

# Test of independence for high-dimensional random vectors based on freeness in block correlation matrices

Zhigang Bao<sup>†</sup>

*Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong*  
*e-mail: [mazgbao@ust.hk](mailto:mazgbao@ust.hk)*

Jiang Hu<sup>‡,\*</sup>

*KLASMOE and School of Mathematics & Statistics, Northeast Normal University,  
Changchun, China*  
*e-mail: [huj156@nenu.edu.cn](mailto:huj156@nenu.edu.cn)*

Guangming Pan<sup>§</sup>

*Division of Mathematical Sciences, School of Physical and Mathematical Sciences,  
Nanyang Technological University, Singapore*  
*e-mail: [gmpan@ntu.edu.sg](mailto:gmpan@ntu.edu.sg)*

and

Wang Zhou<sup>¶</sup>

*Department of Statistics and Applied Probability, National University of Singapore,  
Singapore*  
*e-mail: [stazw@nus.edu.sg](mailto:stazw@nus.edu.sg)*

**Abstract:** In this paper, we are concerned with the independence test for  $k$  high-dimensional sub-vectors of a normal vector, with fixed positive integer  $k$ . A natural high-dimensional extension of the classical sample correlation matrix, namely block correlation matrix, is proposed for this purpose. We then construct the so-called Schott type statistic as our test statistic, which turns out to be a particular linear spectral statistic of the block correlation matrix. Interestingly, the limiting behavior of the Schott type statistic can be figured out with the aid of the Free Probability Theory and the Random Matrix Theory. Specifically, we will bring the so-called real second order freeness for Haar distributed orthogonal matrices, derived in Mingo and Popa (2013)[10], into the framework of this high-dimensional testing

---

\*Corresponding author.

<sup>†</sup>Z.G. Bao was supported by a startup fund from HKUST.

<sup>‡</sup>J. Hu was partially supported by Science and Technology Development Foundation of Jilin (Grant No. 20160520174JH), Science and Technology Foundation of Jilin during the “13th Five-Year Plan” and the National Natural Science Foundation of China (Grant No. 11301063).

<sup>§</sup>G.M. Pan was partially supported by a MOE Tier 2 grant 2014-T2-2-060 and by a MOE Tier 1 Grant RG25/14 at the Nanyang Technological University, Singapore.

<sup>¶</sup>W. Zhou was partially supported by R-155-000-165-112 at the National University of Singapore.

problem. Our test does not require the sample size to be larger than the total or any partial sum of the dimensions of the  $k$  sub-vectors. Simulated results show the effect of the Schott type statistic, in contrast to those statistics proposed in Jiang and Yang (2013)[8] and Jiang, Bai and Zheng (2013)[7], is satisfactory. Real data analysis is also used to illustrate our method.

**MSC 2010 subject classifications:** Primary 62H15; secondary 46L54.

**Keywords and phrases:** Block correlation matrix, independence test, high dimensional data, Schott type statistic, second order freeness, haar distributed orthogonal matrices, central limit theorem, random matrices.

Received September 2016.

## 1. Introduction

Test of independence for random variables is a very classical hypothesis testing problem, which dates back to the seminal work by Pearson in [17], followed by a huge literature regarding this topic and its variants. One frequently recurring variant is the test of independence for  $k$  random vectors, where  $k \geq 2$  is an integer. Comprehensive overview and detailed references on this problem can be found in most of the textbooks on multivariate statistical analysis. For instance, here we recommend the masterpieces by Muirhead in [13] and by Anderson in [1] for more details, in the low-dimensional case. However, due to the increasing demand in the analysis of big data springing up in various fields nowadays, such as genomics, signal processing, microarray, proteomics and finance, the investigation on a high-dimensional extension of this testing problem is much needed, which motivates us to propose a feasible way for it in this work.

Let us take a review more specifically on some representative existing results in the literature, after necessary notation is introduced. For simplicity, henceforth, we will use the notation  $\llbracket m \rrbracket$  to denote the set  $\{1, 2, \dots, m\}$  for any positive integer  $m$ . Assume that  $\mathbf{x} = (\mathbf{x}'_1, \dots, \mathbf{x}'_k)'$  is a  $p$ -dimensional normal vector, in which  $\mathbf{x}_i$  possesses dimension  $p_i$  for  $i \in \llbracket k \rrbracket$ , such that  $p_1 + \dots + p_k = p$ . Denote by  $\boldsymbol{\mu}_i$  the mean vector of the  $i$ th sub-vector  $\mathbf{x}_i$  and by  $\Sigma_{ij}$  the cross covariance matrix of  $\mathbf{x}_i$  and  $\mathbf{x}_j$  for all  $i, j \in \llbracket k \rrbracket$ . Then  $\boldsymbol{\mu} := (\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_k)'$  and  $\Sigma := (\Sigma_{ij})_{i,j=1}^k$  are the mean vector and covariance matrix of  $\mathbf{x}$  respectively. In this work, we consider the following hypothesis testing

$$(T1) \quad \mathbf{H}_0 : \Sigma_{ij} = 0, \quad i \neq j \quad \text{v.s.} \quad \mathbf{H}_1 : \text{not } \mathbf{H}_0.$$

To this end, we draw  $n$  observations of  $\mathbf{x}$ , namely  $\mathbf{x}(1), \dots, \mathbf{x}(n)$ . In addition, the  $i$ th sub-vector of  $\mathbf{x}(j)$  will be denoted by  $\mathbf{x}_i(j)$  for all  $i \in \llbracket k \rrbracket$  and  $j \in \llbracket n \rrbracket$ . Hence,  $\mathbf{x}_i(j), j \in \llbracket n \rrbracket$  are  $n$  independent observations of  $\mathbf{x}_i$ . Conventionally, the corresponding sample means will be written as  $\bar{\mathbf{x}} = n^{-1} \sum_{j=1}^n \mathbf{x}(j)$  and  $\bar{\mathbf{x}}_i = n^{-1} \sum_{j=1}^n \mathbf{x}_i(j)$ . With the observations at hand, we can construct the sample covariance matrix as usual. Set

$$\mathbf{X} := [\mathbf{x}(1) - \bar{\mathbf{x}}, \dots, \mathbf{x}(n) - \bar{\mathbf{x}}], \quad \mathbf{X}_i := [\mathbf{x}_i(1) - \bar{\mathbf{x}}_i, \dots, \mathbf{x}_i(n) - \bar{\mathbf{x}}_i], \quad i \in \llbracket k \rrbracket. \quad (1.1)$$

The sample covariance matrix of  $\mathbf{x}$  and the cross sample covariance matrix of  $\mathbf{x}_m$  and  $\mathbf{x}_\ell$  will be denoted by  $\hat{\Sigma}$  and  $\hat{\Sigma}_{m\ell}$  respectively, to wit,

$$\hat{\Sigma} := \frac{1}{n-1} \mathbf{X}\mathbf{X}', \quad \hat{\Sigma}_{m\ell} := \frac{1}{n-1} \mathbf{X}_m \mathbf{X}_\ell'.$$

In the classical *large n and fixed p* case, the likelihood ratio statistic  $\Lambda_n := W_n^{n/2}$  with

$$W_n := \frac{|\hat{\Sigma}|}{\prod_{i=1}^k |\hat{\Sigma}_{ii}|}$$

is a favorable one for the testing problem **(T1)**. A celebrated limiting law on  $\Lambda_n$  under  $\mathbf{H}_0$  is

$$-2\kappa \log \Lambda_n \implies \chi_\rho^2, \quad \text{as } n \rightarrow \infty \tag{1.2}$$

where

$$\kappa = 1 - \frac{2(p^3 - \sum_{i=1}^k p_i^3) + 9(p^2 - \sum_{i=1}^k p_i^2)}{6n(p^2 - \sum_{i=1}^k p_i^2)}, \quad \rho = \frac{1}{2}(p^2 - \sum_{i=1}^k p_i^2).$$

One can refer to [27] or Theorem 11.2.5 of [13], for instance. The left hand side of (1.2) is known as the Wilks statistic.

Now, we turn to the high-dimensional setting of interest. A commonly used assumption on dimensionality and sample size in the Random Matrix Theory **(RMT)** is that  $p$  is proportionally large as  $n$ , i.e.

$$p := p(n), \quad \frac{p}{n} \rightarrow y \in (0, \infty), \quad \text{as } n \rightarrow \infty.$$

To employ the existing **RMT** apparatus in the sequel, hereafter, we will always work with the above *large n and proportionally large p* setting. This time, resorting to the likelihood ratio statistic in (1.2) directly is obviously infeasible, since the limiting law (1.2) is invalid when  $p$  tends to infinity along with  $n$ . Actually, under this setting, the likelihood ratio statistic can still be employed if an appropriate renormalization is performed a priori. In [8], the authors renormalized the likelihood ratio statistic, and derived its limiting law under  $\mathbf{H}_0$  as a central limit theorem (CLT), under the restriction of

$$n > p + 1, \quad p_i/n \rightarrow y_i \in (0, 1), i \in \llbracket k \rrbracket, \quad \text{as } n \rightarrow \infty. \tag{1.3}$$

One can refer to Theorem 2 in [8] for the details of the normalization and the CLT. One similar result was also obtained in [7], see Theorem 4.1 therein. However, the condition (1.3) is indispensable for the likelihood ratio statistic which will inevitably hit the wall when  $p$  is even larger than  $n$ , owing to the fact that  $\log |\hat{\Sigma}|$  is not well-defined in this situation. In addition, in [7], another test statistic constructed from the traces of F-matrices, the so-called *trace criterion*

test, was proposed. Under  $\mathbf{H}_0$ , a CLT was derived for this statistic under the following restrictions

$$\frac{p_i}{p_1 + \cdots + p_{i-1}} \rightarrow r_1^{(i)} \in (0, \infty), \frac{p_i}{n - 1 - (p_1 + \cdots + p_{i-1})} \rightarrow r_2^{(i)} \in (0, 1), \quad (1.4)$$

for all  $i \in \llbracket k \rrbracket$ , together with  $p - p_1 < n$ . We stress here, condition (1.4) is obviously much stronger than  $p_i/n \rightarrow y_i \in (0, 1)$  for all  $i \in \llbracket k \rrbracket$ .

Roughly speaking, our aim in this paper is to propose a new statistic with both statistical visualizability and mathematical tractability, whose limiting behavior can be derived with the following restriction on the dimensionality and the sample size

$$p_i := p_i(n), \quad p_i/n \rightarrow y_i \in (0, 1), \quad i \in \llbracket k \rrbracket, \quad \text{as } n \rightarrow \infty. \quad (1.5)$$

Especially,  $n$  is not required to be larger than  $p$  or any partial sum of  $p_i$ . More precisely, our multi-facet target consists of the following:

- Introducing a new matrix model tailored to **(T1)**, namely *block correlation matrix*, which can be viewed as a natural high-dimensional extension of the sample correlation matrix.
- Constructing the so-called *Schott type statistic* from the block correlation matrix, which can be regarded as an extension of Schott's statistic for complete independence test in [21].
- Deriving the limiting distribution of the Schott type statistic with the aid of tools from the Free Probability Theory (**FPT**). Specifically, we will channel the so-called *real second order freeness* for Haar distributed orthogonal matrices from [10] into the framework.
- Employing this limiting law to test independence of  $k$  sub-vectors under (1.5) and assessing the statistic via simulations and a real data set, which comes from the daily returns of 258 stocks issued by the companies from S&P 500.

It will be seen that for **(T1)**, it is quite natural and reasonable to put forward the concept of block correlation matrix. Just like that the sample correlation matrix is designed to bypass the unknown mean values and variances of entries, the block correlation matrix can also be employed, without knowing the population mean vectors  $\boldsymbol{\mu}_i$  and covariance matrices  $\Sigma_{ii}$ ,  $i \in \llbracket k \rrbracket$ . Also interestingly, it turns out that, the statistical meaning of the Schott type statistic is rooted in the idea of testing independence based on the Canonical Correlation Analysis (**CCA**). Meanwhile, theoretically, the Schott type statistic is a special type of linear spectral statistic from the perspective of **RMT**. Then methodologies from **RMT** and **FPT** can take over the derivation of the limiting behavior of the proposed statistic. As far as we know, the application of **FPT** in high-dimensional statistical inference is still in its infancy. We also hope this work can evoke more applications of **FPT** in statistical inference in the future. One may refer to [19] for another application of **FPT**, in the context of statistics.

Our paper is organized as follows. In Section 2, we will construct the block correlation matrix. Then we will present the definition of the Schott type statistic and its limiting law in Section 3. Meanwhile, we will discuss the statistical rationality of this test statistic, especially the relationship with CCA. In Section 4, we will detect the utility of our statistic by simulation, and an example about stock prices will be analyzed in Section 5. Finally, Section 6 will be devoted to the illustration of how RMT and FPT machinery can help to establish the limiting behavior of our statistic, and some calculations will be presented in the Appendix.

Throughout the paper,  $\text{tr}\mathbf{A}$  represents the trace of a square matrix  $\mathbf{A}$ . If  $\mathbf{A}$  is  $N \times N$ , we will use  $\lambda_1(\mathbf{A}), \dots, \lambda_N(\mathbf{A})$  to denote its  $N$  eigenvalues. Moreover,  $|\mathbf{A}|$  means the determinant of  $\mathbf{A}$ . In addition,  $\mathbf{0}_{M \times N}$  will be used to denote the  $M \times N$  null matrix, and be abbreviated to  $\mathbf{0}$  when there is no confusion on the dimension. Moreover, for random objects  $\xi$  and  $\eta$ , we use  $\xi \stackrel{d}{=} \eta$  to represent that  $\xi$  and  $\eta$  share the same distribution.

## 2. Block correlation matrix

An elementary property of the sample correlation matrix is that it is invariant under translation and scaling of variables. Such an advantage allows us to discuss the limiting behavior of statistics constructed from the sample correlation matrix without knowing the means and variances of the involved variables, thus makes it a favorable choice in dealing with the real data. Now we are in a similar situation, without knowing explicit information of the population mean vectors  $\boldsymbol{\mu}_i$  and covariance matrices  $\Sigma_{ii}$ , we want to construct a test statistic which is independent of these unknown parameters, in a similar vein. The first step is to propose a high-dimensional extension of sample correlation matrix, namely the block correlation matrix. For simplicity, we use the notation

$$\text{diag}(\mathbf{A}_i)_{i=1}^k = \begin{pmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_k \end{pmatrix}$$

to denote the diagonal block matrix with blocks  $\mathbf{A}_i, i \in \llbracket k \rrbracket$ , i.e. all off-diagonal blocks are  $\mathbf{0}$ .

**Definition 2.1** ((Block correlation matrix)). *With the aid of the notation in (1.1), the block correlation matrix  $\mathbf{B} := \mathbf{B}(\mathbf{X}_1, \dots, \mathbf{X}_k)$  is defined as follows*

$$\begin{aligned} \mathbf{B} &:= \text{diag}((\mathbf{X}_i \mathbf{X}_i')^{-\frac{1}{2}})_{i=1}^k \cdot \mathbf{X} \mathbf{X}' \cdot \text{diag}((\mathbf{X}_i \mathbf{X}_i')^{-\frac{1}{2}})_{i=1}^k \\ &= \left( (\mathbf{X}_i \mathbf{X}_i')^{-1/2} \mathbf{X}_i \mathbf{X}_j' (\mathbf{X}_j \mathbf{X}_j')^{-1/2} \right)_{i,j=1}^k. \end{aligned}$$

**Remark 2.1.** *Note that when  $p_i = 1$  for all  $i \in \llbracket k \rrbracket$ ,  $\mathbf{B}$  is reduced to the classical sample correlation matrix of  $n$  observations of a  $k$ -dimensional random vector.*

In this sense, we can regard  $\mathbf{B}$  as a natural high-dimensional extension of the sample correlation matrix.

**Remark 2.2.** If we take determinant of  $\mathbf{B}$ , we can get the likelihood ratio statistic. However, since one needs to further take logarithm on the determinant, the assumption  $n > p + 2$  is indispensable, in light of [8].

In the sequel, we perform a very standard and well-known transformation for  $\mathbf{X}$  to eliminate the inconvenience caused by subtracting the sample mean. Set the orthogonal matrix

$$A = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{3 \cdot 2}} & \frac{1}{\sqrt{3 \cdot 2}} & -\frac{2}{\sqrt{3 \cdot 2}} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & -\frac{n-1}{\sqrt{n(n-1)}} \end{pmatrix}.$$

Then we can find that there exist i.i.d.  $\mathbf{z}(j) \sim N(0, \Sigma)$ ,  $j \in \llbracket n-1 \rrbracket$ , such that

$$\mathbf{X}A' := (\mathbf{0}, \mathbf{z}(1), \dots, \mathbf{z}(n-1)),$$

Analogously, we denote

$$\mathbf{X}_i A' := (\mathbf{0}, \mathbf{z}_i(1), \dots, \mathbf{z}_i(n-1)), \quad i \in \llbracket k \rrbracket.$$

Obviously,  $\mathbf{z}(j) = (\mathbf{z}'_1(j), \dots, \mathbf{z}'_k(j))'$ . To abbreviate, we further set the matrices

$$\mathbf{Z} = (\mathbf{z}(1), \dots, \mathbf{z}(n-1)), \quad \mathbf{Z}_i = (\mathbf{z}_i(1), \dots, \mathbf{z}_i(n-1)), \quad i \in \llbracket k \rrbracket.$$

Apparently, we have  $\mathbf{Z}\mathbf{Z}' = \mathbf{X}\mathbf{X}'$  and  $\mathbf{Z}_i\mathbf{Z}'_i = \mathbf{X}_i\mathbf{X}'_i$ . Consequently, we can also write

$$\begin{aligned} \mathbf{B} &= \text{diag}(\mathbf{Z}_i\mathbf{Z}'_i)^{-\frac{1}{2}}_{i=1}^k \cdot \mathbf{Z}\mathbf{Z}' \cdot \text{diag}(\mathbf{Z}_i\mathbf{Z}'_i)^{-\frac{1}{2}}_{i=1}^k \\ &= \left( (\mathbf{Z}_i\mathbf{Z}'_i)^{-1/2} \mathbf{Z}_i \mathbf{Z}'_j (\mathbf{Z}_j \mathbf{Z}'_j)^{-1/2} \right)_{i,j=1}^k. \end{aligned}$$

An advantage of  $\mathbf{Z}_i$  over  $\mathbf{X}_i$  is that its columns are i.i.d.

### 3. Schott type statistic and main result

With the block correlation matrix at hand, we can propose our test statistic for **(T1)**, namely the Schott type statistic. Such a nomenclature is motivated by Schott in [21] on another classical independence test problem, the so-called *complete independence test*, which can be described as follows. Given a random vector  $\mathbf{w} = (w_1, \dots, w_p)'$ , we consider the following hypothesis testing:

$$\mathbf{(T2)} \quad \tilde{\mathbf{H}}_0 : w_1, \dots, w_p \text{ are completely independent} \quad \text{v.s.} \quad \tilde{\mathbf{H}}_1 : \text{not } \tilde{\mathbf{H}}_0.$$

When  $\mathbf{w}$  is multivariate normal, the above test problem is equivalent to the so-called test of sphericity of population correlation matrix, i.e. under the null hypothesis, the population correlation matrix of  $\mathbf{w}$  is  $\mathbf{I}_p$ . There is a long list of references devoted to testing complete independence for a random vector under the high-dimensional setting. See, for example, [14], [15], [23], [4], [2] and [21]. Especially, in [21], the author constructed a statistic from the sample correlation matrix. To be specific, denote the sample correlation matrix of  $n$  i.i.d. observations of  $\mathbf{w}$  by  $\mathbf{R} = \mathbf{R}(n, p) := (r_{ij})_{p \times p}$ . Schott's statistic for **(T2)** is then defined as follows

$$\mathbf{s}(\mathbf{R}) := \sum_{i=2}^p \sum_{j=1}^{i-1} r_{ij}^2 = \frac{1}{2} \left( \sum_{i,j=1}^p r_{ij}^2 - p \right) = \frac{1}{2} \text{tr} \mathbf{R}^2 - \frac{p}{2}.$$

Note that under the null hypothesis, all off-diagonal entries of the population correlation matrix should be 0. Hence, it is quite natural to use the summation of squares of all off-diagonal entries to measure the difference between the population correlation matrix and  $I$ . Then Schott's statistic  $\mathbf{s}(\mathbf{R})$  is just the sample counterpart of such a measurement. Now in a similar vein, we define the *Schott type statistic* for the block correlation matrix as follows.

**Definition 3.1** (Schott type statistic). *We define the Schott type statistic of the block correlation matrix  $\mathbf{B}$  by*

$$\mathbf{s}(\mathbf{B}) := \frac{1}{2} \text{tr} \mathbf{B}^2 - \frac{p}{2} = \frac{1}{2} \sum_{\ell=1}^p \lambda_{\ell}^2(\mathbf{B}) - \frac{p}{2}.$$

For simplicity, we introduce the matrix

$$\mathbf{C}(i, j) := (\mathbf{X}_i \mathbf{X}_i')^{-1/2} \mathbf{X}_i \mathbf{X}_j' (\mathbf{X}_j \mathbf{X}_j')^{-1} \mathbf{X}_j \mathbf{X}_i' (\mathbf{X}_i \mathbf{X}_i')^{-1/2}.$$

Stemming from the definition of  $\mathbf{B}$ , we can easily get

$$\mathbf{s}(\mathbf{B}) = \frac{1}{2} \sum_{i,j=1}^k \text{tr} \mathbf{C}(i, j) - \frac{p}{2} = \sum_{i=2}^k \sum_{j=1}^{i-1} \text{tr} \mathbf{C}(i, j).$$

The matrix  $\mathbf{C}(i, j)$  is frequently used in **CCA** towards random vectors  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , known as the canonical correlation matrix. A well-known fact is that the eigenvalues of  $\mathbf{C}(i, j)$  provide meaningful measures of the correlation between these two random vectors. Actually, its singular values (square root of eigenvalues) are the well-known canonical correlations between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . Thus, the summation of all eigenvalues,  $\text{tr} \mathbf{C}(i, j)$ , also known as Pillai's test statistic, can obviously measure the correlation between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . Summing up these measurements over all  $(i, j)$  pairs (subject to  $j < i$ ) yields our Schott type statistic, which can capture the overall correlation among  $k$  parts of  $\mathbf{x}$ . Thus, from the perspective of **CCA**, the Schott type statistic possesses a theoretical validity for the testing problem **(T1)**.

Our main result is the following CLT under  $\mathbf{H}_0$ .

**Theorem 3.1.** Assume that  $p_i := p_i(n)$  and  $p_i/n \rightarrow y_i \in (0, 1)$  for all  $i \in \llbracket k \rrbracket$ , we have

$$\frac{\mathbf{s}(\mathbf{B}) - a_n}{\sqrt{b_n}} \Longrightarrow N(0, 1), \quad \text{as } n \rightarrow \infty,$$

where

$$a_n = \frac{1}{2} \sum_{i,j,i \neq j} \frac{p_i p_j}{n-1}, \quad b_n = \sum_{i,j,i \neq j} \frac{p_i p_j (n-1-p_i)(n-1-p_j)}{(n-1)^4}.$$

**Remark 3.1.** Note that by assumption,  $a_n$  is of order  $O(n)$  and  $b_n$  is of order  $O(1)$ .

**Remark 3.2.** For  $k = 2$ , this theorem coincides with Theorem 2.2 of [28], which does not need the normality assumption. We believe that Theorem 3.1 should be also true under the same conditions as those in Theorem 2.2 of [28].

#### 4. Numerical studies

In this section we compare the performance of the statistics proposed in [8], [7] and our Schott type statistic, under various settings of sample size and dimensionality. For simplicity, we will focus on the case of  $k = 3$ . The Wilks' statistic in (1.2) without any renormalization has been shown to perform badly for  $(\mathbf{T1})$  in [8] and [7], thus will not be considered in this section. Under the null hypothesis  $\mathbf{H}_0$ , due to the scale invariance of the three tests, without loss of generality, the samples are drawn from the following distribution:

$$(I) \quad \boldsymbol{\mu}_i = \mathbf{0}_{p_i \times 1}, \quad \Sigma_{ii} = \mathbf{I}_{p_i};$$

Under the alternative hypothesis  $\mathbf{H}_1$ , we adopt five distributions, including the two used in [8](II) and [7](III) respectively:

$$(II) \quad \boldsymbol{\mu} = \mathbf{0}_{p \times 1}, \quad \Sigma = 0.15 \mathbf{1}_{p \times p} + 0.85 \mathbf{I}_p;$$

$$(III) \quad \boldsymbol{\mu}_i = \mathbf{0}_{p_i \times 1}, \quad \Sigma_{ii} = 26/25 \mathbf{I}_{p_i}, \quad \Sigma_{12} = 1/25 \mathbb{I}_{12}, \quad \Sigma_{13} = 6/25 \mathbb{I}_{13} \text{ and } \Sigma_{23} = 6/25 \mathbb{I}_{23};$$

$$(IV) \quad \boldsymbol{\mu}_i = \mathbf{0}_{p_i \times 1}, \quad \Sigma_{ii} = \mathbf{I}_{p_i}, \quad \Sigma_{12} = 1/2 \mathbb{I}_{12}, \quad \Sigma_{13} = \mathbf{0}_{p_1 \times p_3} \text{ and } \Sigma_{23} = \mathbf{0}_{p_2 \times p_3};$$

$$(V) \quad \boldsymbol{\mu}_i = \mathbf{0}_{p_i \times 1}, \quad \Sigma_{ii} = \mathbf{I}_{p_i}, \quad \Sigma_{12} = \mathbf{0}_{p_1 \times p_2}, \quad \Sigma_{13} = \mathbf{0}_{p_1 \times p_3} \text{ and } \Sigma_{23} = 1/2 \mathbb{I}_{23};$$

$$(VI) \quad \boldsymbol{\mu}_i = \mathbf{0}_{p_i \times 1}, \quad \Sigma_{ii} = \mathbf{I}_{p_i}, \quad \Sigma_{12} = \mathbf{0}_{p_1 \times p_2}, \quad \Sigma_{13} = \mathbf{0}_{p_1 \times p_3} \text{ and } \Sigma_{23} = 3/4 \mathbb{1}_{23}.$$

Here  $\mathbf{1}_{p \times p}$  stands for the matrix whose entries are all equal to 1,  $\mathbb{I}_{ij}$  stands for the  $p_i \times p_j$  rectangular matrix whose main diagonal entries are equal to 1 and the others are equal to 0, and  $\mathbb{1}_{23}$  stands for the  $p_2 \times p_3$  rectangular matrix whose first entry is equal to 1 and the others are equal to 0. The empirical sizes and powers are obtained based on 100,000 replications with 3 digits. In the tables,  $T_0$  denotes the proposed Schott type test,  $T_1$  denotes the renormalized likelihood ratio test of [8] and  $T_2$  denotes the trace criterion test of [7].

The empirical sizes and powers of three tests are presented in Table 1 and Table 2. It shows that the proposed Schott type test  $T_0$  performs quite robust



TABLE 1  
 Empirical sizes (Scenario I) and empirical powers (Scenario II and III) of tests  $T_0$ ,  $T_1$  and  $T_2$  at the 5% significance level.

$(p_1, p_2, p_3)$	$n$	Scenario I			Scenario II			Scenario III		
		$T_0$	$T_1$	$T_2$	$T_0$	$T_1$	$T_2$	$T_0$	$T_1$	$T_2$
(2, 2, 1)	4	0.062	N.A.	N.A.	0.067	N.A.	N.A.	0.066	N.A.	N.A.
	6	0.056	N.A.	0.037	0.070	N.A.	0.042	0.069	N.A.	0.038
	10	0.058	0.077	0.051	0.099	0.110	0.078	0.091	0.105	0.068
	16	0.062	0.074	0.060	0.152	0.152	0.132	0.124	0.132	0.104
	30	0.064	0.071	0.064	0.280	0.262	0.250	0.216	0.221	0.199
(12, 12, 6)	50	0.067	0.070	0.066	0.457	0.427	0.420	0.354	0.360	0.341
	20	0.049	N.A.	N.A.	0.106	N.A.	N.A.	0.087	N.A.	N.A.
	30	0.050	N.A.	0.047	0.229	N.A.	0.126	0.140	N.A.	0.071
	40	0.050	0.059	0.049	0.399	0.293	0.225	0.218	0.167	0.128
	50	0.051	0.058	0.050	0.584	0.478	0.355	0.302	0.246	0.199
(24, 24, 12)	100	0.053	0.056	0.053	0.989	0.976	0.922	0.763	0.725	0.683
	150	0.054	0.055	0.053	1.000	1.000	1.000	0.965	0.957	0.946
	60	0.050	N.A.	0.048	0.416	N.A.	0.190	0.323	N.A.	0.119
	90	0.052	0.055	0.051	0.816	0.757	0.487	0.630	0.478	0.397
	110	0.050	0.053	0.050	0.948	0.936	0.692	0.802	0.692	0.617
(36, 36, 18)	130	0.052	0.054	0.052	0.991	0.990	0.852	0.911	0.850	0.796
	150	0.051	0.054	0.051	0.999	1.000	0.944	0.965	0.938	0.909
	180	0.051	0.053	0.050	1.000	1.000	0.992	0.994	0.989	0.982
	100	0.050	0.054	0.048	0.642	0.480	0.302	0.329	0.344	0.282
	150	0.051	0.054	0.051	0.966	0.976	0.704	0.954	0.875	0.809
	180	0.050	0.052	0.049	0.996	0.999	0.876	0.993	0.975	0.953
	210	0.052	0.052	0.051	1.000	1.000	0.961	0.999	0.997	0.993
	250	0.051	0.052	0.051	1.000	1.000	0.996	1.000	1.000	1.000
	300	0.052	0.053	0.052	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 2  
 Empirical powers of tests  $T_0$ ,  $T_1$  and  $T_2$  at the 5% significance level.

$(p_1, p_2, p_3)$	$n$	Scenario IV			Scenario V			Scenario VI		
		$T_0$	$T_1$	$T_2$	$T_0$	$T_1$	$T_2$	$T_0$	$T_1$	$T_2$
(2, 2, 1)	4	0.064	N.A.	N.A.	0.063	N.A.	N.A.	0.094	N.A.	N.A.
	6	0.062	N.A.	0.043	0.058	N.A.	0.036	0.147	N.A.	0.048
	10	0.074	0.095	0.069	0.068	0.086	0.056	0.289	0.368	0.186
	16	0.099	0.116	0.102	0.081	0.094	0.075	0.571	0.707	0.524
	30	0.161	0.173	0.166	0.111	0.119	0.107	0.962	0.976	0.961
(12, 12, 6)	50	0.261	0.271	0.266	0.155	0.161	0.151	1.000	1.000	1.000
	20	0.063	N.A.	N.A.	0.060	N.A.	N.A.	0.080	N.A.	N.A.
	30	0.091	N.A.	0.104	0.075	N.A.	0.053	0.126	N.A.	0.068
	40	0.128	0.128	0.143	0.092	0.089	0.069	0.178	0.198	0.112
	50	0.176	0.178	0.194	0.112	0.106	0.089	0.249	0.317	0.172
(24, 24, 12)	100	0.515	0.520	0.539	0.242	0.230	0.210	0.682	0.870	0.607
	150	0.821	0.825	0.835	0.406	0.396	0.372	0.949	0.995	0.928
	60	0.159	N.A.	0.207	0.114	N.A.	0.067	0.129	N.A.	0.074
	90	0.341	0.319	0.398	0.183	0.151	0.128	0.217	0.260	0.147
	110	0.498	0.482	0.554	0.243	0.206	0.178	0.293	0.387	0.214
(36, 36, 18)	130	0.644	0.635	0.692	0.310	0.270	0.240	0.376	0.52	0.289
	150	0.771	0.767	0.809	0.381	0.340	0.308	0.466	0.651	0.375
	180	0.904	0.904	0.924	0.490	0.453	0.418	0.607	0.814	0.519
	100	0.322	0.234	0.418	0.191	0.123	0.104	0.146	0.133	0.088
	150	0.699	0.665	0.769	0.352	0.278	0.241	0.254	0.323	0.179
	180	0.863	0.850	0.904	0.466	0.393	0.349	0.336	0.458	0.253
	210	0.952	0.948	0.969	0.579	0.513	0.467	0.424	0.593	0.335
	250	0.992	0.991	0.995	0.719	0.662	0.621	0.544	0.754	0.454
	300	1.000	1.000	1.000	0.854	0.820	0.789	0.694	0.893	0.612

under the null hypothesis. Even when  $p_1 = 2, p_2 = 2, p_3 = 1$  and  $n = 4$  the attained significance levels are also close to 6%. It is obvious that the empirical sizes of  $T_0$  is good enough in the inapplicable cases of  $T_1$  and  $T_2$ . Furthermore, for these applicable cases, when the dimensions and sample sizes are big, all the

three tests perform good. But when the sample sizes  $n$  are close to the total of dimensions  $p$ , test  $T_1$  performs worse than  $T_0$  and  $T_2$ .

For the power comparisons, in Scenario II and III, which are introduced in [8] and [7] respectively, all the powers tend to 1 as the sample size increases, and  $T_0$  performs better than  $T_1$  and  $T_2$  in most of the settings. It is worthy to notice that when the sample size  $n$  is big, the renormalized likelihood ratio test  $T_1$  is also satisfactory, and better than  $T_2$ . What is more, when the dimensions and sample sizes are big,  $T_0$  also shows to be the most powerful one among the three tests in Scenario V. But if the dimensions are small,  $T_1$  seems slightly better. In Scenario IV,  $T_2$  performs extremely good, especially when the dimensions and sample sizes are big. In this situation,  $T_0$  is also acceptable. Scenario VI is the so called spike model in RMT, and from the result we can find that all the three tests failed, because their powers do not tend to 1 as the sample size increases. Therefore, by the numerical results, we recommend the proposed test  $T_0$  for testing the independence of high dimensional random vectors due to its stability.

## 5. An example

For illustration, we apply the proposed test statistic to the daily returns of 258 stocks issued by the companies from S&P 500. The original data are the closing prices or the bid/ask average of these stocks for the trading days of the last quarter in 2013, i.e., from 1 October 2013 to 31 December 2013, with total 64 days. This dataset is derived from the Center for Research in Security Prices Daily Stock in Wharton Research Data Services. According to The North American Industry Classification System (NAICS), which is used by business and government to classify business establishments, the 258 stocks are separated into 11 sectors. Numbers of stocks in each sector are shown in Table 3. A common interest here is to test whether the daily returns for the investigated 11 sectors are independent.

The testing model is established as follows: Denote  $p_i$  as the number of stocks in the  $i$ th sector,  $u_{il}(j)$  as the price for the  $l$ th stock in the  $i$ th sector at day  $j$ . Here  $j \in \llbracket 64 \rrbracket$ . Correspondingly we have  $p = \sum_{i=1}^{11} p_i = 258$ . In order to satisfy the condition of the proposed test statistic, the original data  $u_{il}(j)$  need to be transformed as follows: (i) *logarithmic difference*: Let  $x_{il}(j) = \ln(u_{il}(j) +$

TABLE 3

*Number of stocks in each NAICS Sectors. Sector 1 describes mining, quarrying, and oil and gas extraction; Sector 2 describes utilities; Sector 3 describes wholesale trade; Sector 4 describes retail trade; Sector 5 describes transportation and warehousing; Sector 6 describes information; Sector 7 describes finance and insurance; Sector 8 describes real estate and rental and leasing; Sector 9 describes professional, scientific, and technical services; Sector 10 describes administrative and support and waste management and remediation services; Sector 11 describes arts, entertainment, and recreation.*

Sector	1	2	3	4	5	6	7	8	9	10	11
Number of stocks	30	32	14	32	12	33	55	14	16	10	10

$1)/u_{il}(j)$ ). Notice that  $j \in \llbracket 63 \rrbracket$ . So we denote  $n = 63$ . Logarithmic difference is a very commonly used procedure in finance. There are a number of theoretical and practical advantages of using logarithmic returns, especially we shall assume that the sequence of logarithmic returns are independent of each other for big time scales (e.g.  $\geq 1$  day, see [18]). (ii) *power transform*: It is well known that if a security price follows geometric Brownian motion, then the logarithmic returns of the security are normally distributed. However in most cases, the normalized logarithmic returns  $x_{il}(j)$  are considered to have sharper peaks and heavier tails than the standard normal distribution. Thus we first transform  $x_{il}(j)$  to  $\hat{x}_{il}(j)$  by Box-Cox transformation, and then suppose the transformed data follows a standard normal distribution, that is

$$\tilde{x}_{il}(j) = \left( \frac{\hat{x}_{il}(j) - \bar{x}_{il}}{\hat{\sigma}_{il}} \right)^{\beta_{il}} \sim N(0, 1).$$

Here  $\beta_{il}$  is an unknown parameter,  $\bar{x}_{il}$  and  $\hat{\sigma}_{il}$  are the sample mean and sample standard deviation of  $\hat{x}_{il}(j)$ ,  $j \in \llbracket n \rrbracket$ .  $\beta_{il}$  can be estimated by

$$\frac{1}{n} \sum_{j=1}^n \tilde{x}_{il}^4(j) \approx \int_{a_{il}}^{b_{il}} t^4 d\Phi(t),$$

where  $a_{il} = \min_{j \in \llbracket n \rrbracket} \tilde{x}_{il}(j)$ ,  $b_{il} = \max_{j \in \llbracket n \rrbracket} \tilde{x}_{il}(j)$  and  $\Phi(t)$  is the standard normal distribution function. (iii) *normality test*: we use the Kolmogorov-Smirnov test to test whether the transformed prices of each stock are drawn from the standard normal distribution. By calculation, we get 258  $p$ -values. Among them the minimum is 0.1473. And 91.86% of  $p$ -values are bigger than 0.5. For illustration, we present the smoothed empirical densities of the transformed data  $\tilde{x}_{il}(j)$  for the first four stocks of each sector in Figure 1. From these graphs, we can also see that the transformed data fit the normal density curve well. Observe that, marginal normal cannot guarantee jointly normal. But still, here we perform our method on the transformed data, since as we mentioned in Remark 3.2, our result Theorem 3.1 is believed to hold under more generally distributed data. The transformation is done mainly for getting more moments matching to the Gaussian distribution. Theorem 2.2 in [28] indicates that moment matching condition like the fourth moment equal to 3 may be necessary for Theorem 3.1.

Now we apply  $\mathbf{s}(\mathbf{B})$  to test the independence of every two sectors. The  $p$ -values are shown in Table 4. We find in the total 55 pairs of sectors, there are 23 pairs with  $p$ -values bigger than 0.05 and 18 pairs with  $p$ -values bigger than 0.1. Interestingly, according to these results, if we set the significance level as 5% we find that there are seven sectors which are independent of Sector 7, the finance and insurance sector, which seems to be most independent of other sectors. On the other hand, Sector 11 (the arts, entertainment, and recreation sector) is dependent on all other sectors except Sector 1 (the mining, quarrying, and oil and gas extraction sector). We also investigate the mutual independence of every three sectors. Applying Theorem 3.1, we find that there are 11 groups with  $p$ -values bigger than 0.05. In addition, we find that there is only one group

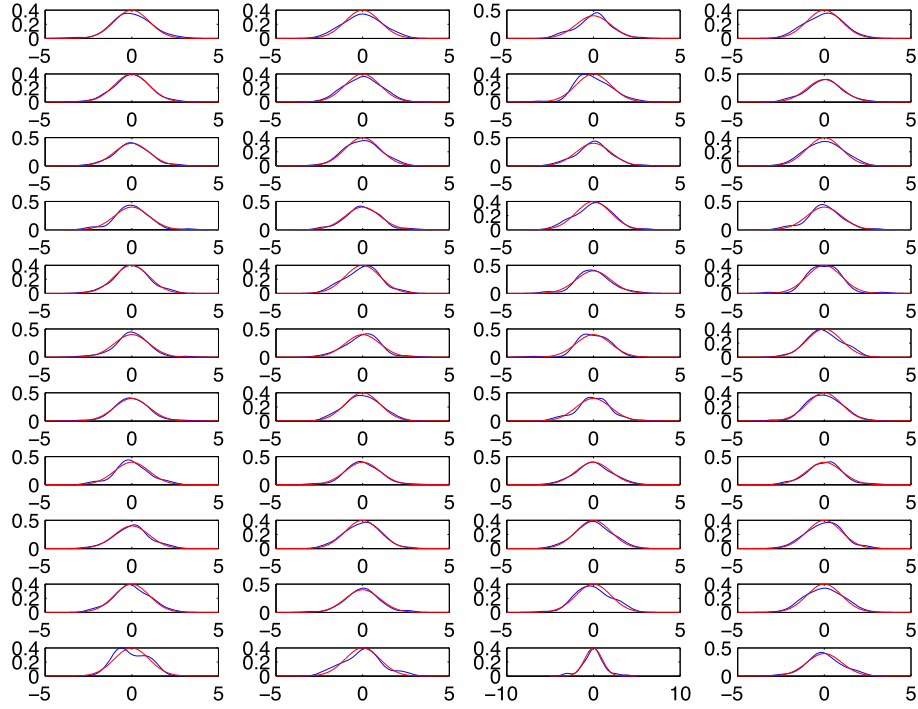


FIG 1. Graphs of the empirical density function of the transformed data versus the standard normal distribution. These graphs contain the empirical density functions of the transformed data for the first four stocks of each sector used in our study. The blue curve is the smoothed density function of the transformed data for one stock and the red curve is standard normal density function.

TABLE 4

The  $p$ -values obtained by the proposed test under  $\mathbf{H}_0$  with  $k = 2$ . Notice that the results are rounded up to the fourth decimal point.

Sector	1	2	3	4	5	6	7	8	9	10
2	0.0405									
3	0.0487	0.2735								
4	0.0604	0	0.0002							
5	0.0012	0.1041	0.0027	0.1639						
6	0.0073	0.0048	0.0008	0.4285	0.0444					
7	0.2299	0.4830	0.0451	0.1053	0.7080	0.6843				
8	0.3558	0.0208	0.2458	0.3547	0.0013	0.0645	0.1127			
9	0.1411	0.0833	0	0.0005	0.2403	0.0124	0.4521	0.0048		
10	0.0418	0.3075	0.0004	0.0847	0.0746	0.0026	0.0036	0.0109	0	
11	0.8689	0.0048	0	0.0470	0.0167	0.0004	0.0490	0	0.0003	0.0003

containing four sectors which are mutually independent. The results are shown in Table 5. Thus, we have strong evidence to believe that every five sectors are dependent.

TABLE 5

The  $p$ -values obtained by the proposed test under  $\mathbf{H}_0$  with  $k = 3$  and  $k = 4$ . Notice that the results are rounded up to the fourth decimal point.

Sectors	(1,4,8)	(1,7,8)	(1,7,9)	(2,5,7)	(2,7,9)	(4,5,7)
$p$ -value	0.0804	0.1135	0.1120	0.2541	0.1413	0.1263
Sectors	(4,6,7)	(4,6,8)	(4,7,8)	(5,7,9)	(6,7,8)	(4,6,7,8)
$p$ -value	0.3088	0.1277	0.0686	0.3913	0.0855	0.0650

### 6. Linear spectral statistics and second order freeness

In this section, we will introduce some **RMT** and **FPT** apparatus with which Theorem 3.1 can be proved. We will just summarize the main ideas on how these tools fit into the framework of the limiting behavior of our proposed statistic, but leave the details of the reasoning and calculation to the Appendix. We start from the following elementary fact. To wit, for two matrices  $\mathbf{S}$  and  $\mathbf{T}$ , we know that  $\mathbf{ST}$  and  $\mathbf{TS}$  share the same non-zero eigenvalues, as long as both  $\mathbf{ST}$  and  $\mathbf{TS}$  are square. Therefore, to study the eigenvalues of  $\mathbf{B}$ , it is equivalent to study the eigenvalues of the  $(n - 1) \times (n - 1)$  matrix

$$\underline{\mathbf{B}} := \mathbf{Z}' \text{diag}[(\mathbf{Z}_i \mathbf{Z}_i')^{-1}]_{i=1}^k \mathbf{Z} = \sum_{i=1}^k \mathbf{Z}_i' (\mathbf{Z}_i \mathbf{Z}_i')^{-1} \mathbf{Z}_i.$$

Setting

$$s(\underline{\mathbf{B}}) = \frac{1}{2} \sum_{\ell=1}^{n-1} \lambda_{\ell}^2(\underline{\mathbf{B}}) - \frac{p}{2},$$

by the above discussion we can assert

$$s(\underline{\mathbf{B}}) = s(\mathbf{B}).$$

A main advantage of  $\underline{\mathbf{B}}$  is embodied in the following proposition, which will be a starting point of the proof of Theorem 3.1.

**Proposition 6.1.** *Assume that  $\mathbf{O}_1, \dots, \mathbf{O}_k$  are i.i.d.  $(n - 1)$ -dimensional random orthogonal matrices possessing Haar distribution on the orthogonal group  $O(n - 1)$ . Let  $\mathbf{P}_i = \mathbf{I}_{p_i} \oplus \mathbf{0}_{n-1-p_i}$  be a diagonal projection matrix with rank  $p_i$ , for each  $i \in \llbracket k \rrbracket$ . Here we denote by  $\mathbf{O}_m$  the  $m \times m$  null matrix. Then under  $\mathbf{H}_0$ , we have*

$$\underline{\mathbf{B}} \stackrel{d}{=} \sum_{i=1}^k \mathbf{O}_i' \mathbf{P}_i \mathbf{O}_i.$$

*Proof.* We perform the singular value decomposition as  $\mathbf{Z}_i := \mathbf{U}_i' \mathbf{D}_i \mathbf{O}_i$ , where  $\mathbf{U}_i$  and  $\mathbf{O}_i$  are  $p_i$ -dimensional and  $(n - 1)$ -dimensional orthogonal matrices respectively, and  $\mathbf{D}_i$  is a  $p_i \times (n - 1)$  rectangular matrix whose main diagonal entries are nonzero while the others are all zero. It is well-known that when  $\mathbf{Z}_i$  has i.i.d. normal entries, both  $\mathbf{U}_i$  and  $\mathbf{O}_i$  are Haar distributed. Then an elementary calculation leads to Proposition 1.  $\square$

Note that Proposition 6.1 allows us to study the eigenvalues of a summation of  $k$  independent random projections instead. Moreover, it is obvious that under  $\mathbf{H}_0$ , this summation of random matrices does not depend on the unknown population mean vectors and covariance matrices of  $\mathbf{x}_i$ ,  $i \in \llbracket k \rrbracket$ .

To compress notation, we set

$$\mathbf{Q} := \sum_{i=1}^k \mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i, \quad \mathbf{Q} \stackrel{d}{=} \mathbf{B}. \quad (6.1)$$

According to the discussion in the last section, we know that our Schott type statistic can be expressed (in distribution) in terms of the eigenvalues of  $\mathbf{Q}$ , since

$$\mathrm{tr} \mathbf{B}^2 \stackrel{d}{=} \mathrm{tr} \mathbf{Q}^2 = \sum_{\ell=1}^{n-1} \lambda_{\ell}^2(\mathbf{Q}).$$

In **RMT**, given an  $N \times N$  random matrix  $\mathbf{A}$  and some test function  $f : \mathbb{C} \rightarrow \mathbb{C}$ , one usually calls the quantity

$$\mathcal{L}_N[f] := \sum_{\ell=1}^N f(\lambda_{\ell}(\mathbf{A}))$$

a *linear spectral statistic* of  $\mathbf{A}$  with test function  $f$ . For some classical random matrix models such as Wigner matrices and sample covariance matrices, linear spectral statistics have been widely studied. Not trying to be comprehensive, one can refer to [3], [9], [16], [25] and [22] for instance. A notable feature in this type of CLTs is that usually the variance of the linear spectral statistic is of order  $O(1)$  when the test function  $f$  is smooth enough, mainly due to the strong correlations among eigenvalues, thus is significantly different from the case of i.i.d variables. Now, in a similar vein, with the random matrix  $\mathbf{Q}$  at hand, we want to study the fluctuation of its linear spectral statistics, focusing on the test function  $f(x) = x^2$ .

In the past few decades, the main stream of **RMT** has focused on the spectral behavior of single random matrix models such as Wigner matrix, sample covariance matrix and non-Hermitian matrix with i.i.d. variables. However, with the rapid development in **RMT** and its related fields, the study of general polynomials with classical single random matrices as its variables is in increasing demand. A favorable idea is to derive the spectral properties of the matrix polynomial from the information of the spectrums of its variables (single matrices). Specifically, the question can be described as

*Given the eigenvalues of  $\mathbf{A}$  and  $\mathbf{B}$ , what can one say about the eigenvalues of  $h(\mathbf{A}, \mathbf{B})$ ?*

Here  $h(\cdot, \cdot)$  is a bivariate polynomial. Usually, for deterministic matrices  $\mathbf{A}$  and

$\mathbf{B}$ , only with their eigenvalues given, it is impossible to write down the eigenvalues of  $h(\mathbf{A}, \mathbf{B})$ . However, for some independent high-dimensional random matrices  $\mathbf{A}$  and  $\mathbf{B}$ , deriving the limiting spectral properties of  $h(\mathbf{A}, \mathbf{B})$  via those of  $\mathbf{A}$  and  $\mathbf{B}$  is possible. To answer this kind of question, the right machinery to employ is **FPT**. In the breakthrough work by Voiculescu in [26], the author proved that if  $(\mathbf{A}_n)_{n \in \mathbb{N}}$  and  $(\mathbf{B}_n)_{n \in \mathbb{N}}$  are two independent sequences of random Hermitian matrices and at least one of them is orthogonally invariant (in distribution), then they satisfy the property of *asymptotic freeness*, which allows one to derive the limit of  $\frac{1}{n} \mathbb{E} \text{tr} h(\mathbf{A}_n, \mathbf{B}_n)$  from the limits of  $(\frac{1}{n} \mathbb{E} \text{tr} \mathbf{A}_n^m)_{m \in \mathbb{N}}$  and  $(\frac{1}{n} \mathbb{E} \text{tr} \mathbf{B}_n^m)_{m \in \mathbb{N}}$  directly. Sometimes we also call the asymptotic freeness of two random matrix sequences as *first order freeness*.

Our aim in this paper, however, is not to derive the limit of the normalized trace of some polynomial in random matrices, but to take a step further to study the fluctuation of the trace. To this end, we need to adopt the concept of *second order freeness*, which was recently raised and developed in the series of work: [11, 12, 5, 10], also see [20]. In contrast, the second order freeness aims at answering how to derive the fluctuation property of  $\text{tr} h(\mathbf{A}_n, \mathbf{B}_n)$  from the limiting spectral properties of  $\mathbf{A}_n$  and  $\mathbf{B}_n$ .

Especially, in [10], the authors established the so-called *real second order freeness* for orthogonal matrices, which is specialized in solving the fluctuation of the linear spectral statistics of polynomials in Haar distributed orthogonal matrices and deterministic matrices. For our purpose, we need to employ Proposition 52 in [10], an ad hoc and simplified version of which can be heuristically sketched as follows. Assume that  $\{\mathbf{A}_n\}_{n \in \mathbb{N}}$  and  $\{\mathbf{B}_n\}_{n \in \mathbb{N}}$  are two independent sequences of random matrices (may be deterministic), where  $\mathbf{A}_n$  and  $\mathbf{B}_n$  are  $n$  by  $n$ , and the limits of  $n^{-1} \mathbb{E} \text{tr} h_1(\mathbf{A}_n)$  and  $n^{-1} \mathbb{E} \text{tr} h_2(\mathbf{B}_n)$  exist for any given polynomials  $h_1$  and  $h_2$ , as  $n \rightarrow \infty$ . Moreover,  $\text{tr} h_1(\mathbf{A}_n)$  and  $\text{tr} h_2(\mathbf{B}_n)$  possess Gaussian fluctuations (may be degenerate) asymptotically for any given polynomials  $h_1$  and  $h_2$ . Then  $\text{tr} q(\mathbf{O}' \mathbf{A}_n \mathbf{O}, \mathbf{B}_n)$  also possesses Gaussian fluctuation asymptotically for any given bivariate polynomial  $q(\cdot, \cdot)$ , where  $\mathbf{O}$  is supposed to be an  $n \times n$  Haar orthogonal matrix independent of  $\mathbf{A}_n$  and  $\mathbf{B}_n$ .

Now as for  $\mathbf{Q}$ , we can start from the case of  $k = 2$ , which fits the above framework quite well. To wit, we can regard  $\mathbf{O}'_1 \mathbf{P}_1 \mathbf{O}_1$  as  $\mathbf{B}_{n-1}$  and  $\mathbf{P}_2$  as  $\mathbf{A}_{n-1}$ , using Proposition 52 of [10] leads to the fact that  $\text{tr} h(\mathbf{O}'_1 \mathbf{P}_1 \mathbf{O}_1 + \mathbf{O}'_2 \mathbf{P}_2 \mathbf{O}_2)$  is asymptotically Gaussian after an appropriate normalization, for any given polynomial  $h$ . Then we take  $\mathbf{O}'_1 \mathbf{P}_1 \mathbf{O}_1 + \mathbf{O}'_2 \mathbf{P}_2 \mathbf{O}_2$  as  $\mathbf{B}_{n-1}$  and regard  $\mathbf{P}_3$  as  $\mathbf{A}_{n-1}$  and repeat the above discussion. By using Proposition 52 in [10] recursively, we can get our CLT for  $\mathbf{Q}$  finally. A formal result which can be derived from Proposition 52 in [10] is as follows, whose proof will be presented in the Appendix.

**Theorem 6.1.** *For our matrix  $\mathbf{Q}$  defined in (6.1), and any deterministic polynomial sequence  $h_1, h_2, h_3, \dots$ , we have*

$$\text{Cov}(\text{tr} h_1(\mathbf{Q}), \text{tr} h_2(\mathbf{Q})) = O(1)$$

and

$$\lim_{n \rightarrow \infty} \kappa_r(\text{tr } h_1(\mathbf{Q}), \dots, \text{tr } h_r(\mathbf{Q})) = 0, \quad \text{if } r \geq 3$$

where  $\kappa_r(\xi_1, \dots, \xi_r)$  represents the joint cumulant of the random variables  $\xi_1, \dots, \xi_r$ .

Now if we set  $h_i(x) = x^2$  for all  $i \in \mathbb{N}$ , we can obtain Theorem 3.1 by proving the following lemma.

**Lemma 6.1.** *Under the above notation, we have*

$$\mathbb{E} \text{tr} \mathbf{Q}^2 = p + \sum_{i,j,i \neq j} \frac{p_i p_j}{n-1}, \quad (6.2)$$

and

$$\text{Var}(\text{tr} \mathbf{Q}^2) = 4 \sum_{i,j,i \neq j} \frac{p_i p_j (n-1-p_i)(n-1-p_j)}{(n-1)^4} + O\left(\frac{1}{n}\right). \quad (6.3)$$

The proof of Lemma 6.1 will also be stated in the Appendix. In the sequel, we prove Theorem 3.1 with Lemma 6.1 granted.

**Proof of Theorem 3.1.** Setting  $h_i(x) = x^2$  for all  $i \in \mathbb{N}$  in Theorem 6.1, we see that all  $r$ th cumulants of  $\text{tr} \mathbf{Q}^2$  tend to 0 if  $r \geq 3$  when  $n \rightarrow \infty$ , which together with Lemma 6.1 implies that

$$\frac{\text{tr} \mathbf{Q}^2 - \mathbb{E} \text{tr} \mathbf{Q}^2}{\sqrt{\text{Var}(\text{tr} \mathbf{Q}^2)}} \implies N(0, 1).$$

Then Theorem 3.1 follows from the definition of  $\mathbf{s}(\mathbf{B})$  directly.  $\square$

## Appendix

In this Appendix, we provide the proofs of Theorem 6.1 and Lemma 6.1.

**Proof of Theorem 6.1.** In Proposition 52 of [10], the authors state that if  $\{\mathbf{A}_n\}_{n \in \mathbb{N}}$  and  $\{\mathbf{B}_n\}_{n \in \mathbb{N}}$  are two independent sequences of random matrices (may be deterministic), where  $\mathbf{A}_n$  and  $\mathbf{B}_n$  are  $n \times n$ , each having a *real second order limit distribution*, and  $\mathbf{O}$  is supposed to be an  $n \times n$  Haar orthogonal matrix independent of  $\mathbf{A}_n$  and  $\mathbf{B}_n$ , then  $\mathbf{B}_n$  and  $\mathbf{O}' \mathbf{A}_n \mathbf{O}$  are *asymptotically real second order free*. By Definition 30 of [10], we see that for a single random matrix sequence  $\{\mathbf{A}_n\}_{n \in \mathbb{N}}$ , the existence of the so-called *real second order limit distribution* means that the following three statements hold simultaneously for any given sequence of polynomials  $h_1, h_2, h_3, \dots$  when  $n \rightarrow \infty$ :

- 1)  $n^{-1} \mathbb{E} \text{tr} h_1(\mathbf{A}_n)$  converges;
- 2)  $\text{Cov}(\text{tr} h_1(\mathbf{A}_n), \text{tr} h_2(\mathbf{A}_n))$  converges (the limit can be 0);
- 3)  $\kappa_r(\text{tr} h_1(\mathbf{A}_n), \dots, \text{tr} h_r(\mathbf{A}_n)) = o(1)$  for all  $r \geq 3$ .



We stress here, the original definition in [10] is given with a language of non-commutative probability theory. To avoid introducing too many additional notions, we just modify it to be the above 1)-3). Then by the Definition 33 and Proposition 52 of [10]), one see that if both  $\{\mathbf{A}_n\}_{n \in \mathbb{N}}$  and  $\{\mathbf{B}_n\}_{n \in \mathbb{N}}$  have real second order limit distributions, we have the following three facts for any given sequences of bivariate polynomials  $q_1, q_2, q_3, \dots$  when  $n \rightarrow \infty$ :

- 1')  $n^{-1} \mathbb{E} \text{tr} q_1(\mathbf{B}_n, \mathbf{O}' \mathbf{A}_n \mathbf{O})$  converges;
- 2')  $\text{Cov}(\text{tr} q_1(\mathbf{B}_n, \mathbf{O}' \mathbf{A}_n \mathbf{O}), \text{tr} q_2(\mathbf{B}_n, \mathbf{O}' \mathbf{A}_n \mathbf{O}))$  converges;
- 3')  $\kappa_r(\text{tr} q_1(\mathbf{B}_n, \mathbf{O}' \mathbf{A}_n \mathbf{O}), \dots, \text{tr} q_r(\mathbf{B}_n, \mathbf{O}' \mathbf{A}_n \mathbf{O})) = o(1)$  for all  $r \geq 3$ .

Here 1') and 2') can be implied by the definitions of the first and second order freeness in [10] respectively, and 3') can be found in the proof of Proposition 52 of [10], where the authors claim that the proof of Theorem 41 therein is also applicable under the setting of Proposition 52. Note that in [10], a more concrete rule to determine the limit of

$$\text{Cov}(\text{tr} q_1(\mathbf{B}_n, \mathbf{O}' \mathbf{A}_n \mathbf{O}), \text{tr} q_2(\mathbf{B}_n, \mathbf{O}' \mathbf{A}_n \mathbf{O}))$$

is given, which can be viewed as the core of the concept of *real second order freeness*. However, here we do not need such a concrete rule, thus do not introduce it.

Now we are at the stage of employing Proposition 52 of [10] to prove Theorem 6.1. Recall our objective  $\mathbf{Q}$  defined in (6.1). We start from the case of  $k = 2$ , to wit, we are considering the linear spectral statistics of the random matrix  $\mathbf{O}'_1 \mathbf{P}_1 \mathbf{O}_1 + \mathbf{O}'_2 \mathbf{P}_2 \mathbf{O}_2$ . Now, we regard  $\mathbf{O}'_1 \mathbf{P}_1 \mathbf{O}_1$  as  $\mathbf{B}_{n-1}$  and  $\mathbf{P}_2$  as  $\mathbf{A}_{n-1}$ , then obviously they both satisfy 1)-3) in the definition of the existence of the *real second order limit distribution*, since the spectrums of  $\mathbf{A}_{n-1}$  and  $\mathbf{B}_{n-1}$  are both deterministic, noticing they are projection matrices with known ranks. Then 1')-3') immediately imply that  $\mathbf{O}'_1 \mathbf{P}_1 \mathbf{O}_1 + \mathbf{O}'_2 \mathbf{P}_2 \mathbf{O}_2$  also has a real second order limit distribution. Next, adding  $\mathbf{O}'_3 \mathbf{P}_3 \mathbf{O}_3$  to  $\mathbf{O}'_1 \mathbf{P}_1 \mathbf{O}_1 + \mathbf{O}'_2 \mathbf{P}_2 \mathbf{O}_2$ , and regarding the latter as  $\mathbf{B}_{n-1}$  and  $\mathbf{P}_3$  as  $\mathbf{A}_{n-1}$ , we can use the above discussion again to conclude that  $\mathbf{O}'_1 \mathbf{P}_1 \mathbf{O}_1 + \mathbf{O}'_2 \mathbf{P}_2 \mathbf{O}_2 + \mathbf{O}'_3 \mathbf{P}_3 \mathbf{O}_3$  also possesses a real second order limit distribution. Recursively, we can finally get that  $\mathbf{Q}$  has a real second order limit distribution, which implies Theorem 6.1. So we conclude the proof.  $\square$

It remains to prove Lemma 6.1. Before commencing the proof, we briefly introduce some technical inputs. Since the trace of a product of matrices can always be expressed in terms of some products of their entries, it is expected that we will need to calculate the quantities of the form

$$\mathbb{E} \mathbf{O}_{i_1 j_1} \cdots \mathbf{O}_{i_m j_m}, \tag{6.4}$$

where  $\mathbf{O}$  is assumed to be an  $N$ -dimensional Haar distributed orthogonal matrix, and  $\mathbf{O}_{ij}$  is its  $(i, j)$ th entry. A powerful tool handling this kind of expectation is the so-called *Weingarten calculus on orthogonal group*, we refer to the seminal paper of [6], formula (21) therein. To avoid introducing too many combinatorics notions for Weingarten calculus, we just list some consequence of it for our

purpose, taking into account the fact that we will only need to handle the case of  $m \leq 4$  in (6.4) in the sequel. Specifically, we have the following lemma.

**Lemma 6.2.** *Under the above notation, we have the following facts for (6.4), assuming  $m \leq 4$  and  $\mathbf{i} = (i_1, \dots, i_m)$  and  $\mathbf{j} = (j_1, \dots, j_m)$ .*

- 1) When  $m = 2$ ,  $i_1 = i_2$ , and  $j_1 = j_2$ , we have (6.4) =  $N^{-1}$ ,
- 2) When  $m = 4$ , we have the following results for four sub-cases.
  - i) If  $i_1 = i_2 = i_3 = i_4$ ,  $j_1 = j_2 = j_3 = j_4$ , we have (6.4) =  $3/(N(N+2))$ ;
  - ii) If  $i_1 = i_2 = i_3 = i_4$ ,  $j_1 = j_2 \neq j_3 = j_4$ , we have (6.4) =  $1/(N(N+2))$ ;
  - iii) If  $i_1 = i_2 \neq i_3 = i_4$ ,  $j_1 = j_2 \neq j_3 = j_4$ , we have (6.4) =  $(N + 1)/(N(N - 1)(N + 2))$ ;
  - iv) If  $i_1 = i_3 \neq i_2 = i_4$ ,  $j_1 = j_2 \neq j_3 = j_4$ , we have (6.4) =  $-1/(N(N - 1)(N + 2))$ .
- 3) Replacing  $\mathbf{O}$  by  $\mathbf{O}'$ , we can obviously switch the roles of  $\mathbf{i}$  and  $\mathbf{j}$  in 1) and 2). Moreover, any permutation on the indices  $\{1, \dots, m\}$  will not change (6.4). Any other triple  $(m, \mathbf{i}, \mathbf{j})$ , which can not be transformed into any case in 1) or 2) via switching the roles of  $\mathbf{i}$  and  $\mathbf{j}$  or performing permutations on the indices  $\{1, \dots, m\}$ , will drives (6.4) to be 0.

With Lemma 6.2 at hand, we can prove Lemma 6.1 in the sequel.

**Proof of Lemma 6.1.** At first, we verify (6.2). Note that by definition,

$$\begin{aligned} \mathbb{E}\text{tr}\mathbf{Q}^2 &= \sum_{i=1}^k \mathbb{E}\text{tr}(\mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i)^2 + \sum_{i,j,i \neq j} \mathbb{E}\text{tr}(\mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i \cdot \mathbf{O}'_j \mathbf{P}_j \mathbf{O}_j) \\ &= p + \sum_{i,j,i \neq j} \mathbb{E}\text{tr}(\mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i \cdot \mathbf{O}'_j \mathbf{P}_j \mathbf{O}_j). \end{aligned}$$

Let  $\mathbf{u}_i(\ell)$  be the  $\ell$ th column of  $\mathbf{O}'_i$  and  $\mathbf{u}_i(\ell, s)$  be the  $s$ th coefficient of  $\mathbf{u}_i(\ell)$ , i.e. the  $(\ell, s)$ th entry of  $\mathbf{O}_i$ , for all  $i \in \llbracket k \rrbracket$ . Then for  $i \neq j$  we have

$$\begin{aligned} \mathbb{E}\text{tr}(\mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i \cdot \mathbf{O}'_j \mathbf{P}_j \mathbf{O}_j) &= \mathbb{E} \sum_{m=1}^{p_i} \sum_{\ell=1}^{p_j} \text{tr} \mathbf{u}_i(m) \mathbf{u}'_i(m) \mathbf{u}_j(\ell) \mathbf{u}'_j(\ell) \\ &= \sum_{m=1}^{p_i} \sum_{\ell=1}^{p_j} \sum_s \sum_t \mathbb{E} \mathbf{u}_i(m, s) \mathbf{u}_i(m, t) \cdot \mathbb{E} \mathbf{u}_j(\ell, s) \mathbf{u}_j(\ell, t) \\ &= \sum_{m=1}^{p_i} \sum_{\ell=1}^{p_j} \sum_s \mathbb{E} (\mathbf{u}_i(m, s))^2 \mathbb{E} (\mathbf{u}_j(\ell, s))^2 = \frac{p_i p_j}{n-1}. \end{aligned} \tag{6.5}$$

Here, in the third step above we used 3) of Lemma 6.2 to discard the terms with  $s \neq t$ , while in the last step we used 1) of Lemma 6.2. Therefore, we have

$$\mathbb{E}\text{tr}\left(\sum_{i=1}^k \mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i\right)^2 = p + \sum_{i,j,i \neq j} \frac{p_i p_j}{n-1}, \tag{6.6}$$

Now we calculate  $\text{Var}(\text{tr}\mathbf{Q}^2)$  as follows. Note that we have

$$\begin{aligned} \text{Var}(\text{tr}\mathbf{Q}^2) &= \mathbb{E}(\text{tr}\left(\sum_{i=1}^k \mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i\right)^2) - (\mathbb{E}\text{tr}\left(\sum_{i=1}^k \mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i\right)^2) \\ &= \mathbb{E}\left(\sum_{i,j,i \neq j} \text{tr}(\mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i \cdot \mathbf{O}'_j \mathbf{P}_j \mathbf{O}_j)\right)^2 - \left(\mathbb{E} \sum_{i,j,i \neq j} \text{tr}(\mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i \cdot \mathbf{O}'_j \mathbf{P}_j \mathbf{O}_j)\right)^2 \\ &= \sum_{i,j,i \neq j} \sum_{m,\ell,m \neq \ell} \text{Cov}(\text{tr}(\mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i \cdot \mathbf{O}'_j \mathbf{P}_j \mathbf{O}_j), \text{tr}(\mathbf{O}'_m \mathbf{P}_m \mathbf{O}_m \cdot \mathbf{O}'_\ell \mathbf{P}_\ell \mathbf{O}_\ell)) \\ &= \sum_{\substack{i,j,m,\ell \\ i \neq j, m \neq \ell \\ \{i,j\} \cap \{m,\ell\} \neq \emptyset}} \text{Cov}(\text{tr}(\mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i \cdot \mathbf{O}'_j \mathbf{P}_j \mathbf{O}_j), \text{tr}(\mathbf{O}'_m \mathbf{P}_m \mathbf{O}_m \cdot \mathbf{O}'_\ell \mathbf{P}_\ell \mathbf{O}_\ell)). \end{aligned}$$

In the sequel, we briefly write

$$\text{Cov}((i, j), (m, \ell)) := \text{Cov}(\text{tr}(\mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i \cdot \mathbf{O}'_j \mathbf{P}_j \mathbf{O}_j), \text{tr}(\mathbf{O}'_m \mathbf{P}_m \mathbf{O}_m \cdot \mathbf{O}'_\ell \mathbf{P}_\ell \mathbf{O}_\ell))$$

Note that the summation in the last step above can be decomposed into the following six cases.

- 1 :  $m = i, \ell = j$ ;      2 :  $m = j, \ell = i$ ;
- 3 :  $m = i, \ell \neq i, j$ ;    4 :  $\ell = i, m \neq i, j$ ;
- 5 :  $m = j, \ell \neq i, j$ ;    6 :  $\ell = j, m \neq i, j$ .

Now given  $i, j$ , we decompose the summation over  $m, \ell$  according to the above 6 cases and denote the sum restricted on these cases by  $\sum_{\alpha(i,j)}, \alpha(i, j) = 1, \dots, 6$  respectively. Therefore,

$$\text{Var}(\text{tr}\mathbf{Q}^2) = \sum_{i,j,i \neq j} \sum_{\alpha(i,j)=1}^6 \sum_{\alpha(i,j)} \text{Cov}((i, j), (m, \ell)) \tag{6.7}$$

Now by definition we have

$$\begin{aligned} \sum_{i,j,i \neq j} \sum_{\alpha(i,j)} \text{Cov}((i, j), (m, \ell)) &= \sum_i \sum_{j \neq i} \text{Cov}((i, j), (i, j)), & \alpha = 1, 2, \\ \sum_{i,j,i \neq j} \sum_{\alpha(i,j)} \text{Cov}((i, j), (m, \ell)) &= \sum_i \sum_{j \neq i} \sum_{\ell \neq i, j} \text{Cov}((i, j), (i, \ell)), & \alpha = 3, 4, \\ \sum_{i,j,i \neq j} \sum_{\alpha(i,j)} \text{Cov}((i, j), (m, \ell)) &= \sum_i \sum_{j \neq i} \sum_{\ell \neq i, j} \text{Cov}((i, j), (j, \ell)), & \alpha = 5, 6. \end{aligned}$$

Now note that for  $i \neq j$

$$\begin{aligned} \text{Cov}((i, j), (i, j)) &= \mathbb{E}(\text{tr}(\mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i \cdot \mathbf{O}'_j \mathbf{P}_j \mathbf{O}_j))^2 - (\mathbb{E} \text{tr}(\mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i \cdot \mathbf{O}'_j \mathbf{P}_j \mathbf{O}_j))^2 \\ &= \mathbb{E}(\text{tr}(\mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i \cdot \mathbf{O}'_j \mathbf{P}_j \mathbf{O}_j))^2 - \left(\frac{p_i p_j}{n-1}\right)^2, \end{aligned}$$

where the last step follows from (6.5). Moreover, we have

$$\begin{aligned} &\mathbb{E}(\text{tr}(\mathbf{O}'_i \mathbf{P}_i \mathbf{O}_i \cdot \mathbf{O}'_j \mathbf{P}_j \mathbf{O}_j))^2 \\ &= \sum_{m_1, m_2=1}^{p_i} \sum_{\ell_1, \ell_2=1}^{p_j} \mathbb{E} \mathbf{u}'_i(m_1) \mathbf{u}_j(\ell_1) \mathbf{u}'_j(\ell_1) \mathbf{u}_i(m_1) \mathbf{u}'_i(m_2) \mathbf{u}_j(\ell_2) \mathbf{u}'_j(\ell_2) \mathbf{u}_i(m_2) \\ &= \sum_{s_1, s_2, t_1, t_2} \left[ \sum_{m_1, m_2=1}^{p_i} \mathbb{E} \mathbf{u}_i(m_1, s_1) \mathbf{u}_i(m_1, t_1) \mathbf{u}_i(m_2, s_2) \mathbf{u}_i(m_2, t_2) \right] \\ &\quad \times \left[ \sum_{\ell_1, \ell_2=1}^{p_j} \mathbb{E} \mathbf{u}_j(\ell_1, s_1) \mathbf{u}_j(\ell_1, t_1) \mathbf{u}_j(\ell_2, s_2) \mathbf{u}_j(\ell_2, t_2) \right]. \end{aligned}$$

To calculate the above expectation, we need to use Lemma 6.2 again. In light of 3) of Lemma 6.2, it suffices to consider the following four cases

$$\begin{aligned} 1: s_1 = s_2 = t_1 = t_2, \quad 2: s_1 = s_2 \neq t_1 = t_2 \\ 3: s_1 = t_1 \neq s_2 = t_2, \quad 4: s_1 = t_2 \neq s_2 = t_1. \end{aligned}$$

Through detailed but elementary calculation, with the aid of Lemma 6.2, we can finally obtain that for  $i \neq j$ ,

$$\text{Cov}((i, j), (i, j)) = \frac{2p_i p_j (n-1-p_i)(n-1-p_j)}{(n-1)^4} + O\left(\frac{1}{n}\right).$$

Moreover, analogously, when  $i, j, \ell$  are mutually distinct, we can get

$$\text{Cov}((i, j), (i, \ell)) = \text{Cov}((i, j), (j, \ell)) = O\left(\frac{1}{n}\right)$$

by using Lemma 6.2. Here we just omit the details of the calculation. Consequently, by (6.7) one has

$$\text{Var}(\text{tr} \mathbf{Q}^2) = \sum_{i, j, i \neq j} \frac{4p_i p_j (n-1-p_i)(n-1-p_j)}{(n-1)^4} + O\left(\frac{1}{n}\right). \quad (6.8)$$

Thus we conclude the proof.  $\square$

### Acknowledgements

The authors would like to thank the anonymous referee and the associate editor for their invaluable and constructive comments, which led to substantial improvement of the paper.

## References

- [1] Anderson, T. (2003). An introduction to multivariate statistical analysis. Third Edition. *Wiley New York*.
- [2] Bai, Z. D., Jiang, D. D., Yao, J. F. and Zheng, S. R. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.*, **37(6B)**, 3822–3840. [MR2572444](#)
- [3] 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. [MR2040792](#)
- [4] Cai, T. T. and Jiang, T. F. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.*, **39(3)**, 1496–1525. [MR2850210](#)
- [5] Collins, B., Mingo, J. A., Śniady, P. and Speicher, R. (2007). Second order freeness and fluctuations of Random Matrices: III. Higher order freeness and free cumulants, *Doc. Math.*, **12**, 1–70.
- [6] Collins, B. and Śniady, P. (2006). Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Commun. Math. Phys.*, **264**, 773–795. [MR2217291](#)
- [7] Jiang, D. D., Bai, Z. D. and Zheng, S. R. (2013). Testing the independence of sets of large-dimensional variables. *Sci. China Math.*, **56(1)**, 135–147.
- [8] Jiang, T. F. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.*, **41(4)**, 2029–2074. [MR3127857](#)
- [9] Johansson, K. (1998). On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, **91**, 151–204. [MR1487983](#)
- [10] Mingo J. A., Popa M. (2013). Real second order freeness and Haar orthogonal matrices. *J. Math. Phys.*, **54(5)**, 051701. [MR3098908](#)
- [11] Mingo, J. A., 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.
- [12] Mingo, J. A., Śniady, P. and Speicher, R. (2007). Second order freeness and fluctuations of random matrices: II. Unitary random matrices. *Adv. Math.*, **209(1)**, 212–240.
- [13] Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory. *Wiley, New York*.
- [14] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29**, 295–327.
- [15] Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30**, 1081–1102.
- [16] Lytova, A. and Pastur, L. (2009). Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.*, **37(5)**, 1778–1840.
- [17] Pearson, Karl (1900). On the criterion that a given system of deviations

- from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5* **50(302)**, 157–175.
- [18] Rama, C. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quant. Finance*, **1**, 223–236.
- [19] Rao, N. R., Mingo, J. A., Speicher, R. and Edelman, A. (2008). Statistical eigen-inference from large Wishart matrices. *Ann. Statist.*, 2850–2885.
- [20] Redelmeier, E. (2013). Real Second-Order Freeness and the Asymptotic Real Second-Order Freeness of Several Real Matrix Models. *Int. Math. Res. Notices*, **2014(12)**, 3353–3395. [MR3217665](#)
- [21] Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika*, **92**, 951–956.
- [22] Shcherbina, M. (2011). Central limit theorem for linear eigenvalue statistics of the Wigner and sample covariance random matrices. *Math. Phys. Anal. Geom.*, **7(2)**, 176–192.
- [23] Srivastava, M. S. (2005). Some Tests Concerning the Covariance Matrix in High Dimensional Data. *J. Japan Statist. Soc.*, **35(2)**, 251–272. [MR2328427](#)
- [24] Srivastava, M. S. and Reid, N. (2012). Testing the structure of the covariance matrix with fewer observations than the dimension. *J. Multivariate Anal.*, **112**, 156–171.
- [25] Sinai, Y. and Soshnikov, A. (1998). Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Bol. Soc. Brasil. Mat.*, **29(1)**, 1–24.
- [26] Voiculescu, D. (1991). Limit laws for random matrices and free products. *Invent. Math.*, **104(1)**, 201–220.
- [27] Wilks, S. S. (1935). On the independence of  $k$  sets of normally distributed statistical variables. *Econometrica*, **3**, 309–326.
- [28] Yang, Y. R. and Pan, G. M. (2015). Independence test for high dimensional data based on regularized canonical correlation coefficients. *Ann. Statist.*, **43**, 467–500.