Proposition 2.4, Theorem 3.2 and 3.6

4.1 Proof of Proposition 2.4

Proof. We give a short proof of Proposition 2.4 using our Theorem 2.3. Let

$$\begin{aligned} U^{(m)} &= \frac{1}{\sqrt{n}}\left[s_1^T (S_1 S_1^T)^m s_1 - \frac{1}{n}\operatorname{tr}(S_1 S_1^T)^m \right] , \ 1 \le m \le i , \\ U^{(i+1)} &= \frac{1}{\sqrt{n}}\left[s_1^T s_1 - 1 \right] , \end{aligned}$$

so we have

$$\begin{aligned}
&k = i+1\ ,\\
&X(1) = s_1 := \frac{1}{\sqrt{n}}(v_{11},\cdots,v_{1n})^T = Y(1)\ ,\ (K=1)\ ,\\
&\rho(l) = \mathbb{E}x_{11}y_{11} = \frac{1}{n}\ ,\\
&A_1 = \frac{1}{n^2}(\nu_4-1),\ A_2 = A_3 = \frac{1}{n^2}\ ,
\end{aligned}$$

and the matrix

$$A^{(m)} := \begin{cases} (S_1S_1^T)^m\ , & 1\le m\le i \\ I_n\ , & m=i+1 \end{cases}\ .$$

Then,

$$\begin{aligned}
\tau_m = \theta_m = \lim_{n\to\infty}\frac{1}{n}\operatorname{tr}[A^{(m)}]^2 &= \begin{cases} \lim\limits_{n\to\infty}\frac{1}{n}\operatorname{tr}(S_1S_1^T)^{2m}\ , & 1\le m\le i \\ 1\ , & m=i+1 \end{cases}\\
&= \begin{cases} y^{2m}\int x^{2m}dG_y(x)\ , & 1\le m\le i \\ 1\ , & m=i+1 \end{cases}\ ,
\end{aligned}$$

$$\begin{aligned}
\tau_{mm'} = \theta_{mm'} &= \begin{cases} \lim\limits_{n\to\infty}\frac{1}{n}\operatorname{tr}\left[(S_1S_1^T)^m(S_1S_1^T)^{m'}\right]\ , & 1\le m,m'\le i\ , \\ \lim\limits_{n\to\infty}\frac{1}{n}\operatorname{tr}(S_1S_1^T)^{m'}\ , & m=i+1, 1\le m'\le i\ , \end{cases}\\
&= \begin{cases} y^{m+m'}\int x^{m+m'}dG_y(x)\ , & 1\le m,m'\le i\ , \\ y^{m'}\int x^{m'}dG_y(x)\ , & m=i+1, 1\le m'\le i\ , \end{cases}
\end{aligned}$$

$$\begin{aligned}
w_m = \lim_{n\to\infty}\frac{1}{n}\operatorname{tr}[A^{(m)}\circ A^{(m)}] &= \begin{cases} (\lim\limits_{n\to\infty}\frac{1}{n}\operatorname{tr}(S_1S_1^T)^m)^2\ , & 1\le m\le i\ , \\ (\lim\limits_{n\to\infty}\frac{1}{n}\operatorname{tr} I_n)^2\ , & m=i+1\ , \end{cases}\\
&= \begin{cases} y^{2m}(\int x^m dG_y(x))^2\ , & 1\le m\le i\ , \\ 1\ , & m=i+1\ , \end{cases}
\end{aligned}$$

$$\begin{aligned}
w_{mm'} &= \lim_{n\to\infty}\frac{1}{n}\operatorname{tr}[A^{(m)}\circ A^{(m')}]\\
&= \begin{cases} \lim\limits_{n\to\infty}\frac{1}{n}\operatorname{tr}(S_1S_1^T)^m\cdot \lim_{n\to\infty}\frac{1}{n}\operatorname{tr}(S_1S_1^T)^{m'}\ , & 1\le m,m'\le i\ , \\ \lim\limits_{n\to\infty}\frac{1}{n}\operatorname{tr}(S_1S_1^T)^m\ , & m'=i+1\ , 1\le m\le i\ , \end{cases}\\
&= \begin{cases} y^m\int x^m dG_y(x)\cdot y^{m'}\int x^{m'}dG_y(x)\ , & 1\le m,m'\le i\ , \\ y^m\int x^m dG_y(x)\ , & m'=i+1, 1\le m\le i\ , \end{cases}
\end{aligned}$$

Combine all that above with (2.8) and (2.9) leads to the result.

□

4.2 Proof of Theorem 3.2

Proof. We prove this result with the help of Corollary 2.2. Consider

$$\begin{pmatrix} \frac{1}{\sqrt{n}}u(i)(I+A_n(\lambda_m))u(j)^T \\ \frac{1}{\sqrt{n}}u(i')(I+A_n(\lambda_{m'}))u(j')^T \end{pmatrix}_{1\le i\le j\le M,\ 1\le i'\le j'\le M}\ ,$$

with $u(i) = \big(\xi_1(i), \cdots, \xi_n(i)\big)$. Moreover, we define $X(l) = u(i)^T, Y(l) = u(j)^T$, $X(l^{'}) = u(i^{'})^T, Y(l^{'}) = u(j^{'})^T$, with $l = (i,j)$, $l^{'} = (i^{'}, j^{'})$, where l and $l^{'}$ both have $K = \frac{M(M+1)}{2}$ options. Recall the definition of R_n, we have:

$$R_n(\lambda_m) = \frac{1}{\sqrt{n}}\Big\{\xi_{1:n}(I + A_n(\lambda_m))\xi_{1:n}^* - \Sigma\,\mathrm{tr}(I + A_n(\lambda_m))\Big\} ,$$
$$R_n(\lambda_{m'}) = \frac{1}{\sqrt{n}}\Big\{\xi_{1:n}(I + A_n(\lambda_{m'}))\xi_{1:n}^* - \Sigma\,\mathrm{tr}(I + A_n(\lambda_{m'}))\Big\} .$$

By applying Corollary 2.2, we have

$$\mathrm{Cov}(R(\lambda_m)(i,j), R(\lambda_{m'})(i^{'}, j^{'})) = w_3 A_1 + (\tau_3 - w_3)A_2 + (\theta_3 - w_3)A_3 .$$

We specify these values in the following:

$$\begin{aligned}
w_3 &= \lim_{n\to\infty} \frac{1}{n}\sum_{u=1}^{n}(I + A_n(\lambda_m))_{uu}(I + A_n(\lambda_{m'}))_{uu} \\
&= 1 + \lim_{n\to\infty}\frac{1}{n}\sum_{u=1}^{n}\Big(A_n(\lambda_m)_{uu} + A_n(\lambda_{m'})_{uu} + A_n(\lambda_m)_{uu}A_n(\lambda_{m'})_{uu}\Big) \\
&= 1 + ym_1(\lambda_m) + ym_1(\lambda_{m'}) + \frac{y^2\big(1+m_1(\lambda_m)\big)\big(1+m_1(\lambda_{m'})\big)}{\big(\lambda_m - y(1+m_1(\lambda_m))\big)\big(\lambda_{m'} - y(1+m_1(\lambda_{m'}))\big)} , \\
&= w(m, m^{'}) , \\
\tau_3 &= \theta_3 = \lim_{n\to\infty}\frac{1}{n}\sum_{u,v=1}^{n}(I + A_n(\lambda_m))_{uv}(I + A_n(\lambda_{m'}))_{vu} \\
&= \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}\big(I + A_n(\lambda_m))(I + A_n(\lambda_{m'})\big) \\
&= 1 + ym_1(\lambda_m) + ym_1(\lambda_{m'}) + y\int\frac{x^2}{(\lambda_{m'} - x)(\lambda_m - x)}F_y(dx) \\
&= 1 + ym_1(\lambda_m) + ym_1(\lambda_{m'}) + y\left(\frac{\lambda_{m'}}{\lambda_m - \lambda_{m'}}m_1(\lambda_{m'}) + \frac{\lambda_m}{\lambda_{m'} - \lambda_m}m_1(\lambda_m)\right) , \\
&= \theta(m, m^{'}) ,
\end{aligned}$$

where we have used Lemma 6.1. in [2]; and

$$\begin{aligned}
A_1 &= \mathbb{E}(\overline{x}_{l1}y_{l1}\overline{x}_{l'1}y_{l'1}) - \rho(l)\rho(l^{'}) = \mathbb{E}[\xi(i)\xi(j)\xi(i^{'})\xi(j^{'})] - \Sigma_{ij}\Sigma_{i'j'} , \\
A_2 &= \mathbb{E}(\overline{x}_{l1}\overline{x}_{l'1})\mathbb{E}(y_{l1}y_{l'1}) = \mathbb{E}[\xi(i)\xi(i^{'})]\mathbb{E}[\xi(j)\xi(j^{'})] , \\
A_3 &= \mathbb{E}(\overline{x}_{l1}y_{l'1})\mathbb{E}(\overline{x}_{l'1}y_{l1}) = \mathbb{E}[\xi(i)\xi(j^{'})]\mathbb{E}[\xi(j)\xi(i^{'})] .
\end{aligned}$$

Combine all these, we have

$$\begin{aligned}
&\mathrm{Cov}(R(\lambda_m)(i,j), R(\lambda_{m'})(i^{'}, j^{'})) \\
&= w(m,m^{'})\Big\{\mathbb{E}[\xi(i)\xi(j)\xi(i^{'})\xi(j^{'})] - \Sigma_{ij}\Sigma_{i'j'}\Big\} \\
&\quad + \big(\theta(m,m^{'}) - w(m,m^{'})\big)\mathbb{E}[\xi(i)\xi(j^{'})]\mathbb{E}[\xi(i^{'})\xi(j)] \\
&\quad + \big(\theta(m,m^{'}) - w(m,m^{'})\big)\mathbb{E}[\xi(i)\xi(i^{'})]\mathbb{E}[\xi(j)\xi(j^{'})] .
\end{aligned}$$

The proof of Theorem 3.2 is complete. □

4.3 Proof of Theorem 3.6

Proof. Since l_i is the extreme eigenvalue of S_n and $(u_i, v_i)^T$ its corresponding eigenvector, we have

$$\begin{pmatrix} l_i I_k - X_1 X_1^* & -X_1 X_2^* \\ -X_2 X_1^* & l_i I_p - X_2 X_2^* \end{pmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix} = 0 \ ,$$

where u_i is the first k components, and v_i the remaining p components, and this leads to

$$\begin{cases} (l_i I_k - X_1 X_1^*) u_i - X_1 X_2^* v_i = 0 \\ -X_2 X_1^* u_i + (l_i I_p - X_2 X_2^*) v_i = 0 \ . \end{cases}$$

Consequently,

$$v_i = (l_i I_p - X_2 X_2^*)^{-1} X_2 X_1^* u_i \ , \tag{4.1}$$
$$(l_i I_k - X_1 \left(I_n + X_2^* (l_i I_p - X_2 X_2^*)^{-1} X_2 \right) X_1^*) u_i = 0 \ . \tag{4.2}$$

(4.2) is equivalent to

$$\left(l_i I_k + l_i \underline{s}(l_i) \begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_k \end{pmatrix} + o(1) \right) u_i = 0 \ , \tag{4.3}$$

where

$$\underline{s}(l_i) = \int \frac{1}{l_i - x} d\underline{F}_y(x),$$

$\underline{F}_y(x)$ is the LSD of $X_2^* X_2$. Since $\underline{s}(l_i) = -1/a_i$, we have (4.3) equivalent to

$$\begin{pmatrix} 1 - \frac{a_1}{a_i} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 - \frac{a_k}{a_i} \end{pmatrix} \begin{pmatrix} u_i(1) \\ \vdots \\ u_i(k) \end{pmatrix} = 0 \ ,$$

and that leads to

$$u_i(1) = \cdots = u_i(i-1) = u_i(i+1) = \cdots = u_i(k) = 0 \ . \tag{4.4}$$

Moreover, combining (4.1) with the fact that

$$\begin{pmatrix} u_i^* & v_i^* \end{pmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix} = 1$$

leads to

$$u_i^* (I_k + X_1 X_2^* (l_i I_p - X_2 X_2^*)^{-2} X_2 X_1^*) u_i = 1 \ ,$$

which is also equivalent to

$$u_i^2(i) [I_k + X_1 X_2^* (l_i I_p - X_2 X_2^*)^{-2} X_2 X_1^*](i,i) = 1 \tag{4.5}$$

if take (4.4) into consideration. Therefore, we have

$$\begin{aligned} u_i^2(i) &= \frac{1}{1 + [X_1 X_2^* (l_i I_p - X_2 X_2^*)^{-2} X_2 X_1^*](i,i)} \\ &= \frac{1}{1 + \mathbb{E}[X_1 X_2^* (\lambda_i I_p - X_2 X_2^*)^{-2} X_2 X_1^*](i,i)} + C \\ &= \frac{1}{1 + a_i y m_3(\lambda_i)} + C + o(\frac{1}{\sqrt{n}}) \ , \end{aligned}$$

where

$$
\begin{aligned}
C &= \frac{1}{1+[X_1X_2^*(l_iI_p-X_2X_2^*)^{-2}X_2X_1^*](i,i)} - \frac{1}{1+\mathbb{E}[X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*](i,i)} \\
&= \frac{1}{1+[X_1X_2^*(l_iI_p-X_2X_2^*)^{-2}X_2X_1^*](i,i)} - \frac{1}{1+[X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*](i,i)} \\
&\quad + \frac{1}{1+[X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*](i,i)} - \frac{1}{1+\mathbb{E}[X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*](i,i)} \\
&:= C_1 + C_2 \ .
\end{aligned}
$$

Next, we simplify the values of C_1 and C_2.

$$
\begin{aligned}
C_1 &= \frac{1}{1+[X_1X_2^*(l_iI_p-X_2X_2^*)^{-2}X_2X_1^*](i,i)} - \frac{1}{1+[X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*](i,i)} \\
&= \frac{X_1X_2^*\left[(\lambda_iI_p-X_2X_2^*)^{-2}-(l_iI_p-X_2X_2^*)^{-2}\right]X_2X_1^*(i,i)}{[1+X_1X_2^*(l_iI_p-X_2X_2^*)^{-2}X_2X_1^*(i,i)]\cdot[1+X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*(i,i)]} \ .
\end{aligned}
$$

First, consider the part in the above numerator:

$$
\begin{aligned}
&(\lambda_iI_p-X_2X_2^*)^{-2}-(l_iI_p-X_2X_2^*)^{-2} \\
&= \left[(\lambda_iI_p-X_2X_2^*)^{-1}-(l_iI_p-X_2X_2^*)^{-1}\right]\cdot\left[(\lambda_iI_p-X_2X_2^*)^{-1}+(l_iI_p-X_2X_2^*)^{-1}\right] \\
&= (l_i-\lambda_i)(\lambda_iI_p-X_2X_2^*)^{-1}(l_iI_p-X_2X_2^*)^{-1}\cdot\left[(\lambda_iI_p-X_2X_2^*)^{-1}+(l_iI_p-X_2X_2^*)^{-1}\right] \ .
\end{aligned}
\tag{4.6}
$$

Since $\sqrt{n}(l_i-\lambda_i)$ has a central limit theorem with the following expression using our notation (see [2]):

$$
\begin{aligned}
&(l_i-\lambda_i)(1+a_iym_3(\lambda_i)+o(1)) \\
&= X_1(I+X_2^*(\lambda_iI_p-X_2X_2^*)^{-1}X_2)X_1^*(i,i)-\mathbb{E}X_1(I+X_2^*(\lambda_iI_p-X_2X_2^*)^{-1}X_2)X_1^*(i,i) \ ,
\end{aligned}
\tag{4.7}
$$

which implies that (4.6) tends to

$$
2(l_i-\lambda_i)(\lambda_iI_p-X_2X_2^*)^{-3}+o(1/\sqrt{n}) \ .
$$

So

$$
\begin{aligned}
C_1 &= 2(l_i-\lambda_i)\frac{X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-3}X_2X_1^*(i,i)}{[1+X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*(i,i)]^2}+o(1/\sqrt{n}) \\
&= \frac{2a_iym_5(\lambda_i)}{(1+a_iym_3(\lambda_i))^2}\cdot(l_i-\lambda_i)+o(1/\sqrt{n}) \ .
\end{aligned}
\tag{4.8}
$$

And

$$
\begin{aligned}
C_2 &= \frac{1}{1+X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*(i,i)} - \frac{1}{1+\mathbb{E}X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*(i,i)} \\
&= -\frac{X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*(i,i)-\mathbb{E}[X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*](i,i)}{(1+X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*(i,i))(1+\mathbb{E}X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*(i,i))} \\
&= -\frac{X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*(i,i)-\mathbb{E}[X_1X_2^*(\lambda_iI_p-X_2X_2^*)^{-2}X_2X_1^*](i,i)}{(1+a_iym_3(\lambda_i))^2} \\
&\quad + o(1/\sqrt{n}) \ .
\end{aligned}
\tag{4.9}
$$

Let

$$A(\lambda) := I_n + X_2^*(\lambda_i I_p - X_2X_2^*)^{-1}X_2 \ , \tag{4.10}$$

$$B(\lambda) := X_2^*(\lambda_i I_p - X_2X_2^*)^{-2}X_2 \ , \tag{4.11}$$

combining with (4.7), (4.8) and (4.9) leads to

$$\begin{aligned} C &= \frac{2a_iym_5(\lambda_i)}{(1+a_iym_3(\lambda_i))^3} \cdot X_1[A - \mathbb{E}A]X_1^*(i,i) - \frac{X_1[B - \mathbb{E}B]X_1^*(i,i)}{(1+a_iym_3(\lambda_i))^2} + o(1/\sqrt{n}) \\ &= \frac{2a_iym_5(\lambda_i)}{(1+a_iym_3(\lambda_i))^3} \cdot \frac{\xi_{1:n}[A - \mathbb{E}A]\xi_{1:n}^*(i,i)}{n} - \frac{\xi_{1:n}[B - \mathbb{E}B]\xi_{1:n}^*(i,i)}{n(1+a_iym_3(\lambda_i))^2} + o(1/\sqrt{n}) \ . \end{aligned} \tag{4.12}$$

Therefore,

$$\begin{aligned} &\sqrt{n} \cdot \left(u_i^2(i) - \frac{1}{1+a_iym_3(\lambda_i)} \right) \\ &= \frac{2a_iym_5(\lambda_i)}{(1+a_iym_3(\lambda_i))^3} \cdot \frac{\xi_{1:n}[A - \mathbb{E}A]\xi_{1:n}^*(i,i)}{\sqrt{n}} - \frac{\xi_{1:n}[B - \mathbb{E}B]\xi_{1:n}^*(i,i)}{\sqrt{n}(1+a_iym_3(\lambda_i))^2} + o(1) \ , \end{aligned}$$

which leads to the fact that

$$\begin{pmatrix} \sqrt{n}\left(u_i^2(i) - \frac{1}{1+a_iym_3(\lambda_i)}\right) \\ \sqrt{n}(l_i - \lambda_i) \end{pmatrix} = \begin{pmatrix} \frac{2a_iym_5(\lambda_i)}{(1+a_iym_3(\lambda_i))^3} & \frac{-1}{(1+a_iym_3(\lambda_i))^2} \\ \frac{1}{1+a_iym_3(\lambda_i)} & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{n}}\xi_{1:n}[A - \mathbb{E}A]\xi_{1:n}^*(i,i) \\ \frac{1}{\sqrt{n}}\xi_{1:n}[B - \mathbb{E}B]\xi_{1:n}^*(i,i) \end{pmatrix} + o(1) \ .$$

If we denote

$$D := \begin{pmatrix} \frac{2a_iym_5(\lambda_i)}{(1+a_iym_3(\lambda_i))^3} & \frac{-1}{(1+a_iym_3(\lambda_i))^2} \\ \frac{1}{1+a_iym_3(\lambda_i)} & 0 \end{pmatrix} \ ,$$

and combining with Lemma 5.2, we have got that

$$\begin{pmatrix} \sqrt{n}\left(u_i^2(i) - \frac{1}{1+a_iym_3(\lambda_i)}\right) \\ \sqrt{n}(l_i - \lambda_i) \end{pmatrix}$$

is asymptotically Gaussian with mean $\mathbf{0}$ and covariance matrix

$$DBD^T = \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix} \ ,$$

where

$$\begin{aligned} v_{11} &= \frac{(2a_iym_5)^2}{(1+a_iym_3)^6}B_{11} - \frac{4a_iym_5}{(1+a_iym_3)^5}B_{12} + \frac{1}{(1+a_iym_3)^4}B_{22} \\ &= \frac{a_i^2y^2(a_i^2+y-1)^2}{(a_i-1)^4(a_i-1+y)^4}\nu_4(i) + \frac{2a_i^2y((a_i+y-1)^2+ya_i^2)}{((a_i-1)^2-y)(a_i-1+y)^4} \\ v_{12} &= \frac{2a_iym_5}{(1+a_iym_3)^4}B_{11} - \frac{1}{(1+a_iym_3)^3}B_{12} \\ &= \frac{ya_i^2(a_i^2-1+y)((a_i-1)^2-y)}{(a_i-1)^4(a_i-1+y)^2}\nu_4(i) + \frac{2a_i^3y}{(a_i-1)(a_i-1+y)^2} \\ v_{22} &= \frac{1}{(1+a_iym_3)^2}B_{11} \\ &= \frac{a_i^2((a_i-1)^2-y)^2}{(a_i-1)^4}\nu_4(i) + \frac{2a_i^2((a_i-1)^2-y)}{(a_i-1)^2} \end{aligned}$$

□

5 Appendix

Lemma 5.1. *For* $a \notin [1-\sqrt{y}, 1+\sqrt{y}]$ *and* $\phi(a) = a + ya/(a-1) \notin [a_y, b_y]$*, we have the following relationship:*

$$\begin{aligned}
m_0 \circ \phi(a) &= \frac{1}{a-1+y} \ , \\
m_1 \circ \phi(a) &= \frac{1}{a-1} \ , \\
m_2 \circ \phi(a) &= \frac{(a-1)+y(a+1)}{(a-1)[(a-1)^2-y]} \ , \\
m_3 \circ \phi(a) &= \frac{1}{(a-1)^2-y} \ , \\
m_4 \circ \phi(a) &= \frac{(a-1)^2}{((a-1)^2-y)(a-1+y)^2} \ , \\
m_5 \circ \phi(a) &= \frac{(a-1)^3}{((a-1)^2-y)^3} \ , \\
m_6 \circ \phi(a) &= \frac{(a-1)^4[(a-1+y)^2+a^2y]}{((a-1)^2-y)^5} \ , \\
m_3 \circ \phi(a) + m_7 \circ \phi(a) &= \frac{a(a-1+y)(a-1)^2}{((a-1)^2-y)^3} \ .
\end{aligned}$$

Proof. (Sketch of the proof) Recall the definitions of these functions in (3.3), which can all be related to the combinations of the Stieltjes transform:

$$m(\lambda) = \int \frac{1}{x-\lambda} dF(x)$$

and it's derivatives. Besides, $\underline{m}(\lambda)$ (definition and properties can be found in [1]) satisfies:

$$\lambda = -\frac{1}{\underline{m}(\lambda)} + \frac{y}{1+\underline{m}(\lambda)} \ ,$$

by taking derivatives on both sides with respect to λ and combing with the relationship between $\underline{m}(\lambda)$ and $m(\lambda)$:

$$\underline{m}(\lambda) = ym(\lambda) - \frac{1}{\lambda}(1-y)$$

will lead to the result. Details of the calculations are omitted. □

Lemma 5.2. *With the matrices* A *and* B *defined in* (4.10) *and* (4.11)*, we have*

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\xi_{1:n}[A-\mathbb{E}A]\xi_{1:n}^*(i,i) \\ \frac{1}{\sqrt{n}}\xi_{1:n}[B-\mathbb{E}B]\xi_{1:n}^*(i,i) \end{pmatrix} \Longrightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} B_{11} & B_{12} \\ B_{12} & B_{22} \end{pmatrix} \right) \ ,$$

where

$$B_{11} = a_i^2 w_1 \nu_4(i) + 2\tau_1 a_i^2 \ , \ B_{22} = a_i^2 w_2 \nu_4(i) + 2\tau_2 a_i^2 \ , \ B_{12} = a_i^2 w_3 \nu_4(i) + 2\tau_3 a_i^2 \ ,$$

and

$$w_1 = \frac{(a_i-1+y)^2}{(a_i-1)^2} \ , \ w_2 = \frac{y^2}{((a_i-1)^2-y)^2} \ , \ w_3 = \frac{y(y+a_i-1)}{(a_i-1)\cdot((a_i-1)^2-y)}$$

$$\tau_1 = \frac{(a_i-1+y)^2}{(a_i-1)^2-y} \ , \ \tau_2 = \frac{y(a_i-1)^4((a_i-1+y)^2+a_i^2y)}{((a_i-1)^2-y)^5} \ , \ \tau_3 = \frac{a_iy(a_i-1+y)(a_i-1)^2}{((a_i-1)^2-y)^3}$$

Proof. Using Corollary 2.2, and let $X(1)^* = Y(1)^* = (\xi_{i1}, \cdots, \xi_{in})$ (i-th row of $\xi_{1:n}$), $l = l^{'} = 1$ and $K = 1$, we have

$$A_1 = \mathbb{E}\xi_i^4 - (\mathbb{E}\xi_i^2)^2 = a_i^2(3 + \nu_4(i)) - a_i^2 \ ,$$
$$A_2 = \mathbb{E}\xi_i^2\mathbb{E}\xi_i^2 = a_i^2 \ ,$$
$$A_3 = \mathbb{E}\xi_i^2\mathbb{E}\xi_i^2 = a_i^2 \ .$$

We only have to calculate these values of w_i and τ_i.

First,

$$\begin{aligned} w_1 &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \left(1 + X_2^*(\lambda I_p - X_2X_2^*)^{-1}X_2(i,i)\right)^2 \\ &= 1 + \left(\frac{y(1+m_1(\lambda))}{\lambda - y(1+m_1(\lambda))}\right)^2 + 2ym_1(\lambda) \\ &= \left(\frac{a_i + y - 1}{a_i - 1}\right)^2 \ , \end{aligned}$$

and

$$\begin{aligned} \theta_1 = \tau_1 &= \lim_{n\to\infty} \frac{1}{n} \operatorname{tr} \left(I_n + X_2^*(\lambda I_p - X_2X_2^*)^{-1}X_2\right)^2 \\ &= 1 + 2ym_1(\lambda) + ym_2(\lambda) \\ &= \frac{(a_i - 1 + y)^2}{(a_i-1)^2 - y} \end{aligned}$$

has been proven in [2].

Next,

$$w_2 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}[B(\lambda)(i,i)]^2 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \left[X_2^*(\lambda I_p - X_2X_2^*)^{-2}X_2(i,i)\right]^2 \ .$$

Since

$$X_2^*(\lambda I_p - X_2X_2^*)^{-2}X_2(i,i) = e_i^*X_2^*(\lambda I_p - X_2X_2^*)^{-2}X_2e_i \ , \tag{5.1}$$

where e_i is the column vector with its i-th coordinate being 1. Recall that

$$X_2 = \frac{1}{\sqrt{n}}(\eta_1, \cdots, \eta_n)_{p\times n} := \frac{1}{\sqrt{n}}\eta_{1:n} \ .$$

then (5.1) reduces to

$$\frac{1}{n}\eta_i^*(\lambda I_p - X_2X_2^*)^{-2}\eta_i \ . \tag{5.2}$$

Denote X_{2i} as the matrix that removing the i-th column of X_2:

$$X_{2i} = \frac{1}{\sqrt{n}}(\eta_1, \cdots, \eta_{i-1}, \eta_{i+1}, \cdots, \eta_n) \ ,$$

then

$$X_2X_2^* = X_{2i}X_{2i}^* + \frac{1}{n}\eta_i\eta_i^* \ .$$

Using the matrix identity that

$$(\lambda I_p - X_2X_2^*)^{-1} - (\lambda I_p - X_{2i}X_{2i}^*)^{-1} = (\lambda I_p - X_2X_2^*)^{-1}\frac{1}{n}\eta_i\eta_i^*(\lambda I_p - X_{2i}X_{2i}^*)^{-1} \ ,$$

we have

$$(\lambda I_p - X_2 X_2^*)^{-1} = \frac{1}{1 - \frac{1}{n}\eta_i^*(\lambda I_p - X_{2i}X_{2i}^*)^{-1}\eta_i} \cdot (\lambda I_p - X_{2i}X_{2i}^*)^{-1} ,$$

which leads to

$$(\lambda I_p - X_2 X_2^*)^{-2} = \frac{1}{\left(1 - \frac{1}{n}\eta_i^*\,(\lambda I_p - X_{2i}X_{2i}^*)^{-1}\eta_i\right)^2} \cdot (\lambda I_p - X_{2i}X_{2i}^*)^{-2} ,$$

and (5.2) equals to

$$\frac{\frac{1}{n}\eta_i^*(\lambda I_p - X_{2i}X_{2i}^*)^{-2}\eta_i}{\left(1 - \frac{1}{n}\eta_i^*\,(\lambda I_p - X_{2i}X_{2i}^*)^{-1}\eta_i\right)^2} ,$$

which tends to the limit:

$$\frac{y\int \frac{1}{(\lambda-x)^2}dF(x)}{(1-y\int \frac{1}{\lambda-x}dF(x))^2} = \frac{ym_4(\lambda)}{(1-ym_0(\lambda))^2} .$$

Therefore,

$$w_2 = \frac{(ym_4(\lambda))^2}{(1-ym_0(\lambda))^2} = \frac{y^2}{((a_i-1)^2-y)^2} .$$

$$\begin{aligned}
w_3 &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} A(\lambda)(i,i)B(\lambda)(i,i) \\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(1 + X_2^*(\lambda I_p - X_2X_2^*)^{-1}X_2(i,i)\right)\cdot X_2^*(\lambda I_p - X_2X_2^*)^{-2}X_2(i,i) \\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} X_2^*(\lambda I_p - X_2X_2^*)^{-1}X_2(i,i)\cdot X_2^*(\lambda I_p - X_2X_2^*)^{-2}X_2(i,i) \\
&\quad + \lim_{n\to\infty}\frac{1}{n}\,\mathrm{tr}\left[X_2^*(\lambda I_p - X_2X_2^*)^{-2}X_2\right] \\
&= \frac{y(1+m_1(\lambda))}{\lambda - y(1+m_1(\lambda))}\cdot\frac{ym_4(\lambda)}{(1-ym_0(\lambda))^2} + ym_3(\lambda) \\
&= \frac{y(y+a_i-1)}{(a_i-1)((a_i-1)^2-y)}
\end{aligned}$$

$$\begin{aligned}
\theta_2 = \tau_2 &= \lim_{n\to\infty}\frac{1}{n}\sum_{i,j=1}^{n}\left(X_2^*(\lambda I_p - X_2X_2^*)^{-2}X_2(i,j)\right)^2 \\
&= \lim_{n\to\infty}\frac{1}{n}\,\mathrm{tr}\left[X_2^*(\lambda I_p - X_2X_2^*)^{-2}X_2\right]^2 \\
&= y\int\frac{x^2}{(\lambda-x)^4}dF(x) \\
&= ym_6(\lambda) \\
&= \frac{y(a_i-1)^4\left((a_i-1+y)^2 + a_i^2 y\right)}{((a_i-1)^2-y)^5}
\end{aligned}$$

$$\begin{aligned}
\theta_3 = \tau_3 &= \lim_{n\to\infty} \frac{1}{n} \operatorname{tr}[A(\lambda)B(\lambda)] \\
&= \lim_{n\to\infty} \frac{1}{n} \operatorname{tr}\left\{\left(I_n + X_2^*(\lambda I_p - X_2X_2^*)^{-1}X_2\right) X_2^*(\lambda I_p - X_2X_2^*)^{-2}X_2\right\} \\
&= y\int \frac{x}{(\lambda - x)^2} dF(x) + y\int \frac{x^2}{(\lambda - x)^3} dF(x) \\
&= y(m_3(\lambda) + m_7(\lambda)) \\
&= \frac{a_i y(a_i - 1 + y)(a_i - 1)^2}{((a_i - 1)^2 - y)^3}
\end{aligned}$$

The proof of Lemma 5.2 is complete. □

Acknowledgement

We thank the anonymous referees for helpful comments. In particular, the application in Proposition 2.4 has been suggested by one of the referees.

References

[1] Bai, Z.D. and Silverstein, J.W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.*, **32**(1A), 553–605. MR-2040792

[2] Bai, Z.D. and Yao, J.F. (2008). CLT for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.*, **44**(3), 447–474. MR-2451053

[3] Baik, J., Ben Arous, G. and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, **33**(5), 1643–1697. MR-2165575

[4] Baik, J. and Silverstein, J.W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.*, **97**, 1382–1408. MR-2279680

[5] Benaych-Georges, F., Guionnet, A. and Maida, M. (2011). Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.*, **16**(60), 1621–1662. MR-2835249

[6] Benaych-Georges, F. and Nadakuditi, R.R. (2011). The eigenvalues and eigenvectors of finite low rank perturbations of large random matrices. *Adv. Math.*, **227**(1), 494–521. MR-2782201

[7] Capitaine, M., Donati-Martin, C. and Féral, D. (2009). The largest eigenvalue of finite rank deformation of large Wigner matrices: Convergence and non university of the fluctuations. *Ann. Probab.*, **37**(1), 1–47. MR-2489158

[8] Capitaine, M., Donati-Martin, C. and Féral, D. (2012). Central limit theorems for eignvalues of deformations of Wigner matrices. Ann. Inst. Henri Poincaré Probab. Stat., **48**(1), 107–133. MR-2919200

[9] De Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields*, **75**(2), 261–277. MR-0885466

[10] Fox, R and Taqqu, M.S.(1987). Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Theory Related Fields*, **74**(2), 213–240. MR-0871252

[11] Hachem, W., Loubaton, P. Najim, J. and Vallet, P. (2013). On bilinear forms based on the resolvent of large random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, **49**(1), 36–63. MR-3060147

[12] Jakubowski, A. and Mémin, J. (1994). Functional central limit theorems for a class of quadratic forms in independent random variables. *translation in Theory Probab. Appl.*, **38**(3), 423–432. MR-1404667

[13] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, **29**(2), 295–327. MR-1863961

[14] Knowles, A. and Yin, J. (2014). The outliers of a deformed Wigner matrix. *Ann. Probab.*, **42**(5), 1980–2031. MR-3262497

[15] Lee, S., Zou, F. and Wright, F.A.(2010). Convergence and prediction of principal component scores in high-dimensional settings. *Ann. Statist.*, **38**(6), 3605–3629. MR-2766862

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

[17] Mikosch, T. (1991). Functional limit theorems for random quadratic forms. *Stochastic Process. Appl.*, **37**, 81–98. MR-1091696

[18] Pan, G.M., Miao, B.Q. and Jin, B. (2008). Central limit theorem of random quadratics forms involving random matrices. *Statistics & Probability Letters.*, **78**, 804–809. MR-2409545

[19] Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica.*, **17**, 1617–1642. MR-2399865

[20] Pizzo, A., Renfrew, D. and Soshnikov, A. (2013). On finite rank deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, **49**(1), 64–94. MR-3060148

[21] Rotar', V.I. (1973). Some limit theorems for polynomials of second degree. *Theory Probab. Appl.*, **18**, 499–507. MR-0326803

[22] Renfrew, D. and Soshnikov, A. (2013). On finite rank deformations of Wigner matrices II: delocalized perturbations. *Random Matrix: Theory and Applciations*, **2**(1), 1250015 (36 pages). MR-3039820

[23] Sevastyanov, B.A. (1961). A class of limit distributions for quadratic forms of normal stochastic variables. *Theor. Probab. Appl.*, **6**, 337–340.

[24] Whittle, P. (1964). On the convergence to normality of quadratic forms in independent variables. *Theor. Probab. Appl.*, **9**, 103–109. MR-0161429