SPECTRAL STATISTICS OF LARGE DIMENSIONAL
SPEARMAN’S RANK CORRELATION MATRIX AND ITS
APPLICATION

ZHIGANG BAO, LIANG-CHING LIN, GUANGMING PAN, AND WANG ZHOU

Abstract. Let $Q = (Q_1, \ldots, Q_n)$ be a random vector drawn from the uniform distribution on the set of all $n!$ permutations of $\{1, 2, \ldots, n\}$. Let $Z = (Z_1, \ldots, Z_n)$, where $Z_j$ is the mean zero variance one random variable obtained by centralizing and normalizing $Q_j$, $j = 1, \ldots, n$. Assume that $X_i, i = 1, \ldots, p$ are i.i.d. copies of $\frac{1}{\sqrt{p}}Z$ and $X = X_{p,n}$ is the $p \times n$ random matrix with $X_i$ as its $i$-th row. Then $S_n = XX^*$ is called the $p \times n$ Spearman’s rank correlation matrix which can be regarded as a high dimensional extension of the classical non-parametric statistic Spearman’s rank correlation coefficient between two independent random variables. In this paper we will establish a CLT for the linear spectral statistics of this non-parametric random matrix model in the scenario of high dimension supposing that $p = p(n)$ and $p/n \to c \in (0, \infty)$ as $n \to \infty$. We propose a novel evaluation scheme to estimate the core quantity in Anderson and Zeitouni’s cumulant method in [1] to bypass the so called joint cumulant summability. In addition, we raise a two-step comparison approach to obtain the explicit formulae for the mean and covariance functions in the CLT. Relying on this CLT we then construct a distribution-free statistic to test complete independence for components of random vectors. Owing to the non-parametric property, we can use this test on generally distributed random variables including the heavy-tailed ones.

1. Introduction

1.1. Matrix model. In this paper, we will consider the large dimensional Spearman’s rank correlation matrices. Firstly, we give the definition of the matrix model. Let $P_n$ be the set consisting of all permutations of $\{1, 2, \ldots, n\}$. Suppose that $Z = (Z_1, \ldots, Z_n)$ is a random vector, where

$$Z_i = \sqrt{\frac{12}{n^2 - 1}} (Q_i - \frac{n + 1}{2})$$

Date: January 28, 2014.

Key words and phrases. Spearman’s rank correlation matrix, non-parametric method, linear spectral statistics, central limit theorem, independence test.

Z.G. Bao was partially supported by the Ministry of Education, Singapore, under grant # ARC 14/11, and NSF of China, grant No. 11371317.; L.-C. Lin was partially supported by the Ministry of Education, Singapore, under grant # ARC 14/11; G.M. Pan was partially supported by the Ministry of Education, Singapore, under grant # ARC 14/11; W. Zhou was partially supported by the Ministry of Education, Singapore, under grant # ARC 14/11, and by a grant R-155-000-116-112 at the National University of Singapore.
and $Q := (Q_1, \ldots, Q_n)$ is uniformly distributed on $P_n$. That is, for any permutation $(\sigma(1), \sigma(2), \ldots, \sigma(n)) \in P_n$, one has $P\{Q = (\sigma(1), \sigma(2), \ldots, \sigma(n))\} = \frac{1}{n!}$. For simplicity, we will use the notation $[N] := \{1, \ldots, N\}$ for any positive integer $N$ in the sequel. Now for $m \in [n]$ we conventionally define the set of $m$ partial permutations of $n$ as

$$P_{nm} := \{(v_1, \ldots, v_m) : v_1, \ldots, v_m \in [n] \text{ and } v_i \neq v_j \text{ if } i \neq j\}.$$ 

For any mutually distinct numbers $l_1, \ldots, l_m \in [n]$, it is elementary to check that $(Q_{l_1}, \ldots, Q_{l_m})$ is uniformly distributed on $P_{nm}$. Such a fact immediately leads to the fact that $\{Z_i\}_{i=1}^n$ is strictly stationary. In addition, by setting $m = 1$ or $2$, it is straightforward to check through calculation that

$$E_{i}Z_i = 0, \quad E_{i}Z_i^2 = 1, \quad \text{Cov}(Z_j, Z_k) = -\frac{1}{n-1}, \quad \text{if } j \neq k.$$ (1.1)

Moreover, it is also easy to see that for any positive integer $l$,

$$E|Z_i|^l \leq C_l$$

for some positive constant $C_l$ depending on $l$. Besides, we note that $Z_i, i \in [n]$ are symmetric random variables.

Assuming that $X_i = (x_{i1}, \ldots, x_{in}), i = 1, \ldots, p$ are i.i.d copies of $\frac{1}{\sqrt{p}}Z$, we set $S_n = XX^*$, where

$$X := \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} = (x_{ij})_{p,n}.$$ 

Then $S_n$ will be called the Spearman’s rank correlation matrix.

1.2. **Motivation from non-parametric statistics.** A main motivation of considering the matrix $S_n$ is from non-parametric statistics. We consider the hypothesis testing problem on some random variable sequence $Y_1, \ldots, Y_p$ as

$\mathbf{H}_0 : \quad Y_1, \ldots, Y_p \text{ are independent }$ \quad v.s. \quad $\mathbf{H}_1 : \quad Y_1, \ldots, Y_p \text{ are not independent}.$

Note that since the covariance matrix and correlation matrix can capture dependence for Gaussian variables, it is natural to compare the sample covariance or correlation matrix with diagonal matrix to detect whether $\mathbf{H}_0$ holds in the classical setting of large $n$ and fixed $p$. While due to the so-called curse of dimensionality, it is now well known that there is no hope to approximate the population covariance or correlation matrix by sample ones in the situation of large $n$ and comparably large $p$ without any beforehand assumption imposed on the population covariance matrix. However, it is still possible to construct independence test statistics from the sample covariance or correlation matrix for Gaussian variables even in the high dimensional case. In the scenario of large $n$ and comparably large $p$, there is a long list of literature devoted to tackling the properties of sample covariance matrix or sample correlation matrix under the null hypothesis $\mathbf{H}_0$ in varying degrees. For example, Bao, Pan and Zhou [4] and Pillai and Yin [12] studied the largest eigenvalue of sample correlation matrix; Ledoit and Wolf...
[10] raised a quadratic form of the spectrum of sample covariance matrix; Schott [14] considered the sum of squares of sample correlation coefficients; Jiang [8] discussed the largest off-diagonal entry of sample correlation matrix; Jiang and Yang [9] studied the likelihood ratio test statistic for sample correlation matrix. However, for non-Gaussian variables, even in the classical large $n$ and fixed $p$ case, the idea to compare the population covariance matrix with diagonal matrix is substantially invalid for independence test for those uncorrelated but dependent variables. Moreover, for random vectors containing at least one heavy tailed component such as Cauchy distributed variable, there is even no population covariance matrix. For the high dimensional data, the philosophy of detecting dependence among variables via studying the statistics constructed from sample covariance or correlation matrix is also doubtful.

In order to tackle the independence hypothesis test for generally distributed variables, we discuss a non-parametric matrix model in this paper and study its spectrum statistics under $H_0$. Assume that we have taken $n$ observations of the vector $(Y_1, \ldots, Y_p)$. Let $Y_{11}, \ldots, Y_{1n}$ be the observations of the first coefficient $Y_1$ and set $Q_{1j}$ to be the rank of $Y_{1j}$ among $Y_{11}, \ldots, Y_{1n}$. We then replace $(Y_{11}, \ldots, Y_{1n})$ by the corresponding normalized rank sequence $(x_{11}, \ldots, x_{1n})$, where

$$x_{1j} = \sqrt{\frac{12}{p(n^2 - 1)}} (Q_{1j} - \frac{n + 1}{2}), \ j \in [n].$$

Analogously, we can define the rank sequence $(x_{i1}, \ldots, x_{in})$ for other $i \in [p]$. For simplicity, in this paper we only consider the case when $Y_i, i \in [p]$ are continuous random variables. In this case, the probability for a tie occurring in the sequence $Q_{i1}, \ldots, Q_{in}$ for any $i \in [p]$ is zero. Then $S_n = XX^*$ with $X = (x_{ij})_{p,n}$ is the so-called Spearman’s rank correlation matrix under $H_0$, which can be regarded as a high dimensional extension of the classical Spearman’s rank correlation coefficient between two random variables. Then we can construct statistics from $S_n$ to tackle the above hypothesis testing problem. By contrast, the parametric models such as Pearson’s sample correlation matrix and sample covariance matrix are well studied by statisticians and probabilists. However, the work on Spearman’s rank correlation matrix is few and far between. In [3], Bai and Zhou proved that the limiting spectral distribution of $S_n$ is also the famous Marchenko-Pastur law (MP law). In [20], Zhou studied the limiting behavior of the largest off-diagonal entry of $S_n$. Our purpose in this paper is to derive the fluctuation (a CLT) of the linear spectral statistics for $S_n$. As an application, we will construct a non-parametric statistic to detect dependence of components of random vectors.

1.3. Main result. We set $\lambda_1 \geq \ldots \geq \lambda_p$ to be ordered eigenvalues of $S_n$. Our main task in this paper is to study the limiting behavior of the so-called linear spectral statistics

$$\mathcal{L}_n[f] = \sum_{i=1}^{p} f(\lambda_i)$$
for some test function $f$. In this paper, we will focus on the polynomial test functions, thus it suffices to study the joint limiting behavior of $\text{tr}S^k_n$, $k = 1, \ldots, \infty$. We state the main result as the following theorem.

**Theorem 1.1.** Assuming that both $n$ and $p := p(n)$ tend to $\infty$ and

$$n/p \to c \in (0, \infty),$$

we have

$$\left\{ \text{tr}S^k_n - \mathbb{E}\text{tr}S^k_n \right\}_{k=2}^{\infty} \Rightarrow \{G_k\}_{k=2}^{\infty}, \quad \text{as } n \to \infty,$$

where $\{G_k\}_{k=2}^{\infty}$ is a Gaussian process with mean zero and covariance function given by

$$\text{Cov}(G_{k_1}, G_{k_2}) = 2c^{k_1+k_2} \sum_{j_1=0}^{k_1-1} \sum_{j_2=0}^{k_2} \binom{k_1}{j_1} \binom{k_2}{j_2} \left( \frac{1-c}{c} \right)^{j_1+j_2}
\times \sum_{l=1}^{k_1-j_1} \left( \frac{2k_1 - 1 - (j_1 + l)}{k_1 - 1} \right) \left( \frac{2k_2 - 1 - j_2 + l}{k_2 - 1} \right)
- 2c^{k_1+k_2+1} \sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} \binom{k_1}{j_1} \binom{k_2}{j_2} \left( \frac{1-c}{c} \right)^{j_1+j_2}
\times \left( \frac{2k_1 - 1}{k_1 - 1} \right) \left( \frac{2k_2 - 1}{k_2 - 1} \right)\right)_{1.2}$$

Moreover, we have the following expansion for the expectation function,

$$\mathbb{E}\text{tr}S^k_n = \frac{n^k}{(n-1)^{k-1}} \sum_{j=0}^{k-1} \frac{1}{j+1} \binom{k}{j} \binom{k-1}{j} \left( \frac{n-1}{p} \right)^j \left( \frac{k}{2} \right)^2 c^j
+ 2c^{1+k} \sum_{j=0}^{k} \binom{k}{j} \left( \frac{1-c}{c} \right)^j \left[ \left( \frac{2k-j}{k-1} \right) - \left( \frac{2k+1-j}{k-1} \right) \right]
+ \frac{1}{4} \left[ (1-\sqrt{c})^{2k} + (1+\sqrt{c})^{2k} \right] + o(1). \quad (1.3)$$

**Remark 1.2.** Note that when $k = 1$, $\text{tr}S_n = n$ which is deterministic. Actually, it can also be checked that the r.h.s. of (1.2) is zero when $k_1 = k_2 = 1$ and the r.h.s. of (1.3) equals $n + o(1)$ when $k = 1$ therein.

1.4. **Methodologies of the proof.** Roughly speaking, we will start with a *cumulant method* raised by Anderson and Zeitouni in [1] to establish the CLT for $\{(\text{tr}S^k_n - \mathbb{E}\text{tr}S^k_n)/\sqrt{\text{Var}(\text{tr}S^k_n)}\}_{k=2}^{\infty}$. In this step, a novel evaluation scheme will be introduced to bypass the main obstacle caused by the absence of the so-called *joint cumulant summability* property imposed in [1]. And for the explicit expressions of mean and covariance functions, we will put forward a *two-step comparison strategy* which allows us to get (1.2) and (1.3) from existing results for large dimensional sample covariance matrices. In the sequel, we will moderately unfold the above strategies.
The so-called cumulant method can be viewed as a modification of the famous moment method which is a celebrated approach in the Random Matrix Theory (RMT). Especially in some weak convergence problems of spectral statistics in RMT, the moment method plays an important role. The main idea of moment method is to verify the limiting properties of a spectral statistic via evaluating the limits of its moments. One main reason for adopting moment method in RMT is that the estimation of moments of spectral statistics of some canonical random matrices can usually be transformed into some friendly combinatoric enumeration problems. Such a powerful strategy can trace back to the groundbreaking work of Wigner [17]. Without trying to be comprehensive we refer to [2, 5, 15, 17] for further reading. Crudely speaking, the so-called cumulant method can be described via replacing moments by cumulants in the above issue. Especially, when one wants to establish a CLT for some random sequence, cumulant method may be more effective. As is well known, except for the cumulants of order 1 and 2, all higher order cumulants vanish for Gaussian distribution. Therefore, if one wants to prove CLT for a random sequence, it suffices to show all higher order cumulants sequence tends to 0 and evaluate the limits of the cumulants of first two orders. Particularly, one can refer to [1, 7, 16] for cumulant method in proving CLTs for linear spectral statistics in RMT.

As mentioned above, we will start with the cumulant method in [1], where the authors established a CLT for the linear spectral statistics of the banded sample covariance matrices for stationary sequences. As explained at the beginning, \( \{Z_i\}_{i=1}^\infty \) defined above is also a stationary sequence, which inspires us to learn from [1]. However, in [1] the stationary sequence is required to satisfy the so-called joint cumulant summability property (see (2.6) below), which was used to bound the core quantity (2.5) below. The property of joint cumulant summability is crucial in the spectral analysis of time series, one can refer to [6, 13, 18, 19] for further reading. Unfortunately, to verify the property of joint cumulant summability for a general stationary sequence is highly nontrivial. As far as we know, only some sequences with particular independence structure have been discussed, see the aforementioned references for instance. In this paper, we will not try to check whether joint cumulant summability holds for \( Z \). Instead we will provide a relatively rough but crucial bound on (2.5) below through a totally novel evaluation scheme, see Proposition 3.2 and Corollary 3.4 below. Such bound will allow us to bypass the property of joint cumulant summability and serve as a main input to pursue the cumulant method to establish the CLT for \( \{ (\text{tr} S_n^k - \text{Etr} S_n^k) / \sqrt{\text{Var}(\text{tr} S_n^k)} \}_{k=2}^\infty \). Provided the above CLT, what remains is therefore to evaluate the non-negligible terms of mean and covariance functions. It will be shown that the mean and covariance functions of \( \{ \text{tr} S_n^k \}_{k=2}^\infty \) can be expressed by some sums with terms indexed by set partitions. For the explicit values of the non-negligible terms of mean and covariance functions, instead of bare-handed combinatoric enumeration, we will adopt a two-step comparison strategy to compare these expressions with the existing results for sample covariance matrices. More specifically, it will be shown that
the main term of either mean or covariance function is some sum over two types of set partitions. The partitions in the first type consist of 2-element blocks only and the ones in the second type contain one 4-element block. For the sum over the partitions in the first type, we compare it with that of the sample covariance matrix with i.i.d. standard Gaussian entries (i.e. Wishart matrix). And then for the sum over the partitions in the second type we compare it with that of sample covariance matrix with more general i.i.d. mean zero variance one entries whose common kurtosis does not vanish (i.e. fourth moment is not 3). Such a two-step comparison strategy also provides us an intuitive way to see how the mean and covariance functions depend on the second and fourth moments of the underlying matrix entries.

1.5. Organization and notation. Our paper is organized as follows. In Section 2, we will introduce some basic notions of joint cumulants and some known results from [1]. Section 3 is our main technical part which will be devoted to providing the required bound for the sum (2.5). Specifically, we will mainly prove Proposition 3.2 therein. In Section 4, we will use the bounds obtained in Section 3 to show that all high order cumulants tend to zero when \( n \to \infty \). And in Section 5, we will combine the bounds in Section 3 and the aforementioned two-step comparison strategy to evaluate the main terms of the mean and covariance functions. Section 6 is devoted to the application of our CLT on independence test on random vectors.

Throughout the paper, we will use \( \#S \) to represent the cardinality of a set \( S \). For any number set \( A \subset [n] \), we will use \( \{j_\alpha\}_{\alpha \in A} \) to denote the set of \( j_\alpha \) with \( \alpha \in A \). Analogously, \( (j_\alpha)_{\alpha \in A} \) means the vector obtained by deleting the components \( j_\beta \) with \( \beta \in [n] \setminus A \) from \( (j_1, \ldots, j_n) \). In addition, we will use the notation \( i = \sqrt{-1} \) to denote the imaginary unit to release \( i \) which will be frequently used as subscript or index. For any vector \( \xi = (\xi_1, \ldots, \xi_N) \), we say the position of \( \xi_i \) in \( \xi \) is \( i \). For example, for vector \( (Z_{j_1}, Z_{j_2}, Z_{j_3}, Z_{j_4}) \), the position of \( Z_{j_1} \) is 2. Moreover, we will use \( C \) to represent some positive constant independent of \( n \) whose value may be different from line to line.

2. Preliminaries and tools from Anderson and Zeitouni [1]

In this section, we will introduce some basic notions concerning cumulants and some necessary results from [1]. For any positive integer \( N \) and random variables \( \xi_1, \ldots, \xi_N \), the joint cumulant \( C(\xi_1, \ldots, \xi_N) \) is defined by

\[
C(\xi_1, \ldots, \xi_N) = i^{-N} \frac{\partial^N}{\partial x_1 \cdots \partial x_N} \log \mathbb{E} \exp\left(\sum_{j=1}^{N} ix_j \xi_j\right)|_{x_1=\ldots=x_N=0}. \tag{2.1}
\]

Firstly, it is straightforward to check via above definition that the following properties hold.

\[ \text{P1 : } C(\xi_1, \ldots, \xi_N) \text{ is a symmetric function of } \xi_1, \ldots, \xi_N. \]

\[ \text{P2 : } C(\xi_1, \ldots, \xi_N) \text{ is a multilinear function of } \xi_1, \ldots, \xi_N. \]
P3 : \( C(\xi_1, \ldots, \xi_N) = 0 \) if the variables \( \xi_i, i \in [N] \) can be split into two groups \( \{\xi_i\}_{i \in S_1} \) and \( \{\xi_i\}_{i \in S_2} \) with \( S_1 \cap S_2 = \emptyset \) and \( S_1 \cup S_2 = [N] \) such that the sigma field \( \sigma\{\xi_i\}_{i \in S_1} \) is independent of \( \sigma\{\xi_i\}_{i \in S_2} \).

It is well known that the joint cumulant can be expressed in terms of moments. For any positive number \( N \), let \( L_N \) be the lattice consisting of all the partitions of \([N]\). We say \( \pi = \{B_1, \ldots, B_m\} \in L_N \) is a partition of the set \([N]\) if

\[
\emptyset \neq B_i \subset [N], \quad i = 1, \ldots, m, \quad \bigcup_{i=1}^m B_i = [N], \quad B_i \cap B_j = \emptyset, \quad \text{if } i \neq j.
\]

We say \( B_i \)'s are blocks of \( \pi \) and \( m \) is the cardinality of \( \pi \). We will also conventionally use the notation \( \#\pi \) to denote the cardinality of a partition \( \pi \) all the way.

\( L_N \) is a poset (partially ordered set) ordered by set inclusion. Specifically, given two partitions \( \pi \) and \( \sigma \) in \( L_N \), we say \( \pi \leq \sigma \) (or \( \pi \) is a refinement of \( \sigma \)) if every block of \( \pi \) is contained in a block of \( \sigma \). Now given two partitions \( \sigma_1, \sigma_2 \in L_N \), with the above order \( \leq \) we define \( \sigma_1 \vee \sigma_2 \) to be the least upper bound of \( \sigma_1 \) and \( \sigma_2 \). For example, let \( N = 8 \) and

\[
\sigma_1 = \{\{1, 2\}, \{3, 4, 5\}, \{6\}, \{7, 8\}\}, \quad \sigma_2 = \{\{1, 3\}, \{2, 5\}, \{4\}, \{6, 8\}, \{7\}\}.
\]

Then we have \( \sigma_1 \vee \sigma_2 = \{\{1, 2, 3, 4, 5\}, \{6, 7, 8\}\} \).

With the aid of these notations, we have the following basic expression of joint cumulant in terms of moments.

\[
C(\xi_1, \ldots, \xi_N) = \sum_{\pi \in L_N} (-1)^{\#\pi - 1} (\#\pi - 1)! \mathbb{E}_\pi(\xi_1, \ldots, \xi_N), \tag{2.2}
\]

where

\[
\mathbb{E}_\pi(\xi_1, \ldots, \xi_N) = \prod_{A \in \pi} \mathbb{E} \prod_{i \in A} \xi_i^*.
\]

Especially, one has

\[
C(\xi) = \mathbb{E}_{\xi}, \quad C(\xi_1, \xi_2) = \text{Cov}(\xi_1, \xi_2).
\]

With the aid of the above concepts, we can now introduce the formula of joint cumulants of \( TrS_n^{k_l}, l = 1, \ldots, r \) with \( \sum_{l=1}^r k_l = k \) derived in [1]. To this end, we need to specify two partitions \( \pi_0, \pi_1 \in L_{2k} \) as

\[
\pi_0 = \{\{1, 2\}, \{3, 4\}, \ldots, \{2k - 1, 2k\}\},
\pi_1 = \{\{2, 3\}, \ldots, \{k_1, 1\}, \{k_1 + 2, k_1 + 3\}, \ldots,
\{k_2, k_1 + 1\}, \ldots, \{k_r - 1 + 2, k_r - 1 + 3\}, \ldots, \{k_r, k_r - 1 + 1\}\},
\]

where \( k_i = 2 \sum_{j=1}^i k_j, i = 1, \ldots, r \) and \( k_j \)'s are nonnegative integers. Observe that \( \#\pi_0 \vee \pi_1 = r \).

Now let \( L_{2k}^+ \subset L_N \) be the set consisting of those partitions in which each block has cardinality larger than 2. If each block of a partition

\[\text{Here we remind that the partition } \pi \text{ in the notation } \mathbb{E}_\pi(\cdot) \text{ takes effect on the positions of the components of the vector, so does the notation } C_\pi(\cdot) \text{ in (2.3). The reader should not confuse the positions with the subscripts of the components of the random vector. For example, for } \pi = \{\{1, 2\}, \{3, 4\}\}, \text{ we have } \mathbb{E}_\pi(Z_{j_2}, Z_{j_3}, Z_{j_4}, Z_{j_5}) = \mathbb{E}(Z_{j_2}Z_{j_3})\mathbb{E}(Z_{j_4}Z_{j_5}).\]
has cardinality of 2, we call such a partition a perfect matching. For even \( N \), we let \( L_N^2(N) \subset L_N^{2+} \) be the set consisting of all perfect matchings. In the sequel, we will also use the notation \( L_{2k}^4 \) to denote the set consisting of the partitions of \([2k]\) containing one 4-element block and \((k - 2)\) 2-element blocks. We use the terminology in [1] to call the index vector \( \mathbf{j} := (j_1, \ldots, j_N) \) with \( j_\alpha \in [n], \alpha \in [N] \) as a \((n,N)\)-word. Moreover, we say an \((n,N)\)-word \( \mathbf{j} \) is \( \pi \)-measurable for some partition \( \pi \in L_N^2 \) when \( j_\alpha = j_\beta \) if \( \alpha, \beta \) are in the same block of \( \pi \). Then by the discussions in [1] (see Proposition 5.2 therein), one has

\[
C(\text{tr} S_{k_1}^{k_1}, \ldots, \text{tr} S_{k_r}^{k_r}) = \sum_{\pi \in L_{2k}^4, \# \pi_0 \vee \pi_1 \vee \pi = 1} \sum_{\mathbf{j} \text{: (n,2k)-word}} C_{\pi}(\mathbf{j}), \tag{2.3}
\]

where

\[
C_{\pi}(\mathbf{j}) := C_{\pi}(Z_{j_1}, \ldots, Z_{j_{2k}}) = \prod_{A \in \pi} C\{Z_{j_\alpha}\}_{\alpha \in A}.
\]

(2.3) was derived by using Möbius inversion formula in [1]. We remind that here we switch the roles of the parameters \( p \) and \( n \) under the setting in [1]. Moreover, \( B(\mathbf{j}) \) therein is always 1 for all \((n,2k)\)-words \( \mathbf{j} \) in our case.

Our aim is to show that for \( k_1, \ldots, k_r \) with \( \sum_{j=1}^r k_j = k \),

\[
C(\text{tr} S_{n}^{k_1}, \ldots, \text{tr} S_{n}^{k_r}) = \begin{cases} 
  o(1), & \text{if } r \geq 3, \\
  \text{r.h.s. of (1.2) } + o(1), & \text{if } r = 2, \\
  \text{r.h.s. of (1.3)}, & \text{if } r = 1.
\end{cases} \tag{2.4}
\]

It is well known that (2.4) can imply Theorem 1.1 directly.

Apparently, by (2.3) we see that to prove (2.4), the main task is to estimate the summation

\[
\sum_{\mathbf{j} \text{: (n,2k)-word}} C_{\pi}(\mathbf{j}) \tag{2.5}
\]

for various \( \pi \). In order to deal with the estimation in the counterparts of [1], the main assumption imposed on the stationary sequence \( \{Y_k\}_{k=-\infty}^\infty \) considered therein is the so called joint cumulant summability

\[
\sum_{j_1} \cdots \sum_{j_r} |C(Y_0, Y_{j_1}, \ldots, Y_{j_r})| = O(1), \quad \text{for all } r \geq 1. \tag{2.6}
\]

Actually, once joint cumulant summability held for random sequence \( \{Z_i\}_{i=1}^\infty \), one could obtain that the summation (2.5) was bounded by \( O(n^{\# \pi_0 \vee \pi}) \) in magnitude. However, the verification of the joint cumulant summability for a general random sequence is quite nontrivial. Without joint cumulant summability, we will provide a weaker bound on (2.5) below directly, see Proposition 3.2 and Corollary 3.4. Such a weaker bound will still make our proof strategy amenable.
proposition proved by Anderson and Zeitouni in [1] will be crucial to ensure that our weaker bound on (2.5) still works well in the proof of (2.4).

Proposition 2.1 (Proposition 3.1, [1]). Let $\Pi \in L_{2k}^{2+}$ and $\Pi_0, \Pi_1 \in L_{2k}^2$ for some positive integer $k$. If $\#\Pi \lor \Pi_0 \lor \Pi_1 = 1$ and $\#\Pi_0 \lor \Pi_1 = r$ for some positive integer $r \leq k$, one has

$$\#\Pi_0 \lor \Pi + \#\Pi_1 \lor \Pi \leq (\#\Pi + 1)1_{\{r=1\}} + \min\{\#\Pi + 1, k + 1 - r/2\}1_{\{r \geq 2\}}. \dagger$$

At the end of this section, we state some elementary properties of the vector $Z$ which will be used in the subsequent sections. We summarize them as the following lemma whose proof will be stated in the Appendix.

Lemma 2.2. Let $Z = (Z_1, \ldots, Z_n)$ be the random vector defined above. Let $m$ and $\alpha_i, i = 1, \ldots, m$ be fixed positive integers. We have the following properties of $Z$.

(i): If $\sum_{i=1}^{m} \alpha_i$ is odd,

$$E(Z_1^{\alpha_1} \cdots Z_m^{\alpha_m}) = 0.$$  \hspace{1cm} (2.7)

(ii): If $\sum_{i=1}^{m} \alpha_i$ is even,

$$E(Z_1^{\alpha_1} \cdots Z_m^{\alpha_m}) = O(n^{-\frac{n_0(\alpha)}{2}}),$$  \hspace{1cm} (2.8)

where $\alpha = (\alpha_1, \ldots, \alpha_m)$ and $n_0(\alpha) = \#\{i \in [m]: \alpha_i \text{ is odd}\}$.

Moreover, by symmetry, (i) and (ii) still hold if we replace $(Z_1, \ldots, Z_m)$ by $(Z_{l_1}, \ldots, Z_{l_m})$ with any mutually distinct indices $l_1, \ldots, l_m \in [n]$.

Note that (2.2) together with (i) of Lemma 2.2 implies that

$$C(Z_{j_1}, \ldots, Z_{j_r}) = 0, \quad \text{if } r \text{ is odd}. \hspace{1cm} (2.9)$$

Consequently, it suffices to consider partition $\pi$ in which each block has even cardinality. We denote $L_{2k}^{\text{even}} \subset L_{2k}^{2+}$ to be the set of such partitions. Then we can rewrite (2.3) as

$$C(\text{tr}S_{k_1}^{\pi_{k_1}}, \ldots, \text{tr}S_{k_1}^{\pi_{k_r}}) = \sum_{\pi \in L_{2k}^{\text{even}}} p^{-k + \#\pi_0 \lor \pi} \sum_{\substack{j:(n,2k)\text{-word} \\text{s.t. } j \text{ is } \pi_1 \text{ measurable}}} C_\pi(j). \hspace{1cm} (2.10)$$

\dagger One might note that in Proposition 3.1 of [1], the authors stated a slightly weaker bound $k + 1 - \lfloor r/2 \rfloor$ in formula (14) therein. However, by the proof of Proposition 3.1 in [1], it is not difficult to see that one can improve it to be $k + 1 - r/2$. We appreciate Prof. Greg W. Anderson’s confirmation on this.
3. A rough bound on the summation (2.5)

3.1. Factorization of the summation. As mentioned in the last section, instead of checking the property of joint cumulant summability \(^{1}\), we will try to provide a rough bound on the summation (2.5) for any given \(\pi \in L_{2k}^{\text{even}}\) directly. Now we observe by definition that

\[ C_{\pi}(j) = \prod_{A \in \pi} C_{\{Z_{j_{\alpha}}\}_{\alpha \in A}} = \prod_{B \in \pi_{1} \cup \pi} \prod_{A \in \pi, s.t. A \subseteq B} C_{\{Z_{j_{\alpha}}\}_{\alpha \in A}}. \]

For simplicity, we denote \(q := \#\pi_{1} \cup \pi\) and index the blocks in \#\pi_{1} \cup \pi\) in any fixed order as \(B_{i}, i \in [q]\) for convenience. We set

\[ b_{i} := \frac{\#B_{i}}{2}, \quad m_{i} := \#\{A \in \pi : A \subseteq B_{i}\}, \quad i \in [q], \]

and index the subsets of \(B_{i}\) in \(\pi\) in any fixed order by \(A_{i}^{(\beta)}, \beta \in [m_{i}]\). Moreover, we define

\[ a_{i}(\beta) := \frac{\#A_{i}^{(\beta)}}{2}. \]

Then we can write

\[ C_{\pi}(j) = \prod_{i=1}^{q} \prod_{\beta=1}^{m_{i}} C_{\{Z_{j_{\alpha}}\}_{\alpha \in A_{i}^{(\beta)}}}. \] (3.1)

To simplify the notation, we use \(J := J(\pi_{1}, n, k)\) to represent the set consisting of all \(\pi_{1}\)-measurable \((n, 2k)\)-word \(j\). Fix \(\alpha \in [2k]\), we can view \(\alpha : J \rightarrow [n]\) defined by \(\alpha(j) := j_{\alpha}\) as a functional on \(J\). Then it is apparent that if \(j_{\alpha} \equiv j_{\beta}\) (i.e. \(\alpha(j) \equiv \beta(j)\)) on \(J\) for fixed \(\alpha, \beta \in [2k]\), one has \(\alpha\) and \(\beta\) are in the same block of \(\pi_{1}\). We have the following lemma on the properties of the triple \((J, \pi, \pi_{1})\).

**Lemma 3.1.** Regarding \(\{j_{\alpha}\}_{\alpha \in [2k]}\) as a class of \(2k\) functionals on \(J\), we have

(i): Given \(i \in [q]\), for any \(\alpha \in B_{i}\), there exists exactly one \(\gamma \in B_{i}\) such that \(\alpha \neq \gamma\) but \(j_{\alpha} \equiv j_{\gamma}\) on \(J\).

(ii): Given \(i \in [q]\) with \(m_{i} \geq 2\), for any proper subset \(P \subseteq [m_{i}]\), there exists at least one \(\alpha_{1} \in \bigcup_{\beta \in PA_{i}^{(\beta)}}\), for any other \(\alpha_{2} \in \bigcup_{\beta \in PA_{i}^{(\beta)}}\) one has \(j_{\alpha_{1}} \neq j_{\alpha_{2}}\) on \(J\).

**Proof.** Note that \(B_{i} \in \pi_{1} \cup \pi\), thus \(B_{i}\) is the union of several blocks in \(\pi_{1}\). Moreover, \(\pi_{1}\) is a perfect matching thus the components of \(j \in J\) with positions in any union of blocks of \(\pi_{1}\) appear in pairs. Each block is corresponding to one pair. Then (i) follows immediately. Now we show (ii). If for some proper subset \(P \subseteq [m_{i}]\) such that for any \(\alpha \in \bigcup_{\beta \in PA_{i}^{(\beta)}}\) there exists exactly one (by (i)) \(\gamma \in \bigcup_{\beta \in PA_{i}^{(\beta)}}\) such

\(^{1}\)As mentioned above, once joint cumulant summability held, the magnitude of (2.5) could be bounded by \(O(n^{\#\pi_{1} \cup \pi})\). To see this, one can refer to Proposition 6.1 of [1] by setting \(b = p\) therein and switching the role of \(p\) by \(n\) to adapt to our notation.
that \( \alpha \neq \gamma \) but \( j_\alpha \equiv j_\gamma \), then \( \bigcup_{\beta \in P} A_i^{(\beta)} \) itself forms a single block of \( \pi_1 \lor \pi \). It contradicts the fact that \( P \) is a proper subset of \( [m_i] \). \( \square \)

For convenience, we use the notation

\[
|j|_B = (j_\alpha)_{\alpha \in B}
\]

(3.2)

for any \( B \subset [2k] \). Analogously, for any partition \( \sigma \) and \( D \subset [2k] \) we will use the notation

\[
|\sigma|_D = \{ E \in \sigma : E \subset D \},
\]

which can be viewed as the partition \( \sigma \) restricted on the set \( D \). Actually, Lemma 3.1 is just a direct consequence of the fact that \( \pi_1|B_i \) is a perfect matching and

\[
\#\pi|B_i \lor \pi_1|B_i = \#(\pi \lor \pi_1)|B_i = 1, \quad i \in [q].
\]

Analogous to the notation \( \mathcal{J}(\pi_1|B_i, n, b_i) \), we denote \( \mathcal{J}(\pi_1|B_i, n, b_i) \) to be the set consisting of all \( \pi_1|B_i \)-measurable words \( |j|_B \). By (3.1) and (i) of Lemma 3.1, we see that for given \( \pi \)

\[
(2.5) := \sum_{j \in \mathcal{J}} C_\pi(j) = \prod_{i=1}^{q} \sum_{j|B_i \in \mathcal{J}(\pi_1|B_i, n, b_i)} C_{\pi|B_i}(j|B_i)
\]

(3.3)

Hence to estimate (2.5), it suffices to provide a bound on the quantity

\[
\sum_{|j|_B \in \mathcal{J}(\pi_1|B_i, n, b_i)} C_{\pi|B_i}(j|B_i) = \sum_{|j|_B \in \mathcal{J}(\pi_1|B_i, n, b_i)} \prod_{\beta=1}^{m_i} C\{Z_{j_\alpha}\}_{\alpha \in A_i^{(\beta)}}
\]

for all fixed \( i \in [q] \). For simplicity, we discard the subscript \( i \) in the discussion below. Thus we will use the notation \( B, A_i^{(\beta)}, a_i(\beta), b, m \) to replace \( B_i, A_i^{(\beta)}, a_i(\beta), b_i, m_i \) temporarily. Our main technical result is the following crucial proposition.

**Proposition 3.2.** Under the above notation, we have

\[
\sum_{|j|_B \in \mathcal{J}(\pi_1|B, n, b)} \prod_{\beta=1}^{m} C\{Z_{j_\alpha}\}_{\alpha \in A_i^{(\beta)}} = O\left(n^{1+\sum_{\beta=1}^{m}(a_i(\beta)-2)1_{\{a_i(\beta)\geq 2\}}} \right).
\]

(3.4)

**Remark 3.3.** Here we draw the attention that by (i) of Lemma 3.1, the indices in \( |j|_B \) are equivalent in pairs on \( \mathcal{J}(\pi_1|B, n, b) \). Under this constraint the number of free \( j \) indices in the summation on the l.h.s. of (3.4) is actually \( b = \sum_{i=1}^{m} a_i(\beta) \).

Now we use \( n_\gamma(\pi) \) to represent the number of blocks in \( \pi \) whose cardinalities are \( 2\gamma \). We have the following corollary from Proposition 3.2.

**Corollary 3.4.** Under the above notation, we have

\[
\sum_{j \in \mathcal{J}} |C_\pi(j)| = O(\gamma \#\pi_1 \lor \pi + \sum_{\gamma \geq 2}(\gamma - 2)n_\gamma(\pi)).
\]

(3.5)

**Proof.** It follows from Proposition 3.2 and (3.3) directly. \( \square \)
Our main task in this section is to prove Proposition 3.2. The tedious proof
will be given in the remaining part of this section which will be further split into
several subsections. In Sections 3.2-3.4, we will provide some preliminary results
for our final evaluation scheme. The formal proof of Proposition 3.2 will be stated
in Section 3.5.

3.2. Bounds on joint cumulants. In this subsection, we will provide some
bounds on single joint cumulants with variables from $\mathbf{Z}$. Such bounds will help
us to reduce all these joint cumulants to some products of 2-element cumulants
which are more friendly for the subsequent combinatoric enumeration. Let $s, t$
be fixed nonnegative integers. Now for $l_1, \ldots, l_s, h_1, \ldots, h_{2t} \in [n]$, we denote the
vectors

$l := (l_1, l_1, l_2, \ldots, l_s, l_s), \quad h := (h_1, \ldots, h_{2t})$

and we will use $l$ index (resp. $h$ index) to refer to $l_i, i \in [s]$ (resp. $h_i, i \in [2t]$).

For simplicity, we will employ the notation

$\hat{l}h = (l_1, l_1, l_2, l_2, \ldots, l_s, l_s, h_1, \ldots, h_1, \ldots, h_{2t})$

which is the concatenation of $l$ and $h$. Note that the $l$ indices appear in pairs.

And in the sequel, we will use the notation

$C(\hat{l}h) := C(Z_{l_1}, Z_{l_1}, \ldots, Z_{l_s}, Z_{l_s}, Z_{h_1}, \ldots, Z_{h_{2t}})$

and

$C(h) := C(Z_{h_1}, \ldots, Z_{h_{2t}})$

to highlight the index sequence. In this manner, for any partition $\tilde{\pi} \in L_{2s+2t}$, we
denote

$E_{\tilde{\pi}}(\hat{l}h) := E_{\tilde{\pi}}(Z_{l_1}, Z_{l_1}, \ldots, Z_{l_s}, Z_{l_s}, Z_{h_1}, \ldots, Z_{h_{2t}})$

Analogously, for any partition $\tilde{\sigma} \in L_{2t}$, we set

$E_{\tilde{\sigma}}(h) := E_{\tilde{\sigma}}(Z_{h_1}, Z_{h_2}, \ldots, Z_{h_{2t}})$.

We remind here the aforementioned convention that the partition $\tilde{\pi}$ in the notation
$E_{\tilde{\pi}}(\cdot)$ takes effect on the positions of components of $\hat{l}h$.

To prove Proposition 3.2, we will need the following two lemmas. Lemma 3.5
gives us an explicit order on the magnitude of 4-element cumulant. And Lemma
3.6 provides some rough bounds on the cumulants with more than 4 elements.

Lemma 3.5. Suppose that $h = (h_1, h_2, h_3, h_4)$. Let $d(h)$ be the number of the
distinct values in $\{h_1, h_2, h_3, h_4\}$. We have

$C(h) = O(n^{-d(h)+1}).$ \hfill (3.7)

The proof of Lemma 3.5 will be stated in the Appendix. From Lemma 3.5, we
can get the following consequences. We see that if there exists a perfect matching
$\sigma = \{A_1, A_2\}$ of $\{1, 2, 3, 4\}$ such that $\{h_i\}_{i \in A_1} \cap \{h_i\}_{i \in A_2} = \emptyset$, then by using
Lemma 3.5 and (1.1) we can get

$|C(h)| \leq O(n^{-1})|E_{\sigma}(h)|$. \hfill (3.8)
If there is no such perfect matching, we have
\[ |C(h)| \leq C|E_\sigma(h)| \]
for any perfect matching \( \sigma \in L^2_4 \) with some positive constant \( C \). Note that the second case occurs if and only if three or four of the indices \( h_1, h_2, h_3, h_4 \) take the same value.

The following Lemma 3.6 provides some crucial bounds on the cumulants with more than 4 underlying elements. The proof of this lemma will also be stated in the Appendix.

**Lemma 3.6.** Under the above notation, we have the following bounds on the joint cumulant \( C(\hat{l}h) \).

(i): (Crude bound) When \( s + t \geq 3 \) and \( t \geq 1 \),
\[ |C(\hat{l}h)| \leq C \sum_{\sigma \in L^2_s} |E_\sigma(h)| \tag{3.10} \]
for some positive constant \( C \).

(ii): When \( s \geq 2, t = 1, l_1, \ldots, l_s \) are mutually distinct and distinct from \( h_1, h_2 \), we have
\[ |C(\hat{l}h)| \leq O(n^{-2}). \tag{3.11} \]

(iii): When \( s \geq 2, t = 1, l_1 = l_2 \) and \( l_2, \ldots, l_s \) are mutually distinct and distinct from \( h_1, h_2 \), we have
\[ |C(\hat{l}h)| = O(n^{-1})|E(Z_{h_1}Z_{h_2})|. \tag{3.12} \]

At the end of this subsection, we need to clear up a potential confusion which may occur when we use Lemma 3.6 in the sequel. Given an index sequence, for example \( (1, 1, 2, 2, 3, 3, 4, 4) \), we consider to use Lemma 3.6 to bound the corresponding cumulant \( C(Z_{l_1}, Z_{l_1}, \ldots, Z_{l_s}, Z_{l_s}) \). Obviously, we can adopt (ii) of Lemma 3.6 by setting \( l_1 = 1, l_2 = 2, l_3 = 3 \) and \( h_1 = h_2 = 4 \). Thus \( s = 3 \) and \( t = 1 \). However, we can also say that \( s = t = 2 \) such that \( l_1 = 1, l_2 = 2 \) while \( h_1 = h_2 = 3 \) and \( h_3 = h_4 = 4 \). We can even say that \( s = 0, t = 4 \) such that \( h_{2i-1} = h_{2i} = i, i = 1, \ldots, 4 \). That means the determination of \( l \) and \( h \) indices as well as \( s \) and \( t \) are not substantially important. Actually in any viewpoint listed above, we can employ (ii) of Lemma 3.6. We state Lemma 3.6 with \( l \) and \( h \) in this way in order to simplify the presentation. However, when we use Lemma 3.6, we only need to check which case of (i)-(iii) is applicable to the given index sequence. Moreover, we can also represent the bounds for (ii) and (iii) in the form of the r.h.s. of (3.10). In case (ii), obviously, we can find a perfect matching \( \tilde{\sigma} \in L^2_{2s+2} \) such that
\[ C_{\tilde{\sigma}}(\hat{l}h) = \prod_{i=1}^{s} C(Z_{l_i}, Z_{l_i}) \cdot C(Z_{h_1}, Z_{h_2}) = E(Z_{h_1}, Z_{h_2}). \]
By (1.1) and (3.11) we observe that in case (ii),
\[
| \mathbf{C}(\mathbf{\hat{h}}) | \leq O(n^{-1}) | \mathbb{E}(Z_{h_1}, Z_{h_2}) | = O(n^{-1}) | \mathbb{E}_\sigma(\mathbf{\hat{h}}) | \\
\leq O(n^{-1}) \sum_{\sigma \in L_2^{2h+2}} | \mathbb{E}_\sigma(\mathbf{\hat{h}}) |.
\]
(3.13)

Note that the above bound is not as strong as (3.11) when \( h_1 = h_2 \). Analogously, it is easy to check that (3.13) also holds in the case (iii) of Lemma 3.6.

3.3. Cyclic product of 2-element cumulants. In this subsection, we introduce the concept of cyclic product of 2-element cumulant (cycle in short) and the summation of this kind of products over involved components of \( j \) words. Such products will serve as canonical factors in the subsequent discussion on the whole product \( \mathbf{C}_{\pi,\beta}(j|\beta) \) in Proposition 3.2. Let \( \ell \) be some positive integer and \( \sigma \in L_2^{2\ell} \). As above, we use the notation \( \mathcal{J}(\sigma, n, \ell) \) to denote the set of all \( \sigma \)-measurable \((n, 2\ell)\)-words \( j \). Moreover, let \( \sigma_0 \in L_2^{2\ell} \). Note that \( \mathbf{C}_{\sigma_0}(j) \) is a product of \( \ell \) 2-element cumulants. Now we define the concept of cycle (with respect to \( \sigma \)) as follows.

**Definition 3.7 (Cycle).** Let \( \sigma_0, \sigma \in L_2^{2\ell} \) and \( j \in \mathcal{J}(\sigma, n, \ell) \). We call the cumulant product \( \mathbf{C}_{\sigma_0}(j) \) a cycle with respect to \( \sigma \) if \( \# \sigma \lor \sigma_0 = 1 \).

**Remark 3.8.** Note that actually for \( j \in \mathcal{J}(\sigma, n, \ell) \), whether \( \mathbf{C}_{\sigma_0}(j) \) is a cycle (with respect to \( \sigma \)) only depends on the perfect matchings \( \sigma \) and \( \sigma_0 \) but not on the choice of \( j \). However, the magnitude of a cycle \( \mathbf{C}_{\sigma_0}(j) \) does depend on the choice of the word \( j \). See Lemma 3.11 below.

We can illustrate the definition of cycle in the following more detailed way through which the meaning of such a nomenclature can be evoked. Provided that \( \# \sigma \lor \sigma_0 = 1 \), it is not difficult to see that there exists a permutation \( \varepsilon \) of \([2\ell]\) such that
\[
\sigma_0 = \{ \{ \varepsilon(2\alpha - 1), \varepsilon(2\alpha) \} \}_{\alpha = 1}^{\ell}, \quad \sigma = \{ \{ \varepsilon(2\alpha), \varepsilon(2\alpha + 1) \} \}_{\alpha = 1}^{\ell}
\]
in which we made the convention of \( \varepsilon(2\ell + 1) = \varepsilon(1) \). Then for all \( j \in \mathcal{J}(\sigma, n, \ell) \), \( \mathbf{C}_{\sigma_0}(j) \) can be written as a product of 2-element cumulants whose indices form a cycle in the sense that if we regard \( V(j) := \bigcup_{\alpha = 1}^{\ell} \{ j_{2\alpha} \} \) as the set of vertices and \( E(j) := \sqcup_{\alpha = 1}^{\ell} \{ j_{2\alpha - 1}, j_{2\alpha} \} \) as the set of edges then the multigraph \( G(j) = (V(j), E(j)) \) is a cycle (i.e. closed walk). Here the notation \( \sqcup \) is the common union while \( \cup \) is the disjoint union. The reader is recommended to take a look at Figure 1 for understanding the definition of a cycle. In this manner, we will also say that the word \( j = (j_1, \ldots, j_{2\ell}) \) forms a cycle under \( \sigma_0 \) with respect to \( \sigma \).

**Definition 3.9 (Product of \( m \) cycles).** Given \( \sigma, \sigma_0 \in L_2^{2\ell} \), if \( \# \sigma_0 \lor \sigma = m \) for some positive integer \( m \geq 2 \), we say that \( \mathbf{C}_{\sigma_0}(j) \) with \( j \in \mathcal{J}(\sigma, n, \ell) \) is a product of \( m \) cycles with respect to \( \sigma \). Actually, if \( \sigma_0 \lor \sigma = \{ D_1, \ldots, D_m \} \), then obviously \( \mathbf{C}_{\sigma_0|D_1}(j|D_i) \) with \( j \in \mathcal{J}(\sigma, n, \ell) \) is a cycle with respect to \( \sigma|D_i \).
Remark 3.10. Note that the definition of product of $m$ cycles also only depends on $\sigma_0$ and $\sigma$. In some sense, unlike the concept of cycle, the product of $m$ cycles is more like a common phrase rather than a new terminology. However, we still raise it as an independent concept to emphasize the relationship between $\sigma_0$ and $\sigma$.

Figure 1. Let $\sigma = \{\{1, 2\}, \{3, 4\}, \{5, 6\}, \{7, 8\}, \{9, 10\}, \{11, 12\}\}$ and $\sigma_0 = \{\{2, 3\}, \{4, 5\}, \{6, 7\}, \{8, 9\}, \{10, 11\}, \{12, 1\}\}$. We take the $\sigma$-measurable word $j$ to be $(1, 1, 2, 2, 1, 1, 3, 3, 4, 4, 4, 4)$. Then the corresponding graph $G(j)$ for the cycle $C_{\sigma_0}(j)$ is as above. Actually, for some specific $j$, the graphical illustration as that in the single cycle case may not evoke the name of product of $m$ cycles any more since different graphs corresponding to different cycles may have coincident vertices thus will be tied together if we define $G(j) = (V(j), E(j))$ as above. To avoid this confusion, we can draw $m$ cycles separately and view the disjoint union of these components as the graph corresponding to the product of $m$ cycles. Moreover, we can use the dash line to connect coincident indices in different components. One can see Figure 2 for example. Actually, when there is some coincidence between indices from different components, we will introduce a merge operation to reduce the number of cycles in the product later. Before commencing this issue, we will raise a fact on the summation of single cycles. Now we have the following lemma.

Lemma 3.11. Let $\ell$ be a fixed positive integer, and $\sigma_0, \sigma \in L^2_{2\ell}$ such that $\#\sigma_0 \vee \sigma = 1$. Assume that $j \in J(\sigma, n, \ell)$, thus $C_{\sigma_0}(j)$ is a cycle with respect to $\sigma$. We have

$$|C_{\sigma_0}(j)| = O(n^{-d(j)}1_{(d(j) \geq 2)}),$$

(3.14)

where $d(j)$ represents the number of distinct values in the collection $\{j_\alpha\}_{\alpha=1}^{2\ell}$.

Proof. As mentioned above, we can assume that $\sigma_0 = \{\{\varepsilon(2\alpha - 1), \varepsilon(2\alpha)\}\}_{\alpha=1}^{\ell}$ and $\sigma = \{\{\varepsilon(2\alpha), \varepsilon(2\alpha + 1)\}\}_{\alpha=1}^{\ell}$ for some permutation $\varepsilon \in P_{2\ell}$. Note that when $d(j) = 1$, all $j$ indices in the cycle take the same value, thus by the fact that $C(Z_1, Z_1) = EZ_1^2 = 1$ we see that the cycle equals 1. Now we consider the
case of \( d(j) \geq 2 \). Note that by the definition of cycle above, we see that there exist at least \( d(j) \) factors \( C(Z_{j_1(2\alpha - 1)}, Z_{j_1(2\alpha)}) \) in the cycle such that two indices in each of these \( d(j) \) factors take different values. To see this we can employ the graphical language. Note that those \( C(Z_{j_1(2\alpha - 1)}, Z_{j_1(2\alpha)}) \) with \( j_1 \in [2\alpha] \) are corresponding to the self-loops in the graph \( G(j) \) defined above. Now if we delete all self-loops in the graph \( G(j) \), the reduced graph is still a cycle and contains edges with two different vertices only. Since the number of the edges of a cycle is always more than or equal to the number of the vertices, we can get (3.14) immediately by invoking the fact that 
\[
C(Z_1, Z_2) = \frac{-1}{n-1}.
\]
\( \square \)

In the sequel, we call a cycle containing at least one factor \( C(Z_{j_1(2\alpha - 1)}, Z_{j_1(2\alpha)}) \) with \( j_1 \in [2\alpha - 1] \) as \textit{in-homogeneous cycle}. Otherwise, we call it \textit{homogeneous cycle}. With the above graphical language, a homogeneous cycle only has a single vertex and all its edges are self-loops. By contrast, an in-homogeneous cycle has at least two vertices. Following from Lemma 3.11 we have

\textbf{Corollary 3.12.} For any given positive integer \( \ell \), and \( \sigma_0, \sigma \in L_{2\ell}^2 \) such that \( \# \sigma \vee \sigma_0 = 1 \), we have

\[
\sum_{j \in J(\sigma,n,\ell)} |C_{\sigma_0}(j)| = n + O(1).
\] (3.15)

\textit{Proof.} By using Lemma 3.11, the leading term of the l.h.s. of (3.15) comes from the homogeneous cycles. Obviously, the total choice of homogeneous cycle is \( n \). Moreover, we can see that the total contribution of the in-homogeneous cycles is \( O(1) \) by (3.14). Thus we obtain (3.15). \( \square \)

Now we use Corollary 3.12 to prove a simple case of Proposition 3.2. That is \( a(\beta) = 1 \) for all \( \beta = 1, \ldots, m \). Note that in this case, by Lemma 3.1, it is not difficult to see that \( \prod_{\beta=1}^m C\{Z_{j_\beta}\}_{\alpha \in A(\beta)} \) is a cycle for \( j_B \in J(\pi_1|_B, n, b) \) since \( \# \pi_B \vee \pi_1|_B = 1 \). Hence, one has

\[
\sum_{j_B \in J(\pi_1|_B, n, b)} \prod_{\beta=1}^m |C\{Z_{j_\beta}\}_{\alpha \in A(\beta)}| = n + O(1).
\] (3.16)

We conclude this subsection by introducing the concept of \textit{merge operation} towards the product of \( m \) cycles when at least two cycles in this product have some coincident indices. Now note that if \( \# \sigma \vee \sigma_0 = 2 \), we see that for \( j \in J(\sigma,n,\ell) \), \( C_{\sigma_0}(j) \) is a product of two cycles by definition. In other words, there exist some permutation \( \varepsilon \) of \( [2\ell] \) and \( \ell_1 \in [\ell] \) such that

\[
\sigma_0 = \{\{\varepsilon(2\alpha - 1), \varepsilon(2\alpha)\}\}_{\alpha=1}^\ell
\]

and

\[
\sigma = \{\varepsilon(2), \varepsilon(3), \ldots, \varepsilon(2\ell_1), \varepsilon(1), \varepsilon(2\ell_1 + 2), \varepsilon(2\ell_1 + 3), \ldots, \varepsilon(2\ell), \varepsilon(2\ell + 1)\}.
\]
Therefore, we have
\[
C_{\sigma_0}(\underline{j}) = \prod_{\alpha=1}^{\ell_1} C(Z_{j_{\ell_1-1}^{2\alpha-1}}, Z_{j_{\ell_1}^{2\alpha}}) \prod_{\alpha=\ell_1+1}^{\ell} C(Z_{j_{\ell_1-1}^{2\alpha-1}}, Z_{j_{\ell_1}^{2\alpha}}).
\]

Now if for some specified \(j \in J(\sigma, n, \ell)\), there exist some \(\beta \in [2\ell_1], \gamma \in [2\ell] \setminus [2\ell_1]\) such that \(j_{\varepsilon(\beta)}\) and \(j_{\varepsilon(\gamma)}\) take the same value, we define the following *merge operation* for the two-cycle product \(C_{\sigma_0}(\underline{j})\). Without loss of generality, we let \(\beta = 2\ell_1, \gamma = 2\ell + 1\). Then in this case we have \(j_{\varepsilon(1)} = j_{\varepsilon(2\ell_1)} = j_{\varepsilon(2\ell_1+1)} = j_{\varepsilon(2\ell)}\) since \(j \in J(\sigma, n, \ell)\). Now by (1.1) we see that
\[
\left|C(Z_{j_{\varepsilon(2\ell_1-1)}}, Z_{j_{\varepsilon(2\ell_1)}})C(Z_{j_{\varepsilon(2\ell_1+1)}}, Z_{j_{\varepsilon(2\ell_1+2)}})\right| \leq \left|C(Z_{j_{\varepsilon(2\ell_1-1)}}, Z_{j_{\varepsilon(2\ell_1+2)}})\right| \tag{3.17}
\]
when \(j_{\varepsilon(2\ell_1)} = j_{\varepsilon(2\ell_1+1)}\). We set
\[
\tilde{\sigma}_0 = (\sigma_0 \setminus \{\varepsilon(2\ell_1-1), \varepsilon(2\ell_1), \varepsilon(2\ell_1+1), \varepsilon(2\ell_1+2)\}) \cup \{\varepsilon(2\ell_1-1), \varepsilon(2\ell_1+2)\}
\]
and
\[
\tilde{\sigma} = (\sigma \setminus \{\varepsilon(1), \varepsilon(2\ell_1)\}) \cup \{\varepsilon(2\ell_1-1), \varepsilon(2\ell_1+2)\} \cup \{\varepsilon(1), \varepsilon(2\ell)\}.
\]
Then we obtain that
\[
C_{\tilde{\sigma}_0}(Z_{j_{\varepsilon(1)}} \cdots Z_{j_{\varepsilon(2\ell_1-1)}}, Z_{j_{\varepsilon(2\ell_1+1)}}, \cdots, Z_{j_{\varepsilon(2\ell)}}) \tag{3.18}
\]
forms a cycle with respect to \(\tilde{\sigma}\). We call (3.18) the *merged cycle* of \(C_{\sigma_0}(\underline{j})\) (See Figure 2 for example). Then by (3.17), we have
\[
\sum_{\substack{j \in J(\sigma, n, \ell) \text{ subject to } j_{\varepsilon(2\ell_1)} = j_{\varepsilon(2\ell_1+1)}}} |C_{\sigma_0}(\underline{j})| \leq \sum_{j \in J(\tilde{\sigma}, n, \ell-1)} |C_{\tilde{\sigma}_0}(Z_{j_{\varepsilon(1)}} \cdots Z_{j_{\varepsilon(2\ell_1-1)}}, Z_{j_{\varepsilon(2\ell_1+1)}}, \cdots, Z_{j_{\varepsilon(2\ell)}})| = n + O(1).
\]

Obviously, the way to do the merge operation may be not unique when there are more than one common value of the vertices from two different cycles. In this case, we can just choose one way to do the merge operation since in the sequel we only care about whether two cycles can be merged but do not care about how to merge them. Analogously, in this manner, when \(\#\sigma \vee \sigma_0 \geq 3\), we can start from two cycles and use the merge operation to merge them into one cycle once there exists at least two indices (one from each cycle) taking the same value. And then we can proceed this merge operation until there is no pair of cycles which can be merged into one.
3.4. Classification of the relationships between indices. Note that the inequalities in Lemma 3.6 rely on the relationships between the underlying indices in the joint cumulants. In order to use Lemma 3.5 and 3.6 in the proof of Proposition 3.2, we will introduce some notation and approach to classify the relationships between the indices (components of $j$).

At first, we introduce the concepts of paired indices and unpaired indices as follows. Note that for any block $\{d_1, d_2\} \in \pi_1$, we have $j_{d_1} \equiv j_{d_2}$ on $J$ by definition. Now for each block $A^{(\beta)} \in \pi_1|_B$, we find out all $\pi_1$’s blocks which are totally contained in $A^{(\beta)}$. We denote the number of such blocks by $s(\beta)$. Specifically, we find out all blocks $D^{(\beta)}_i := \{d^{(\beta)}_1, d^{(\beta)}_2\} \in \pi_1, i = 1, \ldots, s(\beta)$ such that $\bigcup_{i=1}^{s(\beta)} D^{(\beta)}_i \subset A^{(\beta)}$. We then set

$$l^{(\beta)}_i := j_{d^{(\beta)}_1} \equiv j_{d^{(\beta)}_2} \quad \text{on} \quad J, \quad i = 1, \ldots, s(\beta).$$

We call $l^{(\beta)}_i, i = 1, \ldots, s(\beta)$ paired indices from $A^{(\beta)}$ informally. The remaining indices $j_\alpha$ with $\alpha \in A^{(\beta)} \setminus \bigcup_{i=1}^{s(\beta)} D^{(\beta)}_i$ will be ordered (in any fixed order) and denoted by $h^{(\beta)}_i, i = 1, \ldots, 2t(\beta)$ which will be called as unpaired indices from $A^{(\beta)}$. Note that $h^{(\beta)}_i$ should be identical to $h^{(\gamma)}_\ell$ for some $\gamma \neq \beta$ and $\ell \in [2t(\gamma)]$. Obviously, we have $s(\beta) + t(\beta) = a(\beta)$. We remind here the word unpaired means that $h^{(\beta)}_i$ and $h^{(\beta)}_j$ with $i \neq j$ are not identical on $J$. However, for some specific
realization of \( j \in \mathcal{J} \), it is obvious that \( h^{(j)}_{1} \) and \( h^{(j)}_{j} \) may take the same value. Note that since the joint cumulant is a symmetric function of the involved variables, we can work with any specified order of these variables. Therefore, we can write

\[
C\{Z_{j_{\alpha}}\}_{\alpha \in A^{(j)}} = C(Z_{l^{(j)}_{1}}, Z_{l^{(j)}_{1}}, \ldots, Z_{l^{(j)}_{s(j)}}, Z_{h^{(j)}_{1}}, \ldots, Z_{h^{(j)}_{2t(j)}}).
\]

We mimic the notation in Section 3.2 to denote the underlying paired indices and unpaired indices sequence in \( \{Z_{j_{\alpha}}\}_{\alpha \in A^{(j)}} \) by

\[
l^{(j)} := (l^{(j)}_{1}, l^{(j)}_{1}, \ldots, l^{(j)}_{s(j)}, l^{(j)}_{s(j)})
\]

and

\[
h^{(j)} := (h^{(j)}_{1}, h^{(j)}_{2}, \ldots, h^{(j)}_{2t(j)})
\]

respectively. In addition, for simplicity we use the notation

\[
\hat{l}^{(j)} := l^{(j)} h^{(j)}
\]

and write

\[
C(\hat{l}^{(j)} h^{(j)}) := C\{Z_{j_{\alpha}}\}_{\alpha \in A^{(j)}}
\]

by analogy with (3.6). Moreover, we will use the following notation of indices sets

\[
\{l^{(j)}\} := \{l^{(j)}_{1}, \ldots, l^{(j)}_{s(j)}\}, \quad \{h^{(j)}\} := \{h^{(j)}_{1}, \ldots, h^{(j)}_{2t(j)}\}.
\]

We remind here that both \( l^{(j)} \) and \( h^{(j)} \) indices are just \( j \) indices in different notation. We will call an index in \( h^{(j)} \), \( \beta \in [m] \) as \( h \) index. And \( l \) index can be understood analogously. Moreover, when we refer to the position of an \( l \) or \( h \) index, we always mean the position of its corresponding \( j \) index in the word \( j \). In the sequel, we will also employ the notation

\[
B(h) = \{\alpha \in B : j_{\alpha} \text{ is an } h \text{ index}\}.
\]

(3.19)

Note that we can regard \( h_{\alpha}^{(j)} \) and \( h_{\gamma}^{(j)} \) as two different free indices in \( [n] \) when we take sum over \( \mathcal{J} \). However, in Lemma 3.6, the bound on the magnitude of \( C(\hat{l}^{(j)} h^{(j)}) \) may be different according to whether \( h_{\alpha}^{(j)} \) and \( h_{\gamma}^{(j)} \) take the same value or not. Therefore, it is necessary to decompose the summation according to different relationships between \( h \) indices. For example,

\[
\sum_{h_{1}} \sum_{h_{2}} |C(Z_{h_{1}}, Z_{h_{2}})| = \sum_{h_{1}=h_{2}} |C(Z_{h_{1}}, Z_{h_{2}})| + \sum_{h_{1} \neq h_{2}} |C(Z_{h_{1}}, Z_{h_{2}})|. \quad (3.20)
\]

In the above example, the terms from the first summation on the r.h.s. of (3.20) (each term equals 1) are quite different from those from the second summation (each term equals \(-1/(n-1)\)). For more general \( C(\hat{l}) \), we will introduce the following concept of relationship matrix.
Definition 3.13 (Relationship matrix). For some positive integer \( N \), we assume that \( \ell_1, \ldots, \ell_N \in [n] \) and denote \( \bar{\ell} := (\ell_1, \ldots, \ell_N) \). Let \( R_{\bar{\ell}} = (\delta_{ij})_{N,N} \) with

\[
R_{\bar{\ell}}(i, j) := \delta_{ij} = \begin{cases} 
1, & \text{if } \ell_i = \ell_j \\
0, & \text{if } \ell_i \neq \ell_j
\end{cases}
\]

We call \( R_{\bar{\ell}} \) the relationship matrix of \( \bar{\ell} \).

Remark 3.14. Note that not all \( 0 \rightarrow 1 \) matrix can be a relationship matrix. For example, a matrix \( M \) with \( M(1,2) = M(2,3) = 1 \) while \( M(1,3) = 0 \) can not be a relationship matrix.

Example 3.1. For the vector \( \mathbf{j} = (1,1,1,2,3,2) \), we see the relationship matrix of \( \mathbf{j} \) is the block diagonal matrix

\[
R_{\mathbf{j}} = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix} \oplus \begin{pmatrix}
1 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 \\
0 & 1 & 1 & 0 \\
1 & 0 & 0 & 1
\end{pmatrix}.
\]

3.5. Proof of Proposition 3.2. We recall the notation in Section 3.4 to write \( C\{Z_{\alpha}\}_{\alpha \in \mathcal{A}(\beta)} \) as

\[
C(\mathbf{\hat{\mathbf{h}}}^{(\beta)}) = C(Z_{i_1}^{(\beta)}, Z_{t_1}^{(\beta)}, \ldots, Z_{i_t}^{(\beta)}, Z_{s_1}^{(\beta)}, Z_{h_1}^{(\beta)}, Z_{h_2}^{(\beta)}, \ldots, Z_{h_{2s}}^{(\beta)}),
\]

where \( s(\beta), t(\beta) \geq 0 \) are nonnegative integers and \( s(\beta) + t(\beta) = a(\beta) \). Thus our aim is to bound the following quantity

\[
\sum_{\mathbf{j}|B \in \mathcal{J}(\pi_1|B,n,b)} \prod_{\beta=1}^{m} |C(\mathbf{\hat{\mathbf{h}}}^{(\beta)})| = \sum_{\mathbf{h}|B \in \mathcal{J}(\pi_1|B,n,b)} \prod_{\beta=1}^{m} \sum_{1(\beta)}^{n} |C(\mathbf{\hat{\mathbf{h}}}^{(\beta)})|.
\]

(3.21)

Here \( \sum_{\mathbf{h}|B \in \mathcal{J}(\pi_1|B,n,b)} \) represents the summation over all choices of \( \mathbf{h} \) indices along with \( \mathbf{\mathbf{j}|B} \) running all over \( \mathcal{J}(\pi_1|B,n,b) \), and

\[
\sum_{1(\beta)}^{n} = \sum_{t_1(\beta)=1}^{n} \cdots \sum_{t_s(\beta)=1}^{n}.
\]

(3.22)

At first, we can handle the trivial case of \( m = 1 \) for Proposition 3.2 as follows. Observe that in this case, \( A^{(1)} = B \) thus itself forms a block of \( \pi \lor \pi_1 \). Hence \( t(1) = 0 \) and \( s(1) = a(1) \). If \( a(1) \leq 2 \), it is easy to see that Proposition 3.2 holds by employing (1.1) and Lemma 3.5. For the case of \( a(1) \geq 3 \), we need to use Lemma 3.6 by setting \( h_1 = h_2 \) in (ii) and (iii) therein. With the aid of this setting, we can get the conclusion by noticing that except for the cases of (ii) and (iii) in Lemma 3.6 (with \( h_1 = h_2 \) therein), the number of free indices in any other case is at most \( a(1) - 2 \). More specifically, we can split the summation (3.22) as

\[
\sum_{1(1)} = \sum_{1(1)}^{*} + \sum_{1(1)}^{*} + \sum_{1(1)}^{*}
\]
where $\sum^*$ is the summation runs over the sequences $(l^{(1)}_1, \ldots, l^{(1)}_{s(1)}) \in [n]^{s(1)}$ in which all indices are distinct from each other; $\sum^*$ runs over the sequences in which except for one pair of coincidence indices all the others are distinct and distinct from this pair; $\sum^*$ runs over all the remaining cases. Note that the total number of the choices of indices in $\sum^*$ is $O(n^{a(1)-2})$. Then by using (ii) and (iii) of Lemma 3.6, we can actually get the stronger bound as $O(n^{a(1)-2})$ rather than $O(n^{a(1)-1})$.

Therefore, it suffices to consider the case of $m \geq 2$. We summarize our strategy as follows. Obviously we have $t(\beta) \geq 1$ for all $\beta = 1, \ldots, m$ by (ii) of Lemma 3.1. We raise the concepts “Form $\gamma$” to name three different forms of factors. They are as follows.

Form 1: $C(\hat{h}^{(\beta)})$ with $s(\beta) = 0$ and $t(\beta) = 1$,
Form 2: $C(\hat{h}^{(\beta)})$ with $s(\beta) \geq 1$ and $t(\beta) \geq 1$,
Form 3: $C(\hat{h}^{(\beta)})$ with $s(\beta) = 0$ and $t(\beta) \geq 2$.

Observe that the Form 1 factors are 2-element cumulants while Form 2 and 3 factors contain more elements. The main theme of our evaluation scheme in the sequel is to transform Forms 2 and 3 factors into products of 2-element cumulants and finally use the arguments on the sum of cycles in Section 3.3. More specifically, firstly we will pick Form 2 factors out and take sum over paired indices, and bound the sum by the product of factors with unpaired indices with the aid of Lemma 3.5 and 3.6. Next, we need to decompose each Form 3 factor into a product of 2-element cumulants by invoking Lemma 3.5 and 3.6 again. Consequently, we will obtain a product with all factors being 2-element cumulants. Then we can sort these 2-element cumulants into several cycles. Finally, we can get the desired bound by discussions on the summation over these cycles with the aid of Lemma 3.11. Below we unfold this strategy in details.

Now we start with a factor with underlying indices from $A^{(\beta)}$ such that $s(\beta) \geq 1$, i.e. Form 2 factor. For given $h$ indices, we draw up a rule to bound such factor according to various relationships between $l$ and $h$ indices. At first, we classify the $l$ indices for given $h$ indices into the following cases.

(I): $l^{(\beta)}_1, \ldots, l^{(\beta)}_{s(\beta)}$ are not mutually distinct or $\{l^{(\beta)}_1\} \cap \{h^{(\beta)}\} \neq \emptyset$.

(II): $l^{(\beta)}_1, \ldots, l^{(\beta)}_{s(\beta)}$ are mutually distinct and $\{l^{(\beta)}\} \cap \{h^{(\beta)}\} = \emptyset$.

Given $h^{(\beta)}$, we denote the set consisting of the index sequence $l^{(\beta)} \in [n]^{s(\beta)}$ satisfying (I) (resp. II) by $\mathcal{L}_1^{(\beta)} := \mathcal{L}_1^{(\beta)}(h^{(\beta)})$ (resp. $\mathcal{L}_2^{(\beta)} := \mathcal{L}_2^{(\beta)}(h^{(\beta)})$). Note that in case (I), the total choice of $l$ indices is bounded by $O(n^{s(\beta)-1})$ when $h$ is given.
Using Lemma 3.5 and (i) of Lemma 3.6 it is easy to see that
\[
\sum_{I^{(b)} \in L_2^{(b)}} |C(\hat{h}^{(b)})| \leq O(n^{s(b)-1}) \sum_{\tau \in L_2^{2t(b)}} |E_\tau(h^{(b)})|.
\] (3.23)

Now when \( t(b) = 1 \), by using Lemma 3.5 and (ii) of Lemma 3.6 we see that
\[
\sum_{I^{(b)} \in L_2^{(b)}} |C(\hat{h}^{(b)})| \leq O(n^{s(b)-1})|E(Z_{h_1^{(b)}} Z_{h_2^{(b)}})|.
\] (3.24)

For \( t(b) \geq 2 \), we further split \( h \) indices into two cases denoted by \( \hat{S}^{(b)}_{21} \) and \( \hat{S}^{(b)}_{22} \) according to the relationships between components of \( h^{(b)} \in [n]^{2t(b)} \). \( \hat{S}^{(b)}_{21} \) consists of the \( h^{(b)} \) index sequence in one of the following cases: (1): \( h^{(b)} \) is \( \sigma \)-measurable for some \( \sigma \in L_2^{2t(b)} \) and the indices with positions in different blocks are distinct from each other, (2): \( h^{(b)} \) is \( \sigma \)-measurable for some \( \sigma \in L_4^{2t(b)} \) and the indices with positions in different blocks are distinct from each other, (3): there exist \( \alpha, \gamma \in [2t(b)] \) such that \( h_{b_1}^{(b)} \neq h_{b_2}^{(b)} \), and the vector obtained by deleting \( h_{b_1}^{(b)} \) and \( h_{b_2}^{(b)} \) from \( h^{(b)} \) is \( \tilde{\sigma} \)-measurable for some \( \tilde{\sigma} \in L_2^{2t(b)-2} \) and the indices with positions in different blocks are distinct from each other and also distinct from \( h_{b_1}^{(b)} \) and \( h_{b_2}^{(b)} \). Let \( \hat{S}^{(b)}_{22} \) be the complement of \( \hat{S}^{(b)}_{21} \) in the full set of all choices of \( h^{(b)} \) indices.

By using (ii) and (iii) of Lemma 3.6 we can see that when \( h^{(b)} \in \hat{S}^{(b)}_{21} \),
\[
\sum_{I^{(b)} \in L_2^{(b)}} |C(h^{(b)})| \leq O(n^{s(b)-1}) \sum_{\tau \in L_2^{2t(b)}} |E_\tau(h^{(b)})|.
\] (3.25)

by analogy with the bound (3.13). For \( h^{(b)} \in \hat{S}^{(b)}_{22} \), we just use the crude bound (i) in Lemma 3.6 to get
\[
\sum_{I^{(b)} \in L_2^{(b)}} |C(\hat{h}^{(b)})| \leq O(n^{s(b)}) \sum_{\tau \in L_2^{2t(b)}} |E_\tau(h^{(b)})|.
\] (3.26)

Now given \( h \) indices, for each Form 2 factor, we take the sum over the paired \( I^{(b)} \) indices by using (3.23)-(3.26) at first. Then after this step, we can bound (3.21) by sum of products of several Form 1 or 3 factors, i.e. sum over \( h \) indices. If \( t(b) = 1 \) for all \( b \in [m] \), after the first step, the summation over \( h \) indices is just the simple case discussed at the end of Section 3.3, see (3.16). Thus by using (3.23) and (3.24), we easily get that (3.4) holds by noticing (3.16) and the fact that \( s(b) - 1 = a(b) - 2 \).

If not all \( t(b) \) equal 1, we deal with the bound obtained after the first step as follows. For given \( h \) indices, we introduce the following sets of \( \beta \)
\[
\mathfrak{B}_1 = \{ \beta : s(\beta) \geq 1 \text{ and } t(\beta) = 1 \},
\]
\[
\mathfrak{B}_2 = \{ \beta : s(\beta) \geq 1 \text{ and } t(\beta) \geq 2 \}.
\]
Moreover, we further split \( \mathcal{B}_2 := \mathcal{B}_{21} \cup \mathcal{B}_{22} \), where
\[
\mathcal{B}_{21} = \{ \beta : s(\beta) \geq 1 \text{ and } h^{(\beta)} \in \mathcal{S}_{21}^{(\beta)} \}, \\
\mathcal{B}_{22} = \{ \beta : s(\beta) \geq 1 \text{ and } h^{(\beta)} \in \mathcal{S}_{22}^{(\beta)} \}.
\]
Note that \( \mathcal{B}_1 \) and \( \mathcal{B}_2 \) only depend on the given partition \( \pi | _B \). However, \( \mathcal{B}_{21} \) and \( \mathcal{B}_{22} \) depend on the values of \( h \) indices according to the definitions of \( \mathcal{S}_{21}^{(\beta)} \) and \( \mathcal{S}_{22}^{(\beta)} \). In addition, we set
\[
\mathcal{B}_1^0 = \{ \beta : s(\beta) = 0 \text{ and } t(\beta) = 1 \}, \\
\mathcal{B}_2^0 = \{ \beta : s(\beta) = 0 \text{ and } t(\beta) = 2 \}, \\
\mathcal{B}_3^0 = \{ \beta : s(\beta) = 0 \text{ and } t(\beta) \geq 3 \}.
\]
and \( \mathcal{B}^0 = \mathcal{B}_1^0 \cup \mathcal{B}_2^0 \cup \mathcal{B}_3^0 \). Note that given \( h^{(\beta)} , \beta = 1, \ldots, m \), by (3.23)-(3.26) we see
\[
\prod_{\beta = 1}^m \sum_{\beta} | C(\hat{h}^{(\beta)}) | \leq C \prod_{\beta \in \mathcal{B}_1 \cup \mathcal{B}_{21}} n^{s(\beta)-1} \prod_{\beta \in \mathcal{B}_{22}} n^{s(\beta)} | F(h) |, \tag{3.27}
\]
where \( C \) is some large positive constant and
\[
F(h) = \prod_{\beta \in \mathcal{B}^0} C(h^{(\beta)}) \prod_{\beta \in \mathcal{B}_1 \cup \mathcal{B}_{21}} \sum_{\tau(\beta) \in L_{21}^{(\beta)}} E_{\tau(\beta)}(h^{(\beta)}).
\]
We can write
\[
F(h) = \sum_{\tau(\beta) \in L_{21}^{(\beta)}, \beta \in \mathcal{B}^0} \prod_{\beta \in \mathcal{B}_1 \cup \mathcal{B}_{21}} C(h^{(\beta)}) \prod_{\beta \in \mathcal{B}_1 \cup \mathcal{B}_{22}} E_{\tau(\beta)}(h^{(\beta)})
:= \sum_{\tau(\beta) \in L_{21}^{(\beta)}, \beta \in \mathcal{B}_1 \cup \mathcal{B}_{22}} f_\tau(h), \tag{3.28}
\]
where we use \( \tau \) to denote the sequence \( \{ \tau(\beta) \}_{\beta \in \mathcal{B}_1 \cup \mathcal{B}_{21}} \) thus \( f_\tau(h) \) is indexed by the sequence of perfect matchings. Therefore, it suffices to bound
\[
\sum_{h : | h |_B \in F(\pi | _B , n,b)} \prod_{\beta \in \mathcal{B}_1 \cup \mathcal{B}_{21}} n^{s(\beta)-1} \prod_{\beta \in \mathcal{B}_{22}} n^{s(\beta)} | f_\tau(h) |
\]
for any fixed \( \tau \).

Now we use the concept of relationship matrix for the \( h \) indices defined in Section 3.4. At first, we can order all \( h \) indices by the order of the positions of their corresponding original \( j \) indices as components in the word \( j \), and arrange it into a vector with length \( 2 \sum_\beta t(\beta) \) according to this order. We use the notation \( \mathbf{h} \) to denote this vector. Let \( R_{\mathbf{h}} \) be the relationship matrix of \( \mathbf{h} \). Note that for a fixed relationship matrix \( M \), there are a lot of \( \mathbf{h} \in [n]^2 \sum_\beta t(\beta) \) satisfying \( R_{\mathbf{h}} = M \). Moreover, the total number of choices of the relationship matrix is determined by
\( \pi \) and \( \pi_1 \) thus only depends on \( k \) but not \( n \). Hence, it suffices to fix an \( M \) and consider
\[
\sum_{h, j | B \in \mathcal{J}(\pi_1 | B, n, b)} \prod_{b \in \mathcal{B}_1 \cup \mathcal{B}_2} n^{s(b) - 1} \prod_{\beta \in \mathcal{B}_2} n^{s(\beta)} |f_{\mathcal{F}}(h)|,
\]
(3.29)

Now note that in the collection \( \{ h : j | B \in \mathcal{J}(\pi_1 | B, n, b) \text{ s.t. } R_{\mathbb{N}} = M \} \) with given \( M \), the sets \( \mathcal{B}_2 \) and \( \mathcal{B}_2 \) only depend on \( M \). Moreover, note that we can write
\[
f_{\mathcal{F}}(h) = \prod_{\beta \in \mathcal{B}_1} C(h^\beta) \prod_{\beta \in \mathcal{B}_2} E_{\tau(\beta)}(h^\beta).
\]

Now denote
\[
u(M) := \# \mathcal{B}_2, \text{ for } R_{\mathbb{N}} = M.
\]

Then we can write
\[
(3.29) = n^{u(M)} \prod_{\beta \in \mathcal{B}_1 \cup \mathcal{B}_2} n^{s(\beta) - 1}
\times \sum_{h, j | B \in \mathcal{J}(\pi_1 | B, n, b)} \prod_{\beta \in \mathcal{B}_1} |C(h^\beta)| \prod_{\beta \in \mathcal{B}_2} |E_{\tau(\beta)}(h^\beta)| \prod_{\beta \in \mathcal{B}_2} |C(h^\beta)|.
\]
(3.30)

Now for \( \beta \in \mathcal{B}_2^0 \), similar to the definition of \( \mathcal{B}_2 \) and \( \mathcal{B}_2 \), we introduce the sets \( \mathcal{B}_2^0 \) and \( \mathcal{B}_2^0 \) as follows.
\[
\mathcal{B}_2^0 := \{ \beta \in \mathcal{B}_2 : \text{at least three of } h^\beta_i, i = 1, \ldots, 4 \text{ take the same value} \},
\mathcal{B}_2^0 := \mathcal{B}_2 \setminus \mathcal{B}_2^0.
\]

We further denote
\[
u(M) := \# \mathcal{B}_2^0, \text{ for } R_{\mathbb{N}} = M.
\]

Note that for \( \beta \in \mathcal{B}_2^0 \), we can employ (3.8). Hence, we can choose and fix a sequence of perfect matchings \( \tilde{\sigma} = \{ \sigma(\beta) \}_{\beta \in \mathcal{B}_2^0} \) such that for some positive constant \( C \),
\[
\sum_{h, j | B \in \mathcal{J}(\pi_1 | B, n, b)} \prod_{\beta \in \mathcal{B}_1} |C(h^\beta)| \prod_{\beta \in \mathcal{B}_2} |E_{\tau(\beta)}(h^\beta)| \prod_{\beta \in \mathcal{B}_2^0 \cup \mathcal{B}_2^0} |C(h^\beta)|
\leq C n^{-v(M)} \sum_{h, j | B \in \mathcal{J}(\pi_1 | B, n, b)} |g_{\mathcal{F}, \tilde{\sigma}}(h)|,
\]
(3.31)
Now for $R_{\text{h}} = M$, we can further decompose $C(h^{(B)})$ for $\beta \in B_3^0$ according to Lemma 3.6. To this end, we need to further split $B_3^0$ in analogy to what we have done to $B_2^0$. Let

$$B_{31} := \{ \beta \in B_3^0 : h^{(B)}(\beta) \in \hat{\mathcal{H}}_{31} \}, \quad B_{32} := \{ \beta \in B_3^0 : h^{(B)}(\beta) \in \hat{\mathcal{H}}_{32} \}.$$ 

Note that $B_{31}$ and $B_{32}$ are also determined by the relationship matrix $M$. We will use the notation

$$w(M) := \#B_{31}, \text{ for } R_{\text{h}} = M.$$ 

Then by employing Lemma 3.6 we have

$$|g_{\vec{\tau}, \vec{\sigma}}(h)| \leq n^{-w(M)} \prod_{\beta \in B_3^0 \cup B_1} |C(h^{(B)})| \prod_{\beta \in B_2} |E_{\tau(\beta)}(h^{(B)})| \prod_{\beta \in B_3^0} |E_{\sigma(\beta)}(h^{(B)})|$$

$$\times \prod_{\beta \in B_3^0} \sum_{\omega^{(\beta)} \in E_{\omega(\beta)}(h^{(B)})} |E_{\omega(\beta)}(h^{(B)})|$$

$$:= n^{-w(M)} \sum_{\omega^{(\beta)}, \beta \in B_3^0} |w_{\vec{\tau}, \vec{\sigma}, \vec{\omega}}|,$$

where we used $\vec{\omega}$ to denote the sequence of $\omega^{(\beta)}$. Hence it suffices to evaluate

$$\sum_{h, j | h, j \in \mathcal{J}(\pi_1 | B, n, b) \text{ s.t. } R_{\text{h}} = M} |w_{\vec{\tau}, \vec{\sigma}, \vec{\omega}}| = \sum_{h, j | h, j \in \mathcal{J}(\pi_1 | B, n, b) \text{ s.t. } R_{\text{h}} = M} \prod_{\beta \in B_2} |C(h^{(B)})|$$

$$\times \prod_{\beta \in B_2} |E_{\tau(\beta)}(h^{(B)})| \prod_{\beta \in B_3^0} |E_{\sigma(\beta)}(h^{(B)})| \prod_{\beta \in B_3^0} |E_{\omega(\beta)}(h^{(B)})|, (3.33)$$

for fixed $\vec{\tau}, \vec{\sigma}$ and $\vec{\omega}$.

As desired, we see that $w_{\vec{\tau}, \vec{\sigma}, \vec{\omega}}$ is a product of 2-element cumulants. Now note that we can rewrite $w_{\vec{\tau}, \vec{\sigma}, \vec{\omega}}$ as a product of several cycles. If we view the tuple of two positions (in original $j$ word) of underlying $h$ indices in each 2-element cumulant of $w_{\vec{\tau}, \vec{\sigma}, \vec{\omega}}$ as a block, then we have a natural perfect matching. We denote this perfect matching by $\pi(\vec{\tau}, \vec{\sigma}, \vec{\omega})$. Now analogously we use $\pi_1 | B(h)$ to denote the partition of $\pi_1$ restricted on $B(h)$ (see (3.19) for the definition). Assume that

$$\textbf{c} := \#\pi_1 | B(h) \lor \pi(\vec{\tau}, \vec{\sigma}, \vec{\omega}).$$

Then $w_{\vec{\tau}, \vec{\sigma}, \vec{\omega}}$ is a product of $\textbf{c}$ cycles with respect to $\pi_1 | B(h)$ for all $h$ subject to $j_{|B} \in \mathcal{J}(\pi_1 | B, n, b)$. However, for the subset $\{ h : j_{|B} \in \mathcal{J}(\pi_1 | B, n, b) \text{ s.t. } R_{\text{h}} = M \}$, since we imposed additional constraint on the coincidence condition of the $h$ indices, the number of cycles may can be reduced further by the merge operation defined in Section 3.3. We use $\textbf{c}$ to denote the number of cycles after using merge operation on the original product of $\textbf{c}$ cycles for the set of $h$ satisfying $R_{\text{h}} = M$. 

where

$$g_{\vec{\tau}, \vec{\sigma}}(h) := \prod_{\beta \in B_2} C(h^{(B)}) \prod_{\beta \in B_2} E_{\tau(\beta)}(h^{(B)}) \prod_{\beta \in B_3^0} E_{\sigma(\beta)}(h^{(B)}) \prod_{\beta \in B_3^0} C(h^{(B)}).$$
Now by switching rows and columns of $M$, we can get a block diagonal matrix $\bar{M} = \oplus_{i=1}^\ell M_i$ with each block describing the relationships between the indices in one cycle or one product of cycles which can be merged into one. More specifically, we let $\bar{h}_i, i\in[\bar{\ell}]$ be the set of underlying indices in a cycle or a product of cycles which can be merged into one cycle. And analogous to $\tilde{h}$ we can introduce the vector $\bar{h}_i$ of ordered indices of $h_i$ (ordered according to the positions in the original $j$ words). Then we denote the corresponding relationship matrix block of $\bar{h}_i$ by $M_i, i = 1, \ldots, \bar{\ell}$. Now note that

$$\sum_{h, j \mid \pi_j \in \mathcal{J}(\pi_1|B, n, b)} \left| w_{\pi_\ell, \pi_\ell, \pi_\ell} \right| \leq \prod_{i=1}^\ell \sum_{h_i \mid \pi_i \in \mathcal{J}(\pi_1|B, n, b)} \left| w_{\pi_i, \pi_i, \pi_i} \right|,$$

(3.34)

where $w_{\pi_\ell, \pi_\ell, \pi_\ell} | h_i$ represents the cycle or product of cycles with indices in $h_i$. The inequality above arises because we release the constraints between different cycles or product of cycles on the l.h.s. of (3.34).

Now we define a graph $G(M) := (V(M), E(M))$, where $V(M) = \{h_i\}_{i \in [\bar{\ell}]}$ and $(h_i, h_\ell) \in E(M)$ if and only if $i \neq \ell$ and there exists some $\beta \in [m]$ such that both $h_i$ and $h_\ell$ contain at least one index from the common $A(\beta)$. In this case, we say that $h_i$ and $h_\ell$ are connected by $A(\beta)$. We also say that $A(\beta)$ contributes in $h_i$ if $h_i$ has at least one index from $A(\beta)$. Then we have the following lemma.

**Lemma 3.15.** Under the above definition, $G(M)$ is connected.

**Proof.** We can prove it simply by contradiction. We assume that $G(M)$ has more than 2 connected components. Note that by definition those $h$ indices from the same $A(\beta)$ can have contribution in only one connected component of $G(M)$. Thus the set of positions of $h$ indices in one connected component of $G(M)$ is the union of some blocks in $\pi_1|B$. By the definition of cycle with respect to $\pi_1|B$, we know that the set of positions of the indices in one cycle is also the union of some blocks in $\pi_1|B$. Then obviously one has $\# \pi_1|B \vee \pi|B \geq 2$. However, we already know that $\# \pi_1|B \vee \pi|B = 1$. Thus we get the contradiction. 

In the sequel, we regard each $h_i$ as a cycle though some of $h_i$ maybe actually a product of several cycles which can merged into one cycle, since as explained in Section 3.3 when we take sum over such kind of cumulant product we can bound it by the sum over the merged cycle. For simplicity, we will also informally call the set $h_i$ $i$-th cycle in the following discussion. In this manner, we notice that only the set $A(\beta)$ with $\beta \in \mathcal{B}_2 \cup \mathcal{B}_4' \cup \mathcal{B}_3'$ can connect different cycles, since the whole $A(\beta)$ can contribute only in one cycle if $\beta \in \mathcal{B}_2'$. Under the aforementioned viewpoint that the product of cycles which can be merged into one should be regarded as a single cycle. Therefore, by Lemma 3.15 each cycle must contain at least one pair of indices from $A(\beta)$ with $\beta \in \mathcal{B}_2 \cup \mathcal{B}_4' \cup \mathcal{B}_3'$. Relying on this fact, we can bound the product on the r.h.s. of (3.34) through bounding its factors. We employ the following procedure.

- **Step 1:** We start with a cycle and take sum over this cycle, then we can get an
\( O(n) \) factor for the product on the r.h.s. of (3.34). Without loss of generality, we assume the initial cycle is \( h_1 \).

- **Step 2**: Then we find a cycle which is connected to \( h_1 \) through a set \( A^{(\beta)} \) with \( \beta \in \mathcal{B}_2 \cup \mathcal{B}_{21}^0 \cup \mathcal{B}_3^0 \). Such \( A^{(\beta)} \) may be not unique. We can discuss them one by one. Therefore, we can fix one \( A^{(\beta)} \) which contributes in both \( h_1 \) and some other cycles at first. We need to take Step 2 according to the following cases.

  (i): If \( \beta \in \mathcal{B}_{21}^0 \), we can again take the sum over the second cycle then get an \( O(n) \) factor for the whole product (3.34).

  (ii): If \( \beta \in \mathcal{B}_{22} \), we find out all cycles in the product (3.34) which are contributed by this \( A^{(\beta)} \) and evaluate the product of all sums over these cycles. Note that if \( \beta \in \mathcal{B}_{22} \), since one pair of indices in \( h^{(\beta)} \) is already taken into account in the the former cycle (\( h_1 \)), thus the product of the sum over all the remaining cycles (except for \( h_1 \)) contributed by \( A^{(\beta)} \) has a contribution of the order \( O(n^{t(\beta)-1}) \) at most.

  (iii): If \( \beta \in \mathcal{B}_{22} \), there are two cases. One is that beside of the indices in \( h_1 \), the remaining indices from \( A^{(\beta)} \) can be contributed at most in \( t(\beta) - 2 \) cycles if there exists at least three indices in \( A^{(\beta)} \) taking the same value. The other case is that we do have \( t(\beta) - 1 \) cycles contributed by the remaining indices from \( A^{(\beta)} \). However, at least one cycle is in-homogeneous thus the contribution of the sum taken over this cycle is only \( O(1) \). In any case, we get that the product of sum over all the remaining cycles (except for the \( h_1 \)) contributed by \( A^{(\beta)} \) has a contribution of the order \( O(n^{t(\beta)-2}) \) at most.

  (iv): If \( \beta \in \mathcal{B}_{31}^0 \), we can use Lemma 3.6 again and an analogous discussion as (ii) to see that the product of sum over all the remaining cycles (except for \( h_1 \)) contributed by \( A^{(\beta)} \) has a contribution of the order \( O(n^{a(\beta)-1}) \) at most.

  (v): If \( \beta \in \mathcal{B}_{32}^0 \), we can use an analogous discussion as (iii) to see that the product of sum over all the remaining cycles (except for \( h_1 \)) contributed by \( A^{(\beta)} \) has a contribution of the order \( O(n^{a(\beta)-2}) \) at most.

  (vi): If \( h_1 \) is contributed by more than one \( A^{(\beta)} \) with \( \beta \in \mathcal{B}_2 \cup \mathcal{B}_{21}^0 \cup \mathcal{B}_3^0 \), we should use the discussion (i)-(v) on all these \( A^{(\beta)} \)'s one by one.

- **Step 3**: Then we can start from any cycle which has been already evaluated and find a subsequent cycle which is connected to it by a set \( A^{(\beta)} \) with \( \beta \in \mathcal{B}_2 \cup \mathcal{B}_{21}^0 \cup \mathcal{B}_3^0 \) and has not been evaluated yet, and repeat the above discussion.

Finally, we can get that for some positive constant \( C \),

\[
\sum_{h_0|b|v \in \mathcal{J} \setminus \{h_0, n, b\}} |w_{\pi, \tau, \sigma, \varphi}| \quad \text{s.t. } R_n = M
\leq Cn \cdot \prod_{\beta \in \mathcal{B}_{21}^0} n^{t(\beta)-1} \prod_{\beta \in \mathcal{B}_{22}} n^{t(\beta)-2} \prod_{\beta \in \mathcal{B}_{31}^0} n^{a(\beta)-1} \prod_{\beta \in \mathcal{B}_{32}^0} n^{a(\beta)-2}
\]
The first factor $n$ on the r.h.s. of (3.35) arises in Step 1 since in the initial cycle we may have not used up any $A^{(β)}$ with $β \in \mathcal{B}_2 \cup \mathcal{B}_1 \cup \mathcal{B}_3$. Then by using (3.29)-(3.32) we obtain

\[
(3.29) \leq Cn^{u(M)-v(M)-w(M)} \prod_{β \in \mathcal{B}_1 \cup \mathcal{B}_2} n^{s(β)-1} \times n \prod_{β \in \mathcal{B}_{21}} n^{t(β)-1} \prod_{β \in \mathcal{B}_{22}} n^{t(β)-2} \prod_{β \in \mathcal{B}_{31}} n^{a(β)-1} \prod_{β \in \mathcal{B}_{32}} n^{a(β)-2}
\]

for some positive constant $C$. Then by invoking the definition of $u(M), v(M)$ and $w(M)$ and the fact that $s(β) + t(β) = a(β)$, we can immediately get that

\[
(3.29) \leq Cn^{1+\sum_{β:a(β) \geq 2}(a(β)-2)}.
\]

Then Proposition 3.2 follows from (3.21), (3.27) and (3.28). Thus we complete the proof.

3.6. A special case. In the sequel, we also need the following stronger bound for the special case of $\# \{β : a(β) \geq 3\} = 1$ while $\# \{β : a(β) = 2\} = 0$.

**Proposition 3.16.** When $\# \{β : a(β) \geq 3\} = 1$ and $\# \{β : a(β) = 2\} = 0$ for some $B \in \pi \setminus π_1$, we have

\[
\sum_{j∈J(π|B,n,k)} |C_{jA} |_{α∈A(β)}| = O\left(n^{\sum_{β=1}^{m}(a(β)-2)1_{a(β)=2}}\right).
\]

**Remark 3.17.** Such an improvement is possible for more general cases. However, in the sequel we only need this improved bound for the above special case. Moreover, one might have already noticed that in the bounding procedure in the last subsection, we did not make full use of Lemma 3.6. For example, in (3.24), we did not use the better bound $O(n^{-2})$ when $h_1(β) = h_2(β)$. Though the bound in (3.24) is the optimal one following from (ii) of Lemma 3.6 when $h_1(β) \neq h_2(β)$, we did not use the fact that an unequal index pair will bring an in-homogeneous cycle in the later discussion. In the other cases discussed in the last subsection, similar improvement can be made via a more detailed evaluation indeed. That is why the improved bound in Lemma 3.16 can be achieved.

From the above proposition we can immediately get the following corollary.

**Corollary 3.18.** If $π ∈ L_{2k}^{even}$ such that $n_2(π) = 0$ and $\sum_{γ \geq 3} n_γ(π) = 1$, we have

\[
\sum_{j∈J} |C_γ(j)| = O(n^{#π_1 ∨ π - 1 + \sum_{γ \geq 2}(γ - 2)n_γ(π)}).
\]

**Proof of Proposition 3.16.** Now let $A^{(γ)}$ be the unique block of $π|B$ whose cardinality is at least 6 (i.e. $a(γ) \geq 3$). We pursue the strategy in the proof of
Proposition 3.2 to fix a relationship matrix $M$ for $h$ indices and focus on the estimation of the sum
\[
\sum_{h,j \in J(\{ \pi_1, n, b \})} \prod_{\beta=1}^{m} |C(h^{(\beta)})|,
\] (3.36)

At first, we deal with the case of $s(\gamma) = 0, t(\gamma) = a(\gamma)$. In this case (3.36) is reduced to be
\[
\sum_{h,j \in J(\{ \pi_1, n, b \})} \prod_{\beta=1}^{m} |C(h^{(\beta)})|,
\] (3.37)
since by (ii) of Lemma 3.1 we know that except $C(h^{(\gamma)})$ all the other factors are 2-element cumulants with $h$ indices. Analogous to the discussions on $B_0^3$ factors in the proof of Proposition 3.2, we can bound $|C(h^{(\gamma)})|$ by some product of 2-element cumulants. Now recall the definition of the set $\mathcal{B}_{21}^{(\gamma)}$ and its three cases (below (3.24)). Note that by using (ii) of Lemma 3.6 (set $h_1 = h_2$ therein) we can see that if $h^{(\gamma)}$ is in the first case of the definition of $\mathcal{B}_{21}^{(\gamma)}$ when $R^*_h = M$, there exists a perfect matching $\omega \in L_{2a(\gamma)}^2$ such that
\[
|C(h^{(\gamma)})| \leq Cn^{-2}|C_\omega(h^{(\gamma)})|.
\]
Now note that the product
\[
\prod_{\beta \neq \gamma}^{m} C(h^{(\beta)}) \cdot C_\omega(h^{(\gamma)})
\] (3.38)
can be sorted into a product of at most $a(\gamma)$ cycles. Consequently, one has
\[
|C(h^{(\gamma)})| \leq O(n^{a(\gamma)-2}).
\] (3.39)

Now if $h^{(\gamma)}$ is in the second case of the definition of $\mathcal{B}_{21}^{(\gamma)}$, by using (iii) of Lemma 3.6 we can see that there exists a perfect matching $\omega \in L_{2a(\gamma)}^2$ such that
\[
|C(h^{(\gamma)})| \leq Cn^{-1}|C_\omega(h^{(\gamma)})|.
\]
However, in this case (3.38) can be sorted into a product of at most $a(\gamma) - 1$ cycles (after using merge operation) since four of $h^{(\gamma)}$ indices are identical. Therefore, we still have the bound (3.39). Now if $h^{(\gamma)}$ is in the third case of the definition of $\mathcal{B}_{21}^{(\gamma)}$, by using (ii) of Lemma 3.6 (set $h_1 \neq h_2$ therein) we can see that there exists a perfect matching $\omega \in L_{2a(\gamma)}^2$ such that
\[
|C(h^{(\gamma)})| \leq Cn^{-1}|C_\omega(h^{(\gamma)})|,
\]
where only one of 2-element cumulants in $C_\omega(h^{(\gamma)})$ has different underlying indices (i.e. equals $-1/(n-1)$). In this case, though (3.38) may can be sorted into a product of $a(\gamma)$ cycles, there is at least one cycle being in-homogeneous. Therefore, we obtain (3.39) again.
It remains to consider the case of $h^{(\gamma)} \in \mathcal{A}^{(\gamma)}_{22}$. In this case, by using (i) of Lemma 3.6, we only have the crude bound
\[
|C(h^{(\gamma)})| \leq C \sum_{\omega \in L^2_{2a(\gamma)}} |C_{\omega}(h^{(\gamma)})|.
\]
However, for any given $\omega \in L^2_{2a(\gamma)}$, it is not difficult to see that in this case we only have the following three possibilities. The first possible case is that (3.38) can be sorted into a product of $a(\gamma)$ cycles, while there are at least two in-homogeneous cycles. The second case is that (3.38) can be sorted into a product of only have the following three possibilities. The first possible case is that (3.38) can be sorted into a product of $a(\gamma) - 1$ cycles (after using merge operation), while there is at least one in-homogeneous cycle. The third case is that (3.38) can be sorted into a product of at most $a(\gamma) - 2$ cycles (after using merge operation). In any of these cases, we still have (3.39).

Therefore, we conclude the proof of Proposition 3.16 for the situation of $s(\gamma) = 0$.

Now if $s(\gamma) \geq 1$, we need to go back to (3.36). However, the discussion is just analogous to the case of $s(\gamma) = 0$. At first, we can fix a relationship matrix of $l^{(\gamma)}$ as $R_{l^{(\gamma)}} = M$ and consider the partial sum
\[
\sum_{\beta \neq \gamma} \prod_{\beta \neq \gamma} |C(h^{(\beta)})| \cdot C(h^{(\gamma)}),
\]
(3.40)
of (3.36) instead. Now we need to use an analogous discussion as that on $C(h^{(\gamma)})$ to bound $C(h^{(\gamma)})$ at first, then consider the product of sums over cycles. For example, we consider the case that $l_1^{(\gamma)} = l_2^{(\gamma)}$ while $l_i^{(\gamma)}$, $i = 2, \ldots, s(\gamma)$ are mutually distinct when $R_{l^{(\gamma)}} = M$, in addition we assume that $h^{(\gamma)}$ is in the first case of the definition of $\mathcal{A}^{(\gamma)}_{22}$ when $R_{h^{(\gamma)}} = M$. Then the summation (3.40) can be split into two parts. In the first parts, there is $\{l^{(\gamma)}\} \cap \{h^{(\gamma)}\} = \emptyset$. In the second parts, there is $\{l^{(\gamma)}\} \cap \{h^{(\gamma)}\} \neq \emptyset$. Then for the first part, we can employ (iii) of Lemma 3.6 again to see that there exists a perfect matching $\omega \in L^2_{2a(\gamma)}$ such that
\[
|C(h^{(\gamma)}_{\omega})| \leq C n^{-1} |C_{\omega}(h^{(\gamma)})|.
\]
However, in this case the total choice of $l^{(\gamma)}$ indices is at most of the order $O(n^{s(\gamma) - 1})$ since $l_1^{(\gamma)} \equiv l_2^{(\gamma)}$ when $R_{l^{(\gamma)}} = M$. In the second part, we know that there exists one perfect matching $\omega$ such that
\[
|C(h^{(\gamma)}_{\omega})| \leq C |C_{\omega}(h^{(\gamma)}_{\omega})|.
\]
However, if we fix $h^{(\gamma)}$ indices at first, the total choice of $l^{(\gamma)}$ indices is at most of the order $O(n^{s(\gamma) - 2})$ due to $l_1^{(\gamma)} \equiv l_2^{(\gamma)}$ and the coincidence condition between $l^{(\gamma)}$ and $h^{(\gamma)}$. Therefore, we can also get the bound
\[
(3.40) \leq O(n^{s(\gamma) - 2})
\]
by an analogous discussion on the product of sum over cycles.
Actually, the other cases are also analogous to the case of $s(\gamma) = 0$, we do not produce the detailed case-by-case discussion here and just leave them to the reader. Thus we conclude the proof. □

4. High order cumulants

Now with the aid of Corollary 3.4, Corollary 3.18, and Proposition 2.1 (Proposition 3.1 of [1]) we can prove the following lemma.

**Lemma 4.1** (High order cumulants). When $n \to \infty$, we have

$$C(\text{tr} S_{n}^{k_{1}}, \ldots, \text{tr} S_{n}^{k_{r}}) \to 0, \text{ for all } r \geq 3.$$

**Proof.** Now we recall the formula

$$C(\text{tr} S_{n}^{k_{1}}, \ldots, \text{tr} S_{n}^{k_{r}}) = \sum_{\pi \in L_{2k}^{\text{even}}} p^{-k+\# \pi_{0} \lor \pi} \sum_{j \in J} C_{\pi}(j)$$

Using Corollary 3.4, we see that for any given $\pi$ there exists

$$p^{-k+\# \pi_{0} \lor \pi} \sum_{j \in J} |C_{\pi}(j)| \leq Cn^{-k+\# \pi_{0} \lor \pi + \sum_{\gamma \geq 2}(\gamma - 2)n_{\gamma}(\pi)} \quad (4.1)$$

with some positive constant $C$. Noticing the fact that $\# \pi_{0} \lor \pi_{1} \lor \pi = 1$ and both $\pi_{0}$ and $\pi_{1}$ are perfect matchings, we can now employ Proposition 2.1 by setting $\Pi = \pi, \Pi_{0} = \pi_{0}$ and $\Pi_{1} = \pi_{1}$ therein. At first we use the bound

$$\# \pi_{0} \lor \pi + \# \pi_{1} \lor \pi \leq k + 1 - r/2. \quad (4.2)$$

Suppose that

$$1 + \sum_{\gamma \geq 2} (\gamma - 2)n_{\gamma}(\pi) - r/2 < 0, \quad (4.3)$$

by (4.1) we can get

$$p^{-k+\# \pi_{0} \lor \pi} \sum_{j \in J} |C_{\pi}(j)| = o(1)$$

immediately. It remains to consider the case of

$$1 + \sum_{\gamma \geq 2} (\gamma - 2)n_{\gamma}(\pi) - r/2 \geq 0. \quad (4.4)$$

In this case, we can use the bound

$$\# \pi_{0} \lor \pi + \# \pi_{1} \lor \pi \leq \# \pi + 1. \quad (4.5)$$

Note that since $\pi \in L_{2k}^{\text{even}}$, we have

$$\# \pi = \sum_{\gamma \geq 2} n_{\gamma}(\pi) + k - \sum_{\gamma \geq 2} \gamma n_{\gamma}(\pi) = k - \sum_{\gamma \geq 2}(\gamma - 1)n_{\gamma}(\pi). \quad (4.6)$$

Thus we also have

$$\# \pi_{0} \lor \pi + \# \pi_{1} \lor \pi \leq k + 1 - \sum_{\gamma \geq 2}(\gamma - 1)n_{\gamma}(\pi). \quad (4.7)$$
Consequently, by using (4.1) again, we have
\[ p^{-k + \pi_0} \sum_{j \in J} |C_{\pi}(j)| \leq C n^{1 - \sum_{r \geq 2} n_r(\pi)} \]  
(4.8)

Note that when \( r \geq 3 \), (4.4) implies that \( \sum_{\gamma \geq 2} (\gamma - 2) n_\gamma(\pi) \geq 1 \), which yields \( \sum_{\gamma \geq 2} n_\gamma(\pi) \geq 1 \). Now suppose that \( \sum_{\gamma \geq 2} n_\gamma(\pi) \geq 2 \), then by (4.8) we get
\[ p^{-k + \pi_0} \sum_{j \in J} |C_{\pi}(j)| = O(n^{-1}). \]  
(4.9)

Hence, it remains to consider the case of \( \sum_{\gamma \geq 2} n_\gamma(\pi) = 1 \). At first, note that if \( n_2(\pi) = 1 \), \( n_\gamma = 0 \) for \( \gamma \geq 3 \), (4.4) is impossible when \( r \geq 3 \). Thus it suffices to consider the case of
\[ \sum_{\gamma \geq 3} n_\gamma(\pi) = 1, \quad n_2(\pi) = 0. \]

Now we draw attention that in this case we can use the improved bound in Proposition 3.16, which yields
\[ \sum_{j \in J} C_{\pi}(j) \leq O(n^{\pi_1 + \sum_{\gamma \geq 3} (\gamma - 2) n_\gamma(\pi)}). \]  
(4.10)

Then in this special case, we have the improved bound
\[ p^{-k + \pi_0} \sum_{j \in J} C_{\pi}(j) \leq C n^{- \sum_{\gamma \geq 3} n_\gamma(\pi)} = O(n^{-1}). \]  
(4.11)

Consequently, we have shown that
\[ C(\text{tr}S_n^{k_1}, \ldots, \text{tr}S_n^{k_r}) = o(1), \quad r \geq 3. \]

Therefore, we conclude the proof. \( \square \)

5. Mean and covariance functions

In this section we prove (1.2) and (1.3). Before commencing the formal proof, we introduce some necessary notation and results on the population covariance matrix \( E Z^* Z \) at first.

5.1. On the population matrix \( E Z^* Z \). Let
\[ T := T_{n,n} = E Z^* Z = \begin{pmatrix} 1 & -\frac{1}{n-1} & \cdots & -\frac{1}{n-1} \\ -\frac{1}{n-1} & 1 & \cdots & -\frac{1}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n-1} & -\frac{1}{n-1} & \cdots & 1 \end{pmatrix}_{n \times n}. \]

Note that \( T \) has one multiple eigenvalue \( \frac{n}{n-1} \) with multiplicity \( n - 1 \) and one eigenvalue 0. Roughly speaking, our aim is to find some reference sample covariance matrix of the form \( \frac{1}{p} E \Xi^* \Xi \), where \( \Xi := (\xi_{ij})_{p \times n} \) is a random matrix with i.i.d. mean zero variance one entries, and then compare the mean and covariance functions of the spectral statistics of \( S_n \) to those of \( \frac{1}{p} E \Xi^* \Xi \). For the latter, we
can use the existing results from [2] and [11] to obtain the explicit formulae of the mean and covariance functions. To this end, we need to present some notions and properties on \( T \) at first. Now we denote the empirical spectral distribution of \( T \) by

\[
H_n(x) := \frac{n-1}{n} \mathbf{1}(x \geq \frac{n}{n-1}) + \frac{1}{n} \mathbf{1}(x \geq 0).
\]

Let \( c_n = n/p \) and \( m_n(z) : \mathbb{C}^+ \to \mathbb{C}^+ \) satisfy

\[
m_n(z) = \int \frac{1}{t(1-c_n - c_n zm_n(z)) - z} dH_n(t)
= \frac{n-1}{n} \left[ \frac{n}{n-1} \left( 1 - c_n - c_n zm_n(z) \right) \right]^{-1} - \frac{1}{nz}.
\]

Regarding \( m_n(z) \) as a Stieltjes transform, we can denote \( F_{c_n,H_n} \) as its corresponding distribution function. Moreover, we denote that \( H(x) = \mathbf{1}_{\{x \geq 1\}} \). Define \( m(z) : \mathbb{C}^+ \to \mathbb{C}^+ \) by the equation

\[
m(z) = [(1 - c - czm(z)) - z]^{-1}
\]

and set

\[
m(z) = -\frac{1}{z} + cm(z).
\]

In order to use the results in [11] (Theorem 1.4 therein), we need to verify the following lemma on \( T \).

**Lemma 5.1.** Under the above notation, for any fixed \( z, z_1, z_2 \in \mathbb{C}^+ \) we have

\[
\frac{1}{n} \sum_{i=1}^{n} e_i^* T^{1/2} (m(z_1)T + I)^{-1} T^{1/2} e_i e_i^* T^{1/2} (m(z_2)T + I)^{-1} T^{1/2} e_i \to (m(z_1) + 1)^{-1} (m(z_2) + 1)^{-1},
\]

and

\[
\frac{1}{n} \sum_{i=1}^{n} e_i^* T^{1/2} (m(z)T + I)^{-1} T^{1/2} e_i e_i^* T^{1/2} (m(z)T + I)^{-2} T^{1/2} e_i \to (m(z) + 1)^{-3}
\]

when \( n \to \infty \). Here \( e_i \) is the \( n \times 1 \) vector with a 1 in the \( i \)-th coordinate and 0’s elsewhere.

**Proof of Lemma 5.1.** Note that \( T_n \) has \( n-1 \) multiple eigenvalues \( \frac{n}{n-1} \) and one eigenvalue 0. Moreover, the eigenvector corresponding the 0 eigenvalue is \( \frac{1}{\sqrt{n}}(1, 1, \ldots, 1) \).

Thus we can write the spectral decomposition of \( T \) as

\[
T = (u_1, \ldots, u_n) \begin{pmatrix} 0 & \frac{n}{n-1} & \cdots & \frac{n}{n-1} \\ \frac{n}{n-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{n}{n-1} \\ \frac{n}{n-1} & \cdots & \frac{n}{n-1} & 0 \end{pmatrix} \begin{pmatrix} u_1^* \\ u_2^* \\ \vdots \\ u_n^* \end{pmatrix} := UAU^*.
\]
where
\[ u_1 = \frac{1}{\sqrt{n}}(1, 1, \ldots, 1)^* . \]

Thus for \( \alpha = 1, 2 \), we have
\[
e_i^* T^{1/2}(m(z)T + I)^{-\alpha} T^{1/2}e_i = e_i^* U^{1/2}(m(z)\Lambda + I)^{-\alpha} U^* e_i
= \sum_j \left( \Lambda^{1/2}(m(z)\Lambda + I)^{-\alpha} \Lambda^{1/2} \right)_{jj} |(U^* e_i)_j|^2 .
\]

Now note that
\[
\left( \Lambda^{1/2}(m(z)\Lambda + I)^{-\alpha} \Lambda^{1/2} \right)_{11} = 0 ,
\]
\[
\left( \Lambda^{1/2}(m(z)\Lambda + I)^{-\alpha} \Lambda^{1/2} \right)_{jj} = \frac{n}{n-1} \left( \frac{n}{n-1} m(z) + 1 \right)^{-\alpha} , \quad j \neq 1 .
\]
Moreover, we have \(|(U^* e_i)_1|^2 = \frac{1}{n}\) thus
\[
\sum_{j \neq 1} |(U^* e_i)_j|^2 = \frac{n-1}{n} .
\]
Then obviously, for each \( i \) and fixed \( z \in \mathbb{C}^+ \), we have
\[
e_i^* T^{1/2}(m(z)T + I)^{-\alpha} T^{1/2}e_i = \left( \frac{n}{n-1} m(z) + 1 \right)^{-\alpha} \rightarrow (m(z) + 1)^{-\alpha} ,
\]
which implies Lemma 5.1 immediately. \( \square \)

5.2. Mean function. At first we will pursue an argument analogous to that in Section 4 to discard the negligible terms by using Corollary 3.4, Corollary 3.18, and Proposition 2.1. And then we will evaluate the main terms by a two-step comparison strategy whose meaning will be clear later.

We commence with the negligible terms. Note that for the mean function, we have
\[ r = \#\pi_0 \lor \pi_1 = 1, \]
thus we can write
\[
\text{Etr}S_n^k = \sum_{\pi \in L_{2k}^n} p^{-k + \#\pi_0 \lor \pi} \sum_{j \in \mathcal{J}} C_\pi(j) . \tag{5.1}
\]

Analogous to the proof for high order cumulants, we use bound (4.1), (4.5) and (4.7) again. Note that when \( \sum_{\gamma \geq 2} n_\gamma(\pi) \geq 2 \), we can easily get from (4.8) that
\[ p^{-k + \#\pi_0 \lor \pi} \sum_{j \in \mathcal{J}} C_\pi(j) = O(n^{-1}) .
\]

Now we consider the case of \( \sum_{\gamma \geq 2} n_\gamma(\pi) = 1 \). Note that, if \( n_\gamma(\pi) = 1 \) for any \( \gamma \geq 3 \), we can use the improved bound in Corollary 3.18 again to obtain that
\[ p^{-k + \#\pi_0 \lor \pi} \sum_{j \in \mathcal{J}} C_\pi(j) = O(n^{-1}) .
\]
However, the case of $n_0(\pi) = 1, n_* = 0, \gamma \geq 3$ does have a non-negligible contribution to the expectation. Obviously, in this case, by (4.1) and (4.5) we see that only those terms with $\pi$ satisfying

$$#_\pi \lor n_0 + #_\pi \lor n_1 = k$$

have $O(1)$ contribution to the total sum. Now we recall the notation $L^2_{2k}$ and $L^4_{2k}$ defined in Section 2. We can write

$$\mathbb{E}r S^k_n = \sum_{\pi \in L^2_{2k}} p^{-k+\# n_0 \lor n} \sum_{j \in J} C_\pi(j)$$

$$+ \sum_{\pi \in L^4_{2k}} p^{-k+\# n_0 \lor n} \sum_{j \in J} C_\pi(j) + o(1).$$

We will not estimate the r.h.s. of (5.3) by bare-handed calculation and enumeration. Instead, we will adopt a comparison approach. To this end, we need to rely on existing results on sample covariance matrices. At first, we define a reference matrix. Let $\xi, \xi_j, j = 1, \ldots, n$ be i.i.d. symmetric random variables with common mean zero, variance 1 and fourth moment $\nu_4$. Let $V = (\xi_1, \ldots, \xi_n)$. Moreover, for any fixed positive integer $\ell$, we assume $E|\xi|^{\ell} \leq C_\ell$ for some positive constant $C_\ell$. Then we set $Y = (Y_1, \ldots, Y_n) := VT^{1/2}$ and let $V_i, i = 1, \ldots, p$ be i.i.d. copies of $V$. Now let $\Xi$ be the $p \times n$ matrix with $V_i$ as its $i$-th row and let

$$S_n(\xi) = \frac{1}{p} \Xi T \Xi^*.$$

Actually, if we take an analogous discussion on $S_n(\xi)$ as that on $S_n$ in the last sections, it is not difficult to see that as well as (5.3) there exists

$$\mathbb{E}r S^k_n(\xi) = \sum_{\pi \in L^2_{2k}} p^{-k+\# n_0 \lor n} \sum_{j \in J} C_\pi(j, \xi)$$

$$+ \sum_{\pi \in L^4_{2k}} p^{-k+\# n_0 \lor n} \sum_{j \in J} C_\pi(j, \xi) + o(1),$$

where $C_\pi(j, \xi)$ represents the quantity obtained by replacing $Z_i$ by $Y_i$ in the definition of $C_\pi(j)$. Particularly, when $\xi$ is gaussian, we will write $\mathbb{E}r S^k_n(\xi)$ and $C_\pi(j, \xi)$ as $\mathbb{E}r S^k_n(g)$ and $C_\pi(j, g)$ respectively. Actually, since $Y$ is just a linear transform of i.i.d. random sequence, the verification of Lemma 2.2, 3.5 and 3.6 for the vector $Y$ is much easier than that for $Z$. The proofs of these technical results for $Y$ are easily manipulated by invoking the properties P1-P3 of joint cumulants stated in Section 2. However, a more direct way to derive (5.4) is to check the property of joint cumulant summability for $Y$. We sketch it as follows. At first, it is elementary to obtain through calculation that the diagonal entries of $T^{1/2}$ are $t_d := \sqrt{(n-1)/n}$ and the off-diagonal entries are $t_o := -\sqrt{1/n(n-1)}$. Then $Y_i = t_d \xi_i + t_o \sum_{\ell \neq i} \xi_\ell$ by definition. Now let $r$ be a fixed positive integer. Suppose that in the collection of indices $\{j_1, \ldots, j_r\} \in [n]^r$, there are $r_1 \geq 0$ indices taking value of 1, and the remaining $r-r_1$ indices totally take $\alpha - 1$ distinct values with
multiplicities $r_\beta \geq 1$, $\beta = 2, \ldots, \alpha$ such that $\sum_{\beta=1}^{\alpha} r_\beta = r$. Then by symmetry of $Y$ and the properties P1-P3 of joint cumulant we have

$$C(Y_1, Y_{j_1}, \ldots, Y_{j_r}) = C(Y_{r_1+1}, Y_{r_2}, \ldots, Y_{r_\alpha})$$

$$= t_d^{1+1} \sum_{\alpha \neq 1} r_\alpha C(\xi_1, \ldots, \xi_{r+1}) + \sum_{\gamma=2}^{\alpha} t_o^{1+1} \sum_{\beta \neq \gamma} r_\beta C(\xi_\gamma, \ldots, \xi_{r+1}).$$

Obviously, the quantities

$$|C(\xi_\gamma, \ldots, \xi_{r+1})|, \gamma = 1, \ldots, \alpha$$

are all the same and can be bounded by some positive constant from above by invoking the assumption that $E|\xi|^\ell \leq C_\ell$ and the formula (2.2). Moreover, we observe that $\sum_{\beta \neq 1} r_\beta \geq \alpha - 1$ and $1 + \sum_{\beta \neq \gamma} r_\beta \geq \alpha - 1$ for all $\gamma = 2, \ldots, \alpha$. In addition, we have $t_d = O(1)$, $t_o = O(n^{-1})$. Therefore,

$$|C(Y_1, Y_{j_1}, \ldots, Y_{j_r})| = O(n^{-\alpha+1}).$$

Observe that $\alpha - 1$ is the number of distinct values except for 1 in the collection $\{j_1, \ldots, j_r\}$. Consequently, we have that

$$\sum_{j_1=1}^{n} \cdots \sum_{j_r=1}^{n} |C(Y_1, Y_{j_1}, \ldots, Y_{j_r})| = O(1),$$

which implies that the joint cumulant summability holds for $Y$. As explained above, we can get that the stronger bound

$$\sum_{j \in J} |C_{\pi}(j, \xi)| = O(n^{\# \pi_1 \lor \pi})$$

holds by using the result in [1]. With the aid of this stronger bound we can derive (5.4) by a routine discussion as above.

Obviously, we have

$$\sum_{\pi \in L_2^k} p^{-k + \# \pi_0 \lor \pi} \sum_{j \in J} C_{\pi}(j, \xi) = \sum_{\pi \in L_2^k} p^{-k + \# \pi_0 \lor \pi} \sum_{j \in J} C_{\pi}(j)$$

(5.5)

since this term only depends on the covariance structure $T$ which is shared by $Z$ and $Y$.

For the second term on the r.h.s. of (5.4), by Corollary 3.4 and (5.2) we see that it has a contribution of $O(1)$ at most. Hence, it suffices to estimate its leading term. To this end, we use (3.3) with $Z$ replaced by $Y$. Now we consider sum over $j | B_i$ with the block $B_i$ containing the unique 4-element block $A_i^{(\gamma)} \in \pi$. Obviously the sums over the indices with positions in the other blocks of $\pi \lor \pi_1$ are all in the case of (3.16). Without loss of generality, we can fix $i$ and $\gamma$ in the following argument. Recall the notation of paired index and unpaired index. Observe that $C\{Y_{\alpha_a}\}_{\alpha \in A_i^{(\gamma)}}$ may be in one of the following forms. When $m_i = 1$, $C\{Y_{\alpha_a}\}_{\alpha \in A_i^{(\gamma)}}$
must be in the form of \( C(Y_{i_1^{(c)}}, Y_{i_2^{(c)}}, Y_{i_3^{(c)}}, Y_{i_4^{(c)}}) \) (case 1). When \( m_i \geq 2 \), it may be in the form of \( C(Y_{h_1^{(c)}}, Y_{h_2^{(c)}}, Y_{h_3^{(c)}}, Y_{h_4^{(c)}}) \) (case 2) or \( C(Y_{h_1^{(c)}}, Y_{h_2^{(c)}}, Y_{h_3^{(c)}}, Y_{h_4^{(c)}}) \) (case 3). Then we have the following lemma.

**Lemma 5.2.** In any of the above three cases, for given \( \pi |_{B_i} \), there is a unique \( \sigma^{(\gamma)} \in L_4^2 \) such that

\[
\prod_{\beta \neq \gamma} C\{Y_{j_\alpha}\}_{\alpha \in A_i^{(\beta)}} \cdot C_{\sigma^{(\gamma)}}\{Y_{j_\beta}\}_{\alpha \in A_i^{(\gamma)}}
\]

is a product of two cycles. The other two perfect matchings in \( L_4^2 \) will drive the above product to be only one cycle.

**Proof.** In case 1, it is apparent to see that this \( \sigma^{(\gamma)} \) is the one satisfying

\[
C_{\sigma^{(\gamma)}}\{Y_{j_\alpha}\}_{\alpha \in A_i^{(\gamma)}} = C(Y_{i_1^{(c)}}, Y_{i_2^{(c)}}) \cdot C(Y_{i_3^{(c)}}, Y_{i_4^{(c)}}).
\]

And in case 2, \( \sigma^{(\gamma)} \) is the one satisfying

\[
C_{\sigma^{(\gamma)}}\{Y_{j_\alpha}\}_{\alpha \in A_i^{(\gamma)}} = C(Y_{h_1^{(c)}}, Y_{h_2^{(c)}}) \cdot C(Y_{h_3^{(c)}}, Y_{h_4^{(c)}}).
\]

Since \( \#\pi |_{B_i} \cdot \pi |_{B_i} = 1 \), obviously \( \sigma^{(\gamma)} \) defined above cases 1 and 2 will make (5.6) to be a product of two cycles. Now we come to see case 3, we define \( \sigma^{(\gamma)} \) in the following way. At first we can write \( C_{\pi |_{B_i}}(j |_{B_i}) \) (with \( Z \) replaced by \( Y \)) in the form of

\[
\prod_{\beta \neq \gamma} C(Y_{h_1^{(b)}}, Y_{h_2^{(b)}}) \cdot C(Y_{h_3^{(b)}}, Y_{h_4^{(b)}}).
\]

Now since \( j \) is \( \pi_1 \) measurable, we can start from \( h_1^{(\gamma)} \) and find out \( \beta_1 \in [m_i] \) such that \( h_1^{(\beta_1)} = h_1^{(\gamma)} \) (It is also possible that \( h_2^{(\beta_1)} = h_1^{(\gamma)} \). However, we can always switch the roles of \( h_1^{(\beta_1)} \) and \( h_2^{(\beta_1)} \) to simplify the argument.) Then we start with \( h_2^{(\beta_1)} \) and find out \( \beta_2 \in [m_i] \) such that \( h_2^{(\beta_2)} = h_2^{(\gamma)} \). Then repeatedly, we start with \( h_1^{(\beta_2)} \) and then find out the index which is equivalent to it and subsequently start with the index which is from the same 2-element cumulant with the one equivalent to \( h_1^{(\beta_2)} \). We proceed this procedure until we encounter an index in \( \{h_2^{(\gamma)}, h_3^{(\gamma)}, h_4^{(\gamma)}\} \). Without loss of generality, we assume it to be \( h_2^{(\gamma)} \). Then it is not difficult to see the \( \sigma^{(\gamma)} \) defined by

\[
C_{\sigma^{(\gamma)}}(Y_{h_1^{(\gamma)}}, Y_{h_2^{(\gamma)}}, Y_{h_3^{(\gamma)}}, Y_{h_4^{(\gamma)}}) = C(Y_{h_1^{(\gamma)}}, Y_{h_2^{(c)}}) \cdot C(Y_{h_3^{(c)}}, Y_{h_4^{(c)}})
\]

is the one which will make (5.6) be a product of two cycles. Moreover, it is not difficult to check that the other two perfect matchings in \( L_4^2 \) to be a single cycle. \( \square \)

Note that by Proposition 3.2, we have

\[
\sum_{j |_{B_i} \in J(\pi |_{B_i})} \prod_{\beta=1}^{m_i} |C\{Y_{j_\alpha}\}_{\alpha \in A_i^{(\beta)}}| \leq O(n).
\]

(5.7)
Our aim is to get the explicit $O(n)$ term of (5.7). To this end, we recall the discussions in Section 3. In any case of $C\{Y_{j_\alpha}\}_{\alpha \in A_i^{(\gamma)}}$, we only need to consider those $j$ such that $(j_\alpha)_{\alpha \in A_i^{(\gamma)}}$ is $\sigma^{(\gamma)}$-measurable. Since we can see that all the other $j|B_i$ can only make an $O(1)$ contribution totally to the l.h.s. of (5.7) by using Lemma 3.5 and the discussion on summations of in-homogeneous cycles in Section 3.3. Moreover, by the discussions in Section 3, we know that in any of the aforementioned three cases, the main contribution to the summation comes from the terms which can be decomposed into homogeneous cycles. In these terms, each 2-element cumulant is equal to 1, thus we have

$$\sum_{j|B_i \in J(\pi | B_i , n, b_i)} C\{Y_{j_\alpha}\}_{\alpha \in A_i^{(\gamma)}} = \sum_{\alpha_1, \alpha_2 = 1}^n C(Y_{\alpha_1}, Y_{\alpha_1}, Y_{\alpha_2}, Y_{\alpha_2}) + O(1)$$

$$= nC(Y_1, Y_1, Y_1, Y_1) + n(n - 1)C(Y_1, Y_1, Y_2, Y_2) + O(1)$$(5.8)

Note that (5.8) also holds if we replace $Y$ variables by corresponding $Z$ variables.

Now for the Gaussian case we claim

$$\sum_{\pi \in L_4^2} p^{-k + \#\pi \vee \pi} \sum_{j \in J} C_\pi(j, g) = o(1).$$ (5.9)

To see (5.9), it suffices to show that for any fixed $\pi \in L_4^2$, (5.7) can be strengthened to be

$$\sum_{j|B_i \in J(\pi | B_i , n, b_i)} \prod_{\beta = 1}^{m_i} |C\{Y_{j_\alpha}\}_{\alpha \in A_i^{(\gamma)}}| \leq O(1)$$ (5.10)

when $\xi$ is Gaussian. By (5.8), it suffices to evaluate the quantity $C(Y_{\alpha_1}, Y_{\alpha_1}, Y_{\alpha_2}, Y_{\alpha_2})$.

Note that when $\alpha_1 = \alpha_2$,

$$C(Y_{\alpha_1}, Y_{\alpha_1}, Y_{\alpha_2}, Y_{\alpha_2}) = C(Y_1, Y_1, Y_1, Y_1) = \nu_4 - 3 = 0$$ (5.11)

since $Y_1$ is gaussian. If $\alpha_1 \neq \alpha_2$, it is not difficult to get that

$$C(Y_{\alpha_1}, Y_{\alpha_1}, Y_{\alpha_2}, Y_{\alpha_2}) = C(Y_1, Y_1, Y_2, Y_2) = O(n^{-2})$$ (5.12)

by the definition of $Y_i$ and propositions P1-P3 of joint cumulant. Thus (5.10) holds, which directly implies that

$$\sum_{j \in J} C_\pi(j, g) = O(n^{\#\pi \vee \pi - 1}), \quad \pi \in L_4^2,$$

thus further yields (5.9) by combining (4.5) and the elementary fact that $\#\pi = k - 1$ for $\pi \in L_4^2$. Inserting (5.5) and (5.9) into (5.4) we obtain

$$\sum_{\pi \in L_4^2} p^{-k + \#\pi \vee \pi} \sum_{j \in J} C_\pi(j) = EtrS^k_n(g) + o(1).$$
Therefore, for the first term on the r.h.s. of (5.3), it suffices to estimate \( \mathbb{E} \text{tr} S^k_n(g) \). For the latter, we can use the result from [2] or [11] directly to write down
\[
\sum_{\pi \in L^2_{2k}} p^{-k+\# \pi_0} \sum_{j \in J} \mathbb{C}_\pi(j)
= n \int x^k dF_{c_n,H_n}(x) - \frac{1}{2\pi i} \int_C \frac{cz^k m^3(z)/(1 + m(z))^3}{(1 - cm^2(z))/(1 + m(z))^2} \, dz + o(1),
\]
where the contour \( C \) is taken to enclose the interval \([1 - \sqrt{c}^2, 1 + \sqrt{c}^2]\) as interior. See Theorem 1.4 of [11] for instance. However, here we can further simplify (5.13) by the property of orthogonal invariance of standard Gaussian vectors. Note that when \( \xi \) is Gaussian, we have
\[
\frac{1}{p} \mathbb{E} \bar{\xi} \xi^* = \frac{1}{p} \frac{n}{n - 1} GG^*
\]
where \( G := (g_{i,j})_{p,n-1} \) with i.i.d \( N(0,1) \) elements. Now let \( \tilde{c}_n = \frac{n - 1}{p} \) and \( F_{\tilde{c}_n,\text{MP}} \) be Marchenko-Pastur law (MP law) with parameter \( \tilde{c}_n \). Then by Theorem 1.4 of [11] and Lemma 5.1 we can rewrite (5.13) as
\[
\sum_{\pi \in L^2_{2k}} p^{-k+\# \pi_0} \sum_{j \in J} \mathbb{C}_\pi(j) = \frac{n^k}{(n - 1)^{k-1}} \int x^k dF_{\tilde{c}_n,\text{MP}}(x)
- \frac{1}{2\pi i} \int_C \frac{cz^k m^3(z)/(1 + m(z))^3}{(1 - cm^2(z))/(1 + m(z))^2} \, dz + o(1).
\]
Note that by (9.8.14) of [2], we see that
\[
- \frac{1}{2\pi i} \int_C \frac{cz^k m^3(z)/(1 + m(z))^3}{(1 - cm^2(z))/(1 + m(z))^2} = \frac{1}{4} \left( (1 - \sqrt{c})^{2k} + (1 + \sqrt{c})^{2k} \right) - \frac{1}{2} \sum_{j=0}^{k} \binom{k}{j}^2 c^j.
\]
Moreover, by the formula of moments of MP law (see Section 3.1.1 of [2] for instance) one can also get that
\[
\frac{n^k}{(n - 1)^{k-1}} \int x^k dF_{\tilde{c}_n,\text{MP}}(x) = \frac{n^k}{(n - 1)^{k-1}} \sum_{j=0}^{k-1} \frac{1}{j + 1} \binom{k}{j} \left( \frac{k - 1}{j} \right) \left( \frac{n - 1}{p} \right)^{j}.
\]
Now we come to estimate the second term on the r.h.s. of (5.3). Now we choose \( \xi \) satisfying \( \nu_4 \neq 3 \). Note that (5.11) is not valid now. However, (5.12) still holds. Then in this case,
\[
(5.8) = (\nu_4 - 3)n + O(1)
\]
Note that according to (3.3), except for this \( B_1 \) which containing the unique 4-element block of \( \pi \), the summation over the indices with positions in \([2k] \setminus B_1\) only depends on the covariance structure since \( \pi \in L^2_{2k} \). Now for \( C(Z_{l_1}, Z_{l_1}, Z_{l_2}, Z_{l_2}) \), we have
\[
C(Z_{l_1}, Z_{l_1}, Z_{l_2}, Z_{l_2}) = \mathbb{E} Z^4_{l_1} - 3 + o(1), \quad l_1 = l_2
\]
and
\[ C(Z_{l_1}, Z_{l_1}, Z_{l_2}, Z_{l_2}) = \frac{1}{n} (1 - EZ_1^T) + O(n^{-2}), \quad l_1 \neq l_2 \]
which can be checked easily by the distribution of \( Z \). Then we have
\[(5.8)|_{Y \to Z} = -2n + O(1),\]
where \((5.8)|_{Y \to Z}\) represents the quantity obtained by replacing \( Y \) by \( Z \) in (5.8). And all the other factors in (3.3) are the same as those of \( \sum_{j \in J} C_\pi (j, \xi) \) since they only depends on the covariance matrix \( T \). That means
\[ \sum_{\pi \in L_{2k}^4} p^{-k + \#\pi_0 \vee \pi} \sum_{j \in J} C_\pi (j) = -\frac{2}{n_4 - 3} \sum_{\pi \in L_{2k}^4} p^{-k + \#\pi_0 \vee \pi} \sum_{j \in J} C_\pi (j, \xi) + o(1). \]
By using [11] again (see Theorem 1.4 therein), we can get that
\[ \sum_{\pi \in L_{2k}^4} p^{-k + \#\pi_0 \vee \pi} \sum_{j \in J} C_\pi (j, \xi) = \frac{1}{\pi i} \int_C \frac{cz^k m(z) (m(z) + 1)^{-3}}{1 - cm^2(z)/(1 + m(z))^2} dz + o(1). \]
By (1.23) of [11], we obtain
\[
\frac{1}{\pi i} \int_C \frac{cz^k m(z) (m(z) + 1)^{-3}}{1 - cm^2(z)/(1 + m(z))^2} dz = 2c^{1+k} \sum_{j=0}^k \binom{k}{j} \left( \frac{1 - c}{c} \right)^j \binom{2k - j}{k - 1} - 2c^{1+k} \sum_{j=0}^k \binom{r}{j} \left( \frac{1 - c}{c} \right)^j \binom{2k + 1 - j}{k - 1}.
\]

In summary, we use Gaussian matrix as the reference matrix to obtain the value of the summation over \( \pi \in L_{2k}^4 \) and then use general matrix with \( \nu_4 \neq 3 \) as the reference one to obtain the value of the summation over \( \pi \in L_{2k}^4 \). We call such a comparison strategy as a \textit{two-step comparison strategy}.

5.3. \textbf{Covariance function.} Now we estimate the covariance function. Again we start with the formula
\[ C(\text{tr} S_{n_1}^{k_1}, \text{tr} S_{n_1}^{k_2}) = \sum_{\pi \in L_{2k}^2 \text{ s.t.} \#\pi_0 \vee \pi_1 \vee \pi = 1} p^{-k + \#\pi_0 \vee \pi} \sum_{j \in J} C_\pi (j). \]
Similar to the discussion in the last subsection, by using Corollaries 3.4 and 3.18 and (4.2) again we can see that is suffices to evaluate the contribution of the summation over the partitions \( \pi \) satisfying \( n_2(\pi) = 0 \) or 1 and \( n_\gamma(\pi) = 0 \) for all \( \gamma \geq 3 \). Moreover, by (4.1) and (4.2) it is easy to see that
\[ C(\text{tr} S_{n_1}^{k_1}, \text{tr} S_{n_1}^{k_2}) = O(1) \]
since \( \#\pi_0 \vee \pi + \#\pi_1 \vee \pi \leq k \) when \( r = 2 \). Now let
\[ L_{2k}^2 := \{ \pi \in L_{2k}^2 : \#\pi \vee \pi_1 \vee \pi_0 = 1 \}, \quad L_{2k}^4 := \{ \pi \in L_{2k}^4 : \#\pi \vee \pi_1 \vee \pi_0 = 1 \}. \]
For the explicit evaluation, we adopt the aforementioned \textit{two-step comparison strategy} again. We split the summation into the summations over \( L_{2k}^2 \) partitions
and $\hat{L}_k^j$ partitions. For the first part, we compare it with that of the Gaussian case. And for the second part, we compare it with the case of $\nu_3 \neq 3$. Then it is analogous to use Theorem 1.4 of [11] and Lemma 5.1 to obtain that

$$C(\text{tr} S_n^{k_1}, \text{tr} S_n^{k_2}) = -\frac{1}{2\pi^2} \int_{C_1} \int_{C_2} \frac{z_1^{k_1} z_2^{k_2}}{(m(z_1) - m(z_2))^2} m'(z_1) m'(z_2) dz_1 dz_2$$

$$+ \frac{c}{2\pi^2} \int_{C_1} \int_{C_2} z_1^{k_1} z_2^{k_2} \frac{d^2}{dz_1 dz_2} \left( \frac{m(z_1) m(z_2)}{(m(z_1) + 1)(m(z_2) + 1)} \right) dz_1 dz_2 + o(1),$$

where the contours $C_1$ and $C_2$ are disjoint and both enclose the interval $[(1 - \sqrt{\epsilon})^2, (1 + \sqrt{\epsilon})^2]$ as interior. Now by (9.8.15) of [2] and (1.24) of [11], we have

$$-\frac{1}{2\pi^2} \int_{C_1} \int_{C_2} \frac{z_1^{k_1} z_2^{k_2}}{(m(z_1) - m(z_2))^2} m'(z_1) m'(z_2) dz_1 dz_2$$

$$= 2c^{k_1+k_2} \sum_{j_1=0}^{k_1-1} \sum_{j_2=0}^{k_2} \binom{k_1}{j_1} \binom{k_2}{j_2} \left( \frac{1 - c}{c} \right)^{j_1+j_2}$$

$$\times \sum_{l=1}^{k_1-j_1} l \left( 2k_1 - 1 - j_1 + l \right) \left( 2k_2 - 1 - j_2 + l \right) \left( k_1 - 1 \right) \left( k_2 - 1 \right)$$

and

$$\frac{c}{2\pi^2} \int_{C_1} \int_{C_2} z_1^{k_1} z_2^{k_2} \frac{d^2}{dz_1 dz_2} \left( \frac{m(z_1) m(z_2)}{(m(z_1) + 1)(m(z_2) + 1)} \right) dz_1 dz_2$$

$$= -2c^{k_1+k_2+1} \sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} \binom{k_1}{j_1} \binom{k_2}{j_2} \left( \frac{1 - c}{c} \right)^{j_1+j_2} \left( 2k_1 - j_1 \right) \left( 2k_2 - j_2 \right) \left( k_1 - 1 \right) \left( k_2 - 1 \right).$$

Thus we conclude the proof of Theorem 1.1.

6. Application on Independence Test

In this section, as an application of Theorem 1.1 we will construct a nonparametric statistic to test complete independence for components of high dimensional random vectors. Our statistic is $W_7$ defined in (vii) below. We highlight it here at first.

$$W_7 = W_7(k, \delta) := \frac{\text{tr} S_n^k - \mathbb{E} \text{tr} S_n^k}{\sqrt{\text{Var}(G_k)}} + n^{-\delta} \left[ n \left( \max_{1 \leq i < j \leq p} \left| \frac{p}{n} s_{ij} \right| \right)^2 - 4 \log p + \log \log p \right],$$

where $\delta$ is a small positive constant to be chosen. We will experiment on several cases to see the performance of our statistic. Furthermore, we will also compare our method to other parametric or nonparametric methods in the literature. Let $R_n = (r_{ij})_{p \times p}$ be the Pearson’s sample correlation matrix of $n$ observations of a $p$ dimensional random vector $(Y_1, \ldots, Y_p)$, and denote its largest eigenvalue by $\lambda_{\max}(R)$. Assume that the $(i, j)$-th entry of $S_n$ is $s_{ij}$. We will experiment with the following seven statistics.
\[
W_1 = \frac{n\lambda_{\text{max}}(R) - (p^{1/2} + n^{1/2})^2}{(n^{1/2} + p^{1/2})(p^{1/2} + n^{1/2})^{1/3}} \quad (\text{see } [4] \text{ or } [12])
\]

\[
W_2 = \frac{\text{tr} S_n^k - \mathbb{E}\text{tr} S_n^k}{\sqrt{\text{Var}(G_k)}} \quad (\text{see Theorem } 1.1);
\]

\[
W_3 = \sum_{i=2}^p \sum_{j=1}^{i-1} r_{ij}^2 - p(p-1)/(2n) \quad (\frac{p}{n}) \quad (\text{see } [14])
\]

\[
W_4 = \frac{\log(|R_n|) - (p - n + 3/2)\log(1 - \frac{p}{n-1}) + (n - 2\frac{p}{n-1})}{\sqrt{-2[\frac{p}{n-1} + \log(1 - \frac{p}{n-1})]}} \quad (\text{see } [9])
\]

\[
W_5 = n \left( \max_{1 \leq i < j \leq p} |r_{ij}| \right)^2 - 4\log n + \log \log n \quad (\text{see } [8])
\]

\[
W_6 = n \left( \max_{1 \leq i < j \leq p} \left| \frac{p}{n} s_{ij} \right| \right)^2 - 4\log p + \log \log p \quad (\text{see } [20])
\]

\[
W_7 = W_2 + n^{-\delta} W_6, \quad \text{with some small constant } \delta > 0.
\]

Note that \(W_2\) is just the statistic in Theorem 1.1, \(W_7\) is a slight modification of \(W_2\) by adding a small penalty in terms of \(W_6\). Actually the statistic \(W_7\) combines two advantages of \(W_2\) and \(W_6\); thus is expected to have better performance. To see this, we take a \(p\) dimension Gaussian vector \(g\) with standard \(N(0,1)\) components as an example and assume its population covariance matrix to be \(\Sigma_g\). We can explain our construction of \(W_7\) through discussing two extreme cases. The first case is that \(\Sigma_g\) has only one nonzero off-diagonal entry which is significantly large. Then it is reasonable to expect a significantly large off-diagonal entry in the corresponding Spearman’s rank correlation matrix. Consequently, \(W_6\) will be sensitive to detect this type of dependence, since \(W_6\) only cares about the largest off-diagonal entry. However, \(W_2\) may be not so sensitive to detect this kind of dependence since the linear spectral statistics are relatively robust under the disturbance of a single entry of the population covariance matrix. By contrast, \(W_6\) will behave badly in detecting the case that \(\Sigma_g\) contains a lot of small nonzero off-diagonal entries. However, in this case, \(W_2\) may behave well since the limiting spectral statistics can sum up the effects of all these small nonzero off-diagonal entries. We remark here that in the high dimensional case one can not regard \(\Sigma_g\) to be approximately equal to identity if there are a lot of small off-diagonal entries.

Note that only \(W_2, W_5\) and \(W_7\) are non-parametric thus distribution free. Under \(H_0\), it has been shown in [20] that the limiting distribution of \(W_6\) has the c.d.f. of

\[
F_{W_6}(y) = \exp \left\{ -(8\pi)^{-1/2} e^{-y/2} \right\}.
\]

Obviously, the limiting distributions of \(W_2\) and \(W_7\) under \(H_0\) are both \(N(0,1)\). Moreover, for simplicity we only sketch the limiting results of \(W_1, W_3, W_4, W_5\) and the assumptions in their original articles below. The reader is suggested to
for $k$ small simulation. Intuitively, for simplicity of calculation, we always tend to choose large, the effect of adding penalty will be weak. Conversely, if $\delta$ is too small, the effect of adding penalty will be weak. Conversely, if $\delta$ is too small, the penalty term may affect the size under $\mathbf{H}_0$, especially when $n$ is not sufficiently large.

We will use Cauchy($\alpha, \beta$) to denote the Cauchy distribution with location parameter $\alpha$ and scale parameter $\beta$. And we will use $t(\gamma)$ to denote the Student’s $t$-distribution with degree of freedom $\gamma$.

With these, we consider three null hypotheses with the significance level $\alpha = 5\%$:

- $\mathbf{H}_{0,1}$: $Y_{ij}, j \in [n]$ are i.i.d. $N_p(\mathbf{0}, I_p)$ vectors;
- $\mathbf{H}_{0,2}$: $Y_{ij}, i \in [p], j \in [n]$ are i.i.d. Cauchy(0, 1) variables.;
- $\mathbf{H}_{0,3}$: $Y_{ij}, i \in [p], j \in [n]$ are i.i.d. Cauchy(0, 1) variables, $Y_{ij}, j \in [n]$ are standard Cauchy distributions, where $i_1 = 1, \ldots, \lfloor p/3 \rfloor$, $i_2 = \lfloor p/3 \rfloor + 1, \ldots, \lfloor 2p/3 \rfloor$, $i_3 = \lfloor 2p/3 \rfloor + 1, \ldots, p$ and $j \in [n]$.

For each null hypothesis, we consider two alternatives:

- $\mathbf{H}_{n,1,1}$ (one large disturbance): $Y_{ij}, j \in [n]$ are i.i.d. $N_p(\mathbf{0}, I_p + C)$ and $C = (c_{ik})_{p \times p}$ with $c_{ik} = 0, i, k \in [p]$, except $c_{12} = c_{21} = 0.8$.
- $\mathbf{H}_{n,1,2}$ (many small disturbances): $Y_{ij}, j \in [n]$ are i.i.d. $N_p(\mathbf{0}, I_p + D)$ and $D = (d_{ik})_{p \times p}$ with $d_{ik} = 4/p$ except $d_{ii} = 0, i, k \in [p]$.
- $\mathbf{H}_{n,2,1}$ (one large disturbance): $X_{ij}$ are i.i.d. Cauchy(0, 1). We set the observations $Y_{ij} = X_{ij} + 0.8X_{2j}, Y_{2j} = X_{2j} + 0.8X_{1j}$ and $Y_{ij} = X_{ij}$ for all $i = 3, \ldots, p$ and $j \in [n]$;
- $\mathbf{H}_{n,2,2}$ (many small disturbances): $X_{ij}$ are i.i.d. Cauchy(0, 1). We set the observations $Y_{ij} = X_{ij} + (7p)^{-1}\sum_{k \neq i} X_{kj}$ for $i \in [p]$ and $j \in [n]$;
- $\mathbf{H}_{n,3,1}$ (one large disturbance): the vector $(Y_{1,j}, \ldots, Y_{\lfloor p/3 \rfloor,j}, j \in [n]$ are i.i.d. $N_{\lfloor p/3 \rfloor}(0, I_{\lfloor p/3 \rfloor} + C')$ and $C' = (c_{ik})_{\lfloor p/3 \rfloor \times \lfloor p/3 \rfloor}$ with $c_{ik} = 0, i, k =
1, \ldots, \lfloor p/3 \rfloor$, except $c_{12} = c_{21} = 0.8$. Moreover, $Y_{i,j}, i = \lfloor p/3 \rfloor + 1, \ldots, n, j \in [n]$ are the same as those in $H_{0.3}$.

- $H_{a.3-2}$ (many small disturbances): the vector $(Y_{1,j}, \ldots, Y_{\lfloor p/3 \rfloor,j}), j \in [n]$ are i.i.d. $N(\lfloor p/3 \rfloor, I_{\lfloor p/3 \rfloor} + D')$ and $D' = (d_{ik})_{\lfloor p/3 \rfloor \times \lfloor p/3 \rfloor}$ with $d_{ik} = 12/p$ except $d_{ii} = 0, i, k = 1, \ldots, \lfloor p/3 \rfloor$. Moreover, $Y_{ij}, i = \lfloor p/3 \rfloor + 1, \ldots, n, j \in [n]$ are the same as those in $H_{0.3}$.

All of the results are shown in Table 1. When $(n, p) = (120, 160)$, the statistic $W_4$ cannot be evaluated and we omit it in Table 1. Moreover, in the cases of $H_{0.2}$ and $H_{0.3}$, the distribution assumption or the moment assumption is violated for the statistics $W_1, W_3, W_4$ and $W_5$. Thus, we also omit all of the values. The results are summarized as follows:

1. The sizes of $W_2, W_3, W_4$ and $W_7$ are close to the nominal size $5\%$. However, $W_3$ has the same size distortion in the case of $(n, p) = (120, 160)$. Meanwhile, the sizes of $W_1, W_5$ and $W_6$ trend to be smaller than $5\%$.

2. If the alternative hypotheses are set to have one large disturbance, $W_5, W_6$ and $W_7$ have better performance than the others since they consider the largest off-diagonal entry. Conversely, if the alternative hypotheses are set to have many small disturbances, $W_1, W_2, W_3, W_4$ and $W_7$ have better performance than $W_5$ and $W_6$.

Overall speaking, the sizes of statistic $W_7$ are close to the nominal level $\alpha = 5\%$. Also, $W_7$ produces the higher powers than others in most cases of the alternative hypotheses. Most of all, since $W_7$ is non-parametric, no distribution assumption is required for the variables.

**Acknowledgement**

The authors are very grateful to Prof. Greg W. Anderson for his answer to our question on Proposition 3.1 of [1].
Table 1. The sizes and powers (percentage) of $W_1$ to $W_7$ for different hypotheses, sample size $n$ and dimensions $p$.

| $(n,p)$ | $H_{0,1}$ | $H_{0,2}$ | $H_{0,3}$ |
|---------|-----------|-----------|-----------|
|         | $W_1$    | $W_2$    | $W_3$    | $W_4$    | $W_5$    | $W_6$    | $W_7$    | $W_2$    | $W_6$    | $W_7$    | $W_2$    | $W_6$    | $W_7$    |
| (60,40) | 0.4      | 4         | 5         | 4.9      | 2.1      | 3.2      | 4.5      | 4.7      | 1.9      | 4.9      | 4.5      | 2.0      | 4.8      |
| (120,80)| 1.4      | 5         | 5.6       | 4.9      | 2.6      | 2.2      | 5.7      | 4.6      | 2.6      | 5.4      | 3.7      | 3.7      | 4.4      |
| (60,10)| 0.4      | 3.6       | 2.1       | 3.1      | 3.8      | 3.4      | 4.1      | 3.6      | 2.2      | 4.2      | 3.9      | 3.6      | 4.2      |
| (120,160)| 1.7     | 5.8       | 10.9      | –        | 1.9      | 2.3      | 6.7      | 4        | 2.8      | 4.8      | 4.2      | 2.4      | 4.6      |

| $(n,p)$ | $H_{a,1}$ | $H_{a,2}$ | $H_{a,3}$ |
|---------|-----------|-----------|-----------|
|         | $W_1$    | $W_2$    | $W_3$    | $W_4$    | $W_5$    | $W_6$    | $W_7$    | $W_2$    | $W_6$    | $W_7$    | $W_2$    | $W_6$    | $W_7$    |
| (60,40) | 2.5      | 13.9     | 22.6     | 17.4     | 100      | 100      | 66.7     | 25.6     | 100      | 99.7     | 12.4     | 99.9     | 68.2     |
| (120,80)| 5.6      | 13.2     | 25       | 20.5     | 100      | 100      | 95.5     | 26.2     | 100      | 100      | 12.1     | 100      | 96      |
| (60,10)| 29.4     | 84.3     | 96.2     | 99.5     | 100      | 100      | 99.3     | 100      | 100      | 100      | 83.4     | 100      | 99.5     |
| (120,160)| 2.2     | 7.7      | 21.5     | –        | 100      | 100      | 89.6     | 11.4     | 100      | 100      | 6.5      | 100      | 88.6     |

| $(n,p)$ | $H_{a,1}$ | $H_{a,2}$ | $H_{a,3}$ |
|---------|-----------|-----------|-----------|
|         | $W_1$    | $W_2$    | $W_3$    | $W_4$    | $W_5$    | $W_6$    | $W_7$    | $W_2$    | $W_6$    | $W_7$    | $W_2$    | $W_6$    | $W_7$    |
| (60,40) | 99.9     | 99.8     | 99.7     | 67.5     | 15.6     | 18       | 99.8     | 98.2     | 28.7     | 97.8     | 99.9     | 65.4     | 99.9     |
| (120,80)| 100      | 100      | 100      | 72.4     | 11.8     | 12.3     | 100      | 100      | 68.7     | 100      | 100      | 39       | 100      |
| (60,10)| 100      | 100      | 100      | 100      | 99.8     | 99.5     | 100      | 91.7     | 45.6     | 91.4     | 100      | 100      | 100      |
| (120,160)| 100     | 99.3     | 99.2     | –        | 4        | 4.3      | 98.9     | 100      | 73.2     | 100      | 99.6     | 6.1      | 99.4     |

**APPENDIX**

In this Appendix, we prove Lemmas 2.2, 3.5 and 3.6.

**Proof of Lemma 2.2.** Below we denote

$$r_i := \sqrt{12 \left( \frac{n}{2} - 1 \right)} \left( i - \frac{n + 1}{2} \right), \quad i = 1, \ldots, n,$$

and set

$$T_\ell := \sum_{i=1}^{n} r_{\ell i}.$$  

Apparently, we have $T_\ell = 0$ if $\ell$ is odd and $T_\ell = O(n)$ if $\ell$ is even. Especially, we have $T_2 = n$. For (i) in Lemma 2.2, by definition we have

$$E(Z^{a_1} \cdots Z^{a_m}) = \frac{(n-m)!}{n!} \sum_{i_1} a_1^{a_1} \sum_{i_2 \neq i_1} a_2^{a_2} \cdots \sum_{i_m \neq i_{m-1}, \cdots, i_1} r_{*i_m}^{a_m}.$$  

Now we start from the last factor and use the fact

$$\sum_{i_m \neq i_{m-1}, \cdots, i_1} r_{*i_m}^{a_m} = T_{\alpha m} - (r_{i_1}^{a_1} + \cdots + r_{i_{m-1}}^{a_{m-1}}).$$
Let $\delta(j, i) = 1$ if $j = i$ and 0 otherwise. Then we have
\[
\mathbb{E}(Z_1^{\alpha_1} \cdots Z_m^{\alpha_m}) = \frac{(n-m)!}{n!} \left[ T_{\alpha_m} \sum_{i_1} r_{i_1}^{\alpha_1} \sum_{i_2 \neq i_1} r_{i_2}^{\alpha_2} \cdots \sum_{i_m-1 \neq i_{m-2} \cdots \neq i_1} r_{i_m-1}^{\alpha_{m-1}} - \sum_{j=1}^{m-1} \sum_{i_1} r_{i_1}^{\alpha_1 + \delta(j,1)\alpha_m} \sum_{i_2 \neq i_1} r_{i_2}^{\alpha_2 + \delta(j,2)\alpha_m} \cdots \sum_{i_m-1 \neq i_{m-2} \cdots \neq i_1} r_{i_m-1}^{\alpha_{m-1} + \delta(j,m-1)\alpha_m} \right] 
\]
\[
= \frac{1}{n-m+1} \left[ T_{\alpha_m} \mathbb{E}(Z_1^{\alpha_1} \cdots Z_{m-1}^{\alpha_{m-1}}) - \sum_{j=1}^{m-1} \mathbb{E}(Z_1^{\alpha_1 + \delta(j,1)\alpha_m} \cdots Z_{m-1}^{\alpha_{m-1} + \delta(j,m-1)\alpha_m}) \right].
\]
(6.3)

Now note that if $\alpha_m$ is odd, then $T_{\alpha_m} = 0$. Thus in this case only the second term in the above brackets needs to be considered. Moreover, for each $j \in [m-1],$
\[
\sum_{i=1}^{m-1} (\alpha_i + \delta(j, i)\alpha_m) = \sum_{i=1}^{m} \alpha_i.
\]
When $\alpha_m$ is even, we also have that
\[
\text{Parity of } \sum_{i=1}^{m-1} \alpha_i = \text{Parity of } \sum_{i=1}^{m} \alpha_i
\]
which is odd. Note that each term in the brackets on the r.h.s. of (6.3) is a multiple of an expectation of some product of $m - 1$ factors. That means for either odd or even $\alpha_m$, after the step as above to remove $Z_m$ factor we can get that the r.h.s. of (6.3) can be expressed in terms of expectations of some product of $m - 1$ factors $Z_1^{\alpha_1} \cdots Z_{m-1}^{\alpha_{m-1}}$ such that $\sum_{i=1}^{m-1} b_i$ has the same parity as $\sum_{i=1}^{m} \alpha_i$. Therefore, we can use an induction on $m$ to obtain (i) easily since the case of $m = 1$ is obvious.

To verify (ii), without loss of generality, we assume that $\alpha_1, \ldots, \alpha_{m-1}$ are even and $\alpha_m$ is odd for some $m_1 \leq m + 1$. We use (6.3) again. The case of $m_1 = m + 1$ is trivial, thus we can assume that $\alpha_m$ is odd. Then we have
\[
\mathbb{E}(Z_1^{\alpha_1} \cdots Z_m^{\alpha_m}) = \frac{-1}{n-m+1} \sum_{j=1}^{m-1} \mathbb{E}(Z_1^{\alpha_1 + \delta(j,1)\alpha_m} \cdots Z_{m-1}^{\alpha_{m-1} + \delta(j,m-1)\alpha_m}).
\]
(6.4)

Note that we can split the terms in the above sum into two classes: $j \leq m_1 - 1$ and $j \geq m_1$. When $j \leq m_1 - 1$, we see
\[
n_o(\alpha_1 + \delta(j,1)\alpha_m, \ldots, \alpha_{m-1} + \delta(j, m-1)\alpha_m) = n_o(\alpha_1, \ldots, \alpha_m).
\]
(6.5)
If $j \geq m_1$,
\[
n_o(\alpha_1 + \delta(j,1)\alpha_m, \ldots, \alpha_{m-1} + \delta(j, m-1)\alpha_m) = n_o(\alpha_1, \ldots, \alpha_m) - 2.
\]
(6.6)
Note that by symmetry, we can always switch the positions of $Z_i^{\alpha_i}$ and $Z_j^{\alpha_j}$ in (6.2). That means we can switch all $Z_i^{\alpha_i+\delta(j,i)\alpha_m}$ with odd $\alpha_i+\delta(j,i)\alpha_m$ to the end of the product. After this formal switching, we can conduct the calculation as in (6.4). Note that when $\sum \alpha_i$ is even, $n_o(\alpha)$ must be even. Then by using (6.4)-(6.6) we can get (2.8) by noticing that at each step like (6.4) we can gain an $O(n^{-1})$ factor along with a loss of 2 odd exponents at most (see (6.6)). □

**Proof of Lemma 3.5.** The proof of Lemma 3.5 is quite elementary by invoking (2.2). When $d(h_1, h_2, h_3, h_4) = 4$, we note that

$$C(Z_{h_1}, Z_{h_2}, Z_{h_3}, Z_{h_4}) = C(Z_1, Z_2, Z_3, Z_4)$$

$$= E(Z_1Z_2Z_3Z_4) - 3EZ_1Z_2EZ_3Z_4 = E(Z_1Z_2Z_3Z_4) - \frac{3}{n^2} + O(n^{-3}). \text{ (6.7)}$$

Moreover, using the notation $r_i$ defined in (6.1) we have

$$E(Z_1Z_2Z_3Z_4) = \frac{1}{n(n-1)(n-2)(n-3)} \sum_{i_1} r_{i_1} \sum_{i_2 \neq i_1} r_{i_2} \sum_{i_3 \neq i_1, i_2} r_{i_3} \sum_{i_4 \neq i_1, i_2, i_3} r_{i_4}. \text{ (6.8)}$$

Then it is elementary to check that

$$E(Z_1Z_2Z_3Z_4) = \frac{3}{n^2} + O(n^{-3})$$

by noting the facts that $\sum_{i=1}^n r_i^{2\alpha+1} = 0$ and $\sum_{i=1}^n r_i^{2\alpha} = O(n)$ for any fixed nonnegative integer $\alpha$. Thus we can get (3.7) by (6.7) and (6.8) when $d(h) = 4$. The other cases are similar and simpler, and we just leave them to the reader. □

**Proof of Lemma 3.6.** Let $\psi_{2\alpha-1} = \psi_{2\alpha} = l_\alpha$ for $\alpha = 1, \ldots, s$ and $\psi_{2s+\beta} = h_\beta$ for $\beta = 1, \ldots, 2t$. Now for any partition $\pi \in L_{2s+2t}$ and block $\hat{A} \in \pi$, we use the notation $(\hat{h})|_{\hat{A}} := (\psi_\alpha)_{\alpha \in \hat{A}}$. For example, assuming $s = 2, t = 1$, $\hat{A} = \{2, 3, 4, 6\}$, then $(\hat{h})|_{\hat{A}} = (l_1, l_2, l_3, h_2)$. Actually, the discussion below only depends on the collection of the coefficients of $(\hat{h})|_{\hat{A}}$ but not on their positions in the vector. Analogously, we can define $h|_B$ for any set $B \subset [2t]$. Now we prove three statements of Lemma 3.6 one by one.

For (i), recalling the formula (2.2) and (2.7), we have

$$C(\hat{h}) = \sum_{\pi \in P_{\text{sym}}^{2s+2t}} (-1)^{\#\pi-1}(\#\pi - 1)!E_\pi(\hat{h}).$$

Now for any vector of indices $v$ let

$$m_o(v) = \#\{\text{components appearing odd times in } v\}.$$

For example, for $v = (1, 1, 2, 2, 2, 3, 4, 5, 6)$, one has

$$m_o(v) = \#\{2, 3, 5, 6\} = 4.$$

In this manner we can define $m_o((\hat{h}))$ and $m_o(h)$. Then it is not difficult to see

$$m_o((\hat{h})) = m_o(h) \text{ (6.9)}$$
since l indices are paired. Moreover, it is also apparent that
\[ \sum_{A \in \pi} m_o((\hat{h}|_A) \geq m_o((\hat{h})) \] (6.10)
for all \( \pi \in L_{2s+2t}^{\text{even}} \). Here we use \( m_o((\hat{h}|_A) \) to denote the number of the components of \((\hat{h}|_A) \) appearing odd times. Moreover, it is not difficult to check
\[ m_o(h) = \min_{\sigma \in L_{2t}^2} \sum_{B \in \sigma} m_o(h|_{B}) \] (6.11)
To see (6.11), we can find out any pair of coincident components in \( h \) if there exists, then we put them into one block and cancel them in the vector \( h \). Then we proceed to do it until each component in the vector is single. Finally, we can decompose the obtained vector with unpaired components by any perfect matching. Then all the blocks in the above procedure form a natural perfect matching in \( L_{2t}^2 \), which is obvious a minimizer of \( \sum_{B \in \sigma} m_o(h|_{B}) \). Moreover, (6.11) is easy to see from the above construction.

Then by (2.8), (6.9) and (6.10) it is obvious that for any \( \pi \in L_{2s+2t}^{\text{even}} \),
\[ |E_{\pi}((\hat{h}))| \leq O(n^{-m_o(h)/2}). \] (6.12)
Moreover, apparently by (1.1), (2.8) and (6.11) we have
\[ \sum_{\sigma \in L_{2t}^2} |E_{\sigma}(h)| = C n^{-m_o(h)/2}. \] (6.13)
Then (i) follows from (6.12) and (6.13).

Now we check (ii). At first we consider those \( \pi \in L_{2s+2t}^{\text{even}} \) in which there exists \( i \in \{1, \ldots, s\} \) such that two \( l_i \)'s are not in the same block or \( h_1, h_2 \) are not in the same block. By (2.8) and the assumption on l and h, it is easy to see that for these \( \pi \) there exists
\[ E_{\pi}((\hat{h})) = O(n^{-2}). \]

Now according to (2.2), it suffices to consider the sum over those \( \pi \) in which each pair of \( l_i, i = 1, \ldots, s \) is in the same block of \( \pi \), so is the pair \( \{h_1, h_2\} \). Thus by (2.2), we have
\[ C((\hat{h})) = C(Z_{l_1}^2, \ldots, Z_{l_s}^2, Z_{h_1} Z_{h_2}) + O(n^{-2}). \]
Therefore, it suffices to show that
\[ C(Z_{l_1}^2, \ldots, Z_{l_s}^2, Z_{h_1} Z_{h_2}) = O(n^{-2}). \] (6.14)
To verify (6.14), we split the issue into two cases: \( h_1 = h_2 \) and \( h_1 \neq h_2 \). For the first case, it suffices to prove
\[ C(Z_{l_1}^2, \ldots, Z_{l_s}^2) = O(n^{-2}), \quad s \geq 3 \] (6.15)
when \( l_1, \ldots, l_s \) are mutually distinct. By formula (2.2), we see that
\[ C(Z_{l_1}^2, \ldots, Z_{l_s}^2) = \sum_{\pi \in L_s} (-1)^{\#\pi-1}(\#\pi - 1)!E_{\pi}(Z_{l_1}^2, \ldots, Z_{l_s}^2). \] (6.16)
Now let $t_\nu(\pi)$ be the cardinality of the $\nu$-th block of the partition $\pi$ for $\nu = 1, \ldots, \#\pi$. We do not need to specify any order of the blocks of $\pi$, since the discussion below only depends on the set $\{t_1(\pi), \ldots, t_{\#\pi}(\pi)\}$ itself but not on the order of its components. Let

$$t^+_\nu(\pi) := \frac{t_\nu(\pi)(t_\nu(\pi) - 1)}{2} = 1 + 2 + \cdots + (t_\nu(\pi) - 1).$$

Then we have the following lemma.

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

$$C(Z^2_{l_1}, \ldots, Z^2_{l_s}) = \sum_{\pi \in L_s} (-1)^{\#\pi - 1} (\#\pi - 1)!$$

$$+ \frac{1}{n} \sum_{\pi \in L_s} (-1)^{\#\pi - 1} (\#\pi - 1)! \sum_{\nu=1}^{\#\pi} t^+_\nu(\pi) (1 - EZ^2_1) + O(n^{-2}).$$

The proof of Lemma 6.1 will be stated later. At this stage we just admit it and proceed to prove Lemma 3.6. Hence, to show (6.15), it suffices to verify

$$\sum_{\pi \in L_s} (-1)^{\#\pi - 1} (\#\pi - 1)! = 0 \quad (6.17)$$

and

$$\sum_{\pi \in L_s} (-1)^{\#\pi - 1} (\#\pi - 1)! \sum_{\nu=1}^{\#\pi} t^+_\nu(\pi) = 0. \quad (6.18)$$

To this end, we study the coefficients of the expansion formula for the following cumulant

$$C(1 + \Delta, \ldots, 1 + \Delta), \quad (6.19)$$

where $\Delta := \Delta_n$ is a mean zero real random variable with

$$\mathbb{E}|\Delta|^k = C_k n^{-\frac{k}{2}}$$

for all fixed nonnegative integer $k$ and some positive constant $C_k$ depending on $k$. Now by shift invariance of joint cumulant, we know

$$C(1 + \Delta, \ldots, 1 + \Delta) = C(\Delta, \ldots, \Delta) = O(n^{-\frac{7}{2}}) \quad (6.20)$$

on one hand. On the other hand, we can also expand (6.19) by (2.2). Then it is easy to find that the l.h.s. of (6.17) is the constant term in the expansion formula of (6.19) and the l.h.s. of (6.18) is just the coefficient of $\mathbb{E}\Delta^2$. Now since $s \geq 3$, by (6.20) we know that (6.17) and (6.18) hold, which implies (6.14) in the case of $h_1 = h_2$.

Now we deal with the case of $h_1 \neq h_2$. In this case it is easy to see that

$$C(Z^2_{l_1}, \ldots, Z^2_{l_s}, Z_{h_1}Z_{h_2})$$
\[\begin{align*}
&= \sum_{\pi \in L_{s+1}} (-1)^{\#\pi - 1} (\#\pi - 1)!E_\pi(Z_{l_1}^2, \ldots, Z_{l_{s+1}}^2, Z_{h_1}Z_{h_2}) \\
&= \sum_{\pi \in L_{s+1}} (-1)^{\#\pi - 1} (\#\pi - 1)! \prod_{i=1}^{s} \mathbb{E}Z_i^2 \mathbb{E}Z_{h_1}Z_{h_2} + O(n^{-2})
\end{align*}\]

which follows from the fact that when \(l_i, i = 1, \ldots, s\) are mutually distinct and distinct from \(h_1, h_2\),

\[\mathbb{E}_\pi(Z_{l_1}^2, \ldots, Z_{l_s}^2, Z_{h_1}Z_{h_2}) = \prod_{i=1}^{s} \mathbb{E}Z_i^2 \mathbb{E}Z_{h_1}Z_{h_2}(1 + O(n^{-1})) \quad (6.21)\]

for any \(\pi \in L_{s+1}\). (6.21) can be verified easily and directly via the distribution of \(Z\). Actually the proof of (6.21) is analogous to and simpler than the proof of (6.32) below, we just leave it to the reader. Now by using (6.17) again we can get (6.14).

It remains to show (iii). Similarly, we split the discussion into two cases: \(h_1 = h_2\) and \(h_1 \neq h_2\). At first, we check the first case, it suffices to show that

\[|\text{C}(\hat{\pi})| \leq O(n^{-1}).\]

Let \(L_{2s+2}^0 \subseteq L_{2s+2}^{\text{even}}\) consist of the partitions such that each pair of \(l_i: l = 3, \ldots, s\) is in the same block and \(h_1, h_2\) are in the same block. Analogous to the discussions above, it is not difficult to see that

\[|\text{C}(\hat{\pi})| = \sum_{\pi \in L_{2s+2}^{\text{even}}} (-1)^{\#\pi - 1} (\#\pi - 1)!E_\pi(\hat{\pi}) + O(n^{-2}). \quad (6.22)\]

Now we further split the partitions in \(L_{2s+2}^0\) into two types: \(L_{2s+2}^{0,1}\) and \(L_{2s+2}^{0,2}\), where \(L_{2s+2}^{0,2}\) consists of all the partitions in \(L_{2s+2}^0\) such that four \(h_1(= h_2)\) are in the same block. Otherwise, we say \(\pi \in L_{2s+2}^{0,1}\). In other words, if \(\pi \in L_{2s+2}^{0,1}\), four \(l_1\)'s are separated into two pairs and assigned in two different blocks. We write

\[\begin{align*}
&\sum_{\pi \in L_{2s+2}^{0,1}} (-1)^{\#\pi - 1} (\#\pi - 1)!E_\pi(\hat{\pi}) \\
&= \sum_{\pi \in L_{2s+2}^{0,1}} (-1)^{\#\pi - 1} (\#\pi - 1)!E_\pi(\hat{\pi}) \\
&+ \sum_{\pi \in L_{2s+2}^{0,2}} (-1)^{\#\pi - 1} (\#\pi - 1)!E_\pi(\hat{\pi}). \quad (6.23)
\end{align*}\]

Since here \(h_1 = h_2\), we can write \(h_1\) and \(h_2\) as \(l_{s+1}\). In this manner, we have

\[\begin{align*}
\sum_{\pi \in L_{2s+2}^{0,1}} (-1)^{\#\pi - 1} (\#\pi - 1)!E_\pi(\hat{\pi}) &= \text{C}(Z_{l_1}^4, Z_{l_3}^2, \ldots, Z_{l_{s+1}}^2) \\
&= \sum_{\pi \in L_s} (-1)^{\#\pi - 1} (\#\pi - 1)!E_\pi(Z_{l_1}^4, Z_{l_3}^2, \ldots, Z_{l_{s+1}}^2). \quad (6.24)
\end{align*}\]
Now we claim
\[ \sum_{\pi \in L_{2s+2}^1} (-1)^{\#\pi-1} (\#\pi - 1)! = 0 \]  \hspace{1cm} (6.25)
and
\[ \sum_{\pi \in L_{2s+2}^2} (-1)^{\#\pi-1} (\#\pi - 1)! = 0. \]  \hspace{1cm} (6.26)

Note that by (6.24) and (6.17), we know
\[ \sum_{\pi \in L_{2s+2}^1} (-1)^{\#\pi-1} (\#\pi - 1)! = \sum_{\pi \in L_{2s+2}^2} (-1)^{\#\pi-1} (\#\pi - 1)! = 0. \]  \hspace{1cm} (6.27)

Hence, it suffices to verify either (6.25) or (6.26). We choose to show (6.25) below.

Now let \( \xi_i, i = 1, \ldots, s \) be i.i.d Bernoulli variables with mean zero and variance 1. Observe that
\[ \sum_{\pi \in L_{2s+2}^1} (-1)^{\#\pi-1} (\#\pi - 1)! = \sum_{\pi \in L_{2s+2}^2} (-1)^{\#\pi-1} (\#\pi - 1)! = 0. \]  \hspace{1cm} (6.28)

Hence, it suffices to verify either (6.25) or (6.26). We choose to show (6.25) below.

Analogous to (6.21), for any fixed positive integers \( r \) and \( \ell_1, \ldots, \ell_r \), it is elementary to check that
\[ \mathbb{E} \prod_{i=1}^{r} Z_i^{2\ell_i} = \prod_{i=1}^{r} \mathbb{E} Z_i^{2\ell_i} + O(n^{-1}). \]

Hence one has
\[ \mathbb{E}_\pi(\hat{h}) = \mathbb{E} Z_1^{s+1} \prod_{i=3}^{r} \mathbb{E} Z_i^{2\ell_i} + O(n^{-1}), \]  \hspace{1cm} (6.28)
and
\[ \mathbb{E}_\pi(\hat{h}) = (\mathbb{E} Z_i^2)^2 \prod_{i=3}^{s+1} \mathbb{E} Z_i^2 + O(n^{-1}), \]  \hspace{1cm} (6.29)

(6.28) and (6.29) together with (6.22)-(6.26) imply that (iii) holds in the case of \( h_1 = h_2 \).
Now note that (6.22) and (6.23) hold in the case of \( h_1 \neq h_2 \) as well. Thus by (6.26) and (6.27) it suffices to show for any \( \pi \in L_{2s+1}^2 \), there exists

\[
    \mathbb{E}_{\pi}(\hat{h}) = \mathbb{E} Z_1^2 \prod_{l=2}^{s-1} \mathbb{E} Z_l^2 \mathbb{E}(Z_s Z_{s+1}) + O(n^{-2}),
\]

and for \( \pi \in L_{2s+2}^2 \) there exists

\[
    \mathbb{E}_{\pi}(\hat{h}) = (\mathbb{E} Z_1^2) \prod_{l=2}^{s-1} \mathbb{E} Z_l^2 \mathbb{E}(Z_s Z_{s+1}) + O(n^{-2}).
\]

The proofs of (6.30) and (6.31) are also quite elementary and analogous to that for (6.32) below, we just leave them to the reader. Therefore, we conclude the proof. \( \square \)

Now we prove Lemma 6.1

**Proof of Lemma 6.1.** By (1.1) and (6.16), it suffices to verify for \( s \geq 1 \)

\[
    \mathbb{E}(\prod_{l=1}^{s} Z_l^2) = \prod_{l=1}^{s} \mathbb{E} Z_l^2 + \frac{1}{n} \frac{s(s-1)}{2} \left[ \prod_{l=1}^{s} \mathbb{E} Z_l^2 - \mathbb{E}^4 Z_1^2 \prod_{l=2}^{s} \mathbb{E} Z_l^2 \right] + O(n^{-2}).
\]

To this end, we start with the definition

\[
    \mathbb{E}(\prod_{l=1}^{s} Z_l^2) = \frac{(n-s)!}{n!} \left[ T_2 \sum_{i_1} r_{i_1}^2 \sum_{i_2 \neq i_1} r_{i_2}^2 \cdots \sum_{i_{s-1} \neq i_s-2,\ldots,i_1} r_{i_{s-1}}^2
\]

\[
- \sum_{j=1}^{s-1} \sum_{i_1} r_{i_1}^{2(1+\delta(j,i))} \sum_{i_2 \neq i_1} r_{i_2}^{2(1+\delta(j,i))} \cdots \sum_{i_{s-1} \neq i_{s-2},\ldots,i_1} r_{i_{s-1}}^{2(1+\delta(j,i))} \right].
\]

Note that by symmetry, for any \( j \in [s-1] \), we have

\[
    \frac{(n-s+1)!}{n!} \left[ \sum_{i_1} r_{i_1}^{2(1+\delta(j,i))} \sum_{i_2 \neq i_1} r_{i_2}^{2(1+\delta(j,i))} \cdots \sum_{i_{s-1} \neq i_{s-2},\ldots,i_1} r_{i_{s-1}}^{2(1+\delta(j,i))} \right] = \mathbb{E}(Z_1^4 Z_2^2 \cdots Z_{s-1}^2).
\]

By taking the fact that \( T_2 = n \) into account, we have

\[
    \mathbb{E}(\prod_{l=1}^{s} Z_l^2) = \frac{n}{n-s+1} \mathbb{E}(Z_1^4 Z_2^2 \cdots Z_{s-1}^2) \mathbb{E} Z_s^2 - \frac{s-1}{n-s+1} \mathbb{E}(Z_1^4 Z_2^2 \cdots Z_{s-1}^2)
\]

\[
= \mathbb{E}(\prod_{l=1}^{s-1} Z_l^2) \cdot \mathbb{E} Z_s^2.
\]
\[ + \frac{s-1}{n} \left[ \mathbb{E} \left( \prod_{l=1}^{s-1} Z_l^2 \right) \cdot \mathbb{E} Z_s^2 - \mathbb{E} \left( \prod_{l=2}^{s-1} Z_l^2 \right) \right] + O(n^{-2}) \quad (6.33) \]

from which we see that
\[ \mathbb{E} \left( \prod_{l=1}^{s} Z_l^2 \right) = \mathbb{E} \left( \prod_{l=1}^{s-1} Z_l^2 \right) \cdot \mathbb{E} Z_s^2 + O(n^{-1}). \]

Recursively, one has
\[ \mathbb{E} \left( \prod_{l=1}^{s} Z_l^2 \right) = \prod_{l=1}^{s} \mathbb{E} Z_l^2 + O(n^{-1}). \quad (6.34) \]

Analogously, we can also get through calculation that
\[ \mathbb{E} (Z_1^4 Z_2^2 \cdots Z_s^2) = \mathbb{E} Z_1^4 \prod_{l=2}^{s} \mathbb{E} Z_l^2 + O(n^{-1}). \quad (6.35) \]

Substituting (6.34) and (6.35) into (6.33) we have
\[ \mathbb{E} \left( \prod_{l=1}^{s} Z_l^2 \right) = \mathbb{E} \left( \prod_{l=1}^{s-1} Z_l^2 \right) \cdot \mathbb{E} Z_s^2 + \frac{s-1}{n} \left[ \prod_{l=1}^{s} \mathbb{E} Z_l^2 - \mathbb{E} Z_1^4 \prod_{l=2}^{s-1} \mathbb{E} Z_l^2 \right] + O(n^{-2}). \]

Now we proceed to expand \( \mathbb{E} (\prod_{l=1}^{s-1} Z_l^2) \) on the r.h.s. of the above equation as that in (6.33). Recursively, we can finally get (6.32). Thus we conclude the proof. \( \square \)

**References**

[1] G.W. Anderson, O. Zeitouni. A CLT for regularized sample covariance matrices. Ann. Stat. **36** (6): 2553-2576. (2008)
[2] Z.D. Bai and J.W. Silverstein. Spectral analysis of large dimensional random matrices. Science Press, 2nd ed., Beijing. (2006)
[3] Z.D. Bai and W. Zhou. Large sample covariance matrices without independence structures in columns. Statistica Sinica **18**, 425-442. (2008)
[4] Z. G. Bao, G. M. Pan, and W. Zhou. Tracy-Widom law for the extreme eigenvalues of sample correlation matrices. Electron. J. Probab., **17**, 1-32. (2012)
[5] A. Bose and A. Sen. Another look at the moment method for large dimensional random matrices. Electron. J. Probab. **13**(21), 588628. (2008)
[6] D.R. Brillinger. Time series. Classics in Applied Mathematics 36. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Data analysis and theory, Reprint of the 1981 edition. (2001)
[7] O. Costin and J. Lebowitz. Gaussian fluctuations in random matrices. Phys. Rev. Lett. **75**(1): 6972. (1995)
[8] T. Jiang. The asymptotic distributions of the largest entries of sample correlation matrices. Ann. Appl. Probab. **14**: 865-880. (2004)
[9] T. Jiang and F. Yang. Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. Ann. Statist. 41(4), 2029-2074. (2013)

[10] O. Ledoit and M. Wolf. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. Ann. Stat. 30, 1081-1102. (2002)

[11] G.M. Pan, W. Zhou. Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. Ann. Applied Probab. 18(3): 12321270. (2008)

[12] N. S. Pillai, J. Yin, Universality of correlation matrices. Ann. Statist., 40(3), 1737-1763. (2012)

[13] M. Rosenblatt. Stationary sequences and random fields. Birkhäuser Boston Inc., Boston, MA. (1998)

[14] J.R. Schott. Testing for complete independence in high dimensions. Biometrika 92, 951-956. (2005)

[15] Ya. Sinai, A. Soshnikov. Central limit theorem for traces of large random symmetric matrices with independent matrix elements. Bol. Soc. Bras. Mat. 29: 1-24. (1998)

[16] A. Soshnikov. Gaussian Limit for Determinantal Random Point Fields. Ann. Probab. 30(1): 171-187. (2002)

[17] E.P. Wigner. Characteristic Vectors of Bordered Matrices With Infinite Dimensions. Ann. Math. 62 (3): 548-564. (1955)

[18] W. B. Wu, X. Shao. Limit theorems for iterated random functions. J. Appl. Probab. 41 425-436. (2004)

[19] H. Xiao, W.B. Wu. Asymptotic Inference of Autocovariances of Stationary Processes. arXiv:1105.3423.

[20] W, Zhou. Asymptotic distribution of the largest off-diagonal entry of correlation matrices. Transactions of the American Mathematical Society 359, 5345-5363. (2007)