On eigenvalues of a high-dimensional Kendall’s rank correlation matrix with dependence

Zeng Li¹, Cheng Wang²,³,* & Qinwen Wang⁴

¹Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen 518055, China;
²School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China;
³Key Laboratory of Scientific and Engineering Computing (Shanghai Jiao Tong University), Shanghai 200240, China;
⁴School of Data Science, Fudan University, Shanghai 200433, China

Email: liz9@sustech.edu.cn, chengwang@sjtu.edu.cn, wqw@fudan.edu.cn

Received March 22, 2022; accepted August 30, 2022; published online September 14, 2023

Abstract In this paper, we investigate the limiting spectral distribution of a high-dimensional Kendall’s rank correlation matrix. The underlying population is allowed to have a general dependence structure. The result no longer follows the generalized Marcenko-Pastur law, which is brand new. It is the first result on rank correlation matrices with dependence. As applications, we study Kendall’s rank correlation matrix for multivariate normal distributions with a general covariance matrix. From these results, we further gain insights into Kendall’s rank correlation matrix and its connections with the sample covariance/correlation matrix.

Keywords Hoeffding decomposition, Kendall’s rank correlation matrix, limiting spectral distribution, random matrix theory

MSC(2020) 62H12, 62G35, 60F05

Citation: Li Z, Wang C, Wang Q W. On eigenvalues of a high-dimensional Kendall’s rank correlation matrix with dependence. Sci China Math, 2023, 66: 2615–2640, https://doi.org/10.1007/s11425-022-2031-2

1 Introduction

Covariance and correlation matrices play a vital role in multivariate statistical analysis because they provide the most direct way to characterize the relation between different variables. Many statistical estimation or inference methods involve the covariance/correlation matrix, such as principal component analysis, multivariate analysis of variance and factor analysis. In high-dimensional data analysis, studying the eigenvalues and eigenvectors of such covariance/correlation matrices is a fundamental problem.

In the random matrix theory, the sample covariance matrix has been thoroughly studied in the past decades. For an $n \times n$ Hermitian matrix $H_n$, the empirical spectral distribution (ESD) of $H_n$ is defined as

$$F^{H_n}(x) = \frac{1}{n} \sum_{i=1}^{n} I(\lambda_i \leq x),$$

where $\lambda_1, \ldots, \lambda_n$ are eigenvalues of $H_n$ and $I(\cdot)$ is the indicator function. If $F^{H_n}$ converges to a deterministic distribution function $F$, then $F(x)$ is called the limiting spectral distribution (LSD) of $H_n$. Marčenko and Pastur [27] first derived the LSD of the sample covariance
matrix. Furthermore, Bai and Silverstein [2] studied the central limit theorem (CLT) for its linear spectral statistics (LSSs) defined as
\[ \frac{1}{n} \sum_{i=1}^{n} f(\lambda_i) = \int f(x) dF_n^H(x), \]
where \( f(\cdot) \) is a function on \( \mathbb{R}^+ \). As we know, many important statistics in multivariate analysis can be expressed through the ESD, for example,
\[ \frac{1}{n} \text{tr}(H_n) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i = \int x dF_n^H(x) \]
and
\[ \frac{1}{n} \log |H_n| = \frac{1}{n} \sum_{i=1}^{n} \log (\lambda_i) = \int \log(x) dF_n^H(x). \]
Generally, the LSD describes the first-order limits of these LSSs and the CLT then characterizes their second-order asymptotics. The two are the analogs of the law of large numbers and the central limit theorem in the classical probability theory, respectively. As applications, the CLT for LSSs provides an important tool for many hypothesis testing problems in multivariate analysis with the diverging data dimension, e.g., Bai et al. [1] derived the distribution of the likelihood ratio test for high-dimensional data. Both the LSD and the CLT for LSSs study the global law of the empirical eigenvalues. Another fundamental problem is the local law [23], e.g., the asymptotic behaviors of the smallest and largest eigenvalues. The well-known Bai-Yin law [4] derived the limits of the extreme eigenvalues. Johnstone [19] further established the Tracy-Widom law for the largest eigenvalues, which plays a fundamental role in the principal component analysis. For a more comprehensive overview of this topic, one may refer to [3].

In practice, the data normalization is a standard procedure and after that, we are actually dealing with the sample correlation matrix [12]. Parallel to the study of the sample covariance matrix, Jiang [18] first obtained the LSD of the Pearson-type sample correlation matrix and Gao et al. [15] developed the CLT for its LSSs. Bao et al. [10] established the Tracy-Widom law for its extreme eigenvalues and Pillai and Yin [30] extended the result to general cases. However, due to the complex structure of the sample correlation matrix, most results only consider the case where sample data has independent components so that its population covariance matrix is diagonal and then the correlation matrix is an identity matrix. From the perspective of applications, the independence assumption is however too strong so that such results have very limited applicability [12]. On the other hand, for general dependent or correlated data, little work (e.g., [12, 29]) has been done on the sample correlation matrix. As far as the CLT for LSSs, the existing works include [28] that considered the case for Gaussian distributions and [35] that studied the trace moments.

For the sample covariance/correlation matrix, due to the congenital sensitivity of the Pearson-type correlation, finite fourth-order or even higher-order moments of the data distribution are usually required to guarantee the convergence of the limiting distributions. However, most of the results applicable to light-tailed distributions cannot be directly extended to heavy-tailed cases, e.g., Heiny and Yao [17] explored the spectral behavior of Pearson-type correlations for heavy-tailed distributions where the story becomes completely different.

As a remedy for dealing with heavy-tailed data samples, some non-parametric correlation matrices, such as Kendall’s \( \tau \) and Spearman’s \( \rho \), have received considerable attention in recent years. Kendall’s \( \tau \) and Spearman’s \( \rho \) are rank-based and thus there is no need to impose any moment restrictions on the underlying distribution. What is more, the classical theory of non-parametric statistics shows that only partial information will be lost while robustness can be retained if we only use the ranks of the data. In the random matrix theory, Bai and Zhou [5] first derived the LSD of Spearman’s rank correlation matrix, which turns out to be the same as the standard Marčenko-Pastur law. For Kendall’s \( \tau \), Bandeira et al. [6] proved that its LSD is an affine transformation of the standard Marčenko-Pastur law. For the CLT for LSSs, Bao et al. [9] considered Spearman’s rank correlation matrix and Li et al. [25] studied
Kendall’s rank correlation matrix. The Tracy-Widom law for the extreme eigenvalues of the two matrices can be found in [7] and [8], respectively. However, all these asymptotic results are for the data sample with independent components, i.e., all the components are independent. To the best of our knowledge, there are no available results on such rank correlation matrices when the underlying distribution has the general dependent structure. We summarize the developments of the sample covariance matrix, the sample correlation matrix, Kendall’s \( \tau \) and Spearman’s \( \rho \) in Table 1.

As can be seen from Table 1, the asymptotic behaviors of the eigenvalues of the rank correlation matrices under the general dependent structure are still unclear. In this paper, we take the first step to fill this gap and focus on Kendall’s rank correlation matrix with high-dimensional correlated data. Specifically, for the data sample \( x_1, \ldots, x_n \in \mathbb{R}^p \), we define the sign vector

\[
A_{ij} = \text{sign}(x_i - x_j) = (\text{sign}(x_{i1} - x_{j1}), \ldots, \text{sign}(x_{ip} - x_{jp}))^T,
\]

where \( \text{sign}(\cdot) \) denotes the sign function and the sample Kendall’s rank correlation matrix \( K_n \) is given by

\[
K_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} A_{ij}A_{ij}^T. \tag{1.1}
\]

Our goal is to study the spectral properties of \( K_n \) when \( x_i \)'s have a general dependent structure. This is a problem of its own significant interest in the random matrix theory. To study Kendall’s rank correlation matrix with dependence, the LSD is the cornerstone for further derivations of the CLT for LSSs [2] and local laws including the asymptotic distribution of extreme eigenvalues [23]. It is also a key step to solving many high-dimensional statistical problems with heavy-tailed observations. Take the high-dimensional independent test as an example, and many test statistics based on covariance/correlation matrices have been proposed to test complete independence among the components of \( x_i \)'s (see [7, 9, 15, 24, 25, 31]). In particular, Leung and Drton [24] and Li et al. [25] considered test statistics based on linear functions of the eigenvalues of Kendall’s \( \tau \), e.g., \( \text{tr}(K_n^2) \) and \( \log |K_n| \). However, the test power is still unclear since the limiting properties of Kendall’s \( \tau \) (and also Spearman’s \( \rho \)) under general dependent alternatives remain largely unknown.

To answer such questions, in this paper, we take the first step to deriving the limiting spectral distribution of \( K_n \) under the asymptotic regime, where \( p, n \to \infty \) and \( p/n \to c \in (0, \infty) \). One major challenge is the nonlinear dependent structure among the sign-based summands \( A_{ij} \) of \( K_n \). Hence, we apply the Hoeffding decomposition to \( A_{ij} \) to locate the leading terms. In this way, we obtain the equation which the Stieltjes transform of the limiting spectral distribution of \( K_n \) satisfies. It is a brand-new distribution which relies heavily on both the covariance and the conditional covariance structures of \( A_{ij} \). As an illustration, we study the normal distribution, where Kendall’s rank correlation has a specific relation with Pearson’s correlation and then we derive explicit LSDs for some cases with the common dependent structure. Simulation experiments also lend full support to the accuracy of our theoretical results.

| Table 1 | Developments of sample covariance/correlation matrices in the random matrix theory |
|---------|-------------------------------------------------|
|          | Sample covariance | Sample correlation | Kendall’s \( \tau \) | Spearman’s \( \rho \) |
| Independent case (\( \Sigma = I \)) | | | |
| LSD | [27] | [18] | [6] | [5] |
| CLT for LSSs | [2] | [15] | [25] | [9] |
| Tracy-Widom | [19] | [10] | [7] | [8] |
| Dependent case (general \( \Sigma \)) | | | |
| LSD | [27] | [12] | | |
| CLT for LSSs | [2] | [28] | | |
| Tracy-Widom | [14] | | | |
The rest of the paper is organized as follows. In Section 2, we introduce some preliminary knowledge on Kendall’s rank correlation matrix and Hoeffding decomposition. In Section 3, we give our main results on the LSD of Kendall’s rank correlation matrix for correlated data. In Section 4, we consider Gaussian distributions. In Section 5, we collect all the numerical experiments. We give the proofs of the main results in Appendix A.

2 Background on Kendall’s rank correlation matrix

2.1 Kendall’s rank correlation matrix

From the definition of Kendall’s rank correlation matrix (1.1), we can write

$$K_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \text{sign}\left(\frac{x_i - x_j}{\sqrt{2}}\right)\text{sign}\left(\frac{x_i - x_j}{\sqrt{2}}\right)^T,$$

which looks similar to the sample covariance matrix. To be specific, the sample covariance matrix of the data sample $x_1, \ldots, x_n \in \mathbb{R}^p$ can be written as a U-statistic of order two, i.e.,

$$S_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \left(\frac{x_i - x_j}{\sqrt{2}}\right)\left(\frac{x_i - x_j}{\sqrt{2}}\right)^T. \quad (2.1)$$

Despite this similarity in their forms, the inner structure of the two matrices does not follow the same pattern. The sign function introduces the nonlinear correlation into the matrix $K_n$ and it is nontrivial to analyze such a correlation even for binary random variables. For example, Esscher [13] spent a lot of efforts to derive the variance of Kendall’s rank correlation for bi-normal distributions. As can be seen from [11], it is already quite complicated to calculate the integral of the sign function over fourth order even for the normal distribution. Thus, for high-dimensional Kendall’s rank correlation matrix, it is very challenging to study its asymptotic properties.

One appealing property of Kendall’s rank correlation is that it is monotonically invariant [34].

**Proposition 2.1** (Monotonic invariance). For any strictly increasing monotonic functions $f_j(\cdot)$ ($j = 1, \ldots, p$), $K_n$ is invariant for the monotonic component transformation

$$x_i = (x_{i1}, \ldots, x_{ip})^T \mapsto (f_1(x_{i1}), \ldots, f_p(x_{ip}))^T.$$

The reason is that Kendall’s rank correlation is rank-based and the monotonic transformation does not change the order statistics. One special case is that for any linear transformation of the data

$$x_i = (x_{i1}, \ldots, x_{ip})^T \mapsto (\mu_1 + \sigma_1 x_{i1}, \ldots, \mu_p + \sigma_p x_{i1})^T,$$

their corresponding Kendall’s rank correlations remain unchanged, i.e., Kendall’s $\tau$ is a correlation matrix which is invariant to the location and scale. More importantly, for any distribution, Kendall’s rank correlation always exists and this fact makes it an important tool to characterize data with heavy tails.

In previous works on Kendall’s rank correlation matrix for high-dimensional data (e.g., [6, 7, 24, 25]), they assumed that all the components were independent with the absolutely continuous density. By the monotonic invariance, we can always transform each component into a standard normal distribution and all the components are still independent. Thus, it can be formulated as that $x_1, \ldots, x_n$ are independent and identically distributed (i.i.d.) from a standard multivariate normal distribution $N(0, I_p)$, from which we can see that these results are very limited. In this work, we consider more general cases where the data components are allowed to have dependence.

2.2 Hoeffding decomposition

In this subsection, to deal with the nonlinear dependent structure of high-dimensional Kendall’s rank correlation matrix, we apply the Hoeffding decomposition to find out the leading terms. Specifically,
denote by $A_i$ the conditional expectation of $A_{ij}$ given $x_i$, i.e.,
\[
A_i \triangleq E\{\text{sign}(x_i - x) \mid x_i\}. \tag{2.2}
\]
The Hoeffding decomposition for $A_{ij}$ can be written as
\[
A_{ij} = A_i - A_j + \epsilon_{ij}, \tag{2.3}
\]
where
\[
\epsilon_{ij} = \text{sign}(x_i - x_j) - E\{\text{sign}(x_i - x_j) \mid x_i\} + E\{\text{sign}(x_i - x_j) \mid x_j\}.
\]

Throughout this paper, we assume that $x_1, \ldots, x_n$ are i.i.d. from a population with the absolutely continuous density. Then we have
\[
E(A_{ij}) = E(A_i) = E(\epsilon_{ij}) = 0,
\]
and the covariance matrices of $A_{ij}$, $A_i$ and $\epsilon_{ij}$ exist. Specifically, we define
\[
\Sigma_1 \triangleq \text{cov}(A_{ij}), \quad \Sigma_2 \triangleq \text{cov}(A_i), \quad \text{cov}(\epsilon_{ij}) = \Sigma_1 - 2\Sigma_2 \triangleq \Sigma_3. \tag{2.4}
\]

### 2.3 Preliminary results

With the Hoeffding decomposition of $A_{ij}$ described in (2.3), Kendall’s rank correlation matrix $K_n$ can be decomposed accordingly, i.e.,
\[
K_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} (A_i - A_j)(A_i - A_j)^T + \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \epsilon_{ij}(A_i - A_j)^T + \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \epsilon_{ij}\epsilon_{ij}^T =: M_1 + M_2 + M_2^T + M_3, \tag{2.5}
\]
where
\[
M_1 \triangleq \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} (A_i - A_j)(A_i - A_j)^T = \frac{2}{n-1} \sum_{i=1}^{n}(A_i - \bar{A})(A_i - \bar{A})^T,
\]
\[
M_2 \triangleq \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \epsilon_{ij}(A_i - A_j)^T,
\]
\[
M_3 \triangleq \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \epsilon_{ij}\epsilon_{ij}^T,
\]
and $\bar{A} = \frac{1}{n} \sum_{i=1}^{n} A_i$. In particular, $M_1$ is the sample covariance matrix formed by i.i.d. random vectors $A_1, \ldots, A_n$.

Our first result is to show that the error terms $M_2$ and $M_3 - E(M_3)$ can be controlled so that the dominant contribution in terms of the LSD of $K_n$ is from $M_1$. As a result, we define the following matrix:
\[
W_n \triangleq M_1 + \Sigma_3 = \frac{2}{n-1} \sum_{i=1}^{n}(A_i - \bar{A})(A_i - \bar{A})^T + \Sigma_3. \tag{2.6}
\]
Throughout this paper, we use $\| \cdot \|$ and $\| \cdot \|_2$ to denote the common spectral norm and the Frobenius norm of a matrix, respectively.
Proposition 2.2. Assume \( \| \Sigma_1 \| \leq C \) for some universal constant \( C \) and
\[
\frac{1}{p^2} \text{var}(A_{12}^T A_{13}) \to 0.
\]

Then we have
\[
L(F_{K_n}, F_{W_n}) \to 0 \quad \text{in probability},
\]
where \( L(\cdot, \cdot) \) is the Lévy distance between two distributions.

Remark 2.3. The assumptions on \( \| \Sigma_1 \| \) and \( \text{var}(A_{12}^T A_{13}) \) are to avoid too strong dependence among the components of the data. In the random matrix theory, it is a regular condition to assume that the norm of the population covariance matrix is uniformly bounded (e.g., [5, Condition 3]). Here, this condition is also required for bounding the difference between \( K_n \) and \( W_n \). To show that such a condition \( \| \Sigma_1 \| \leq C \) is necessary, we conduct a toy example in the following. Specifically, we generate the data sample \( x_1, \ldots, x_n \) i.i.d. \( \sim N(0, \Sigma) \), where \( \Sigma \) is a matrix with \( \Sigma_{ii} = 1 \) and \( \Sigma_{ij} = \rho \). In this case,
\[
\| \Sigma_1 \| = 1 + \frac{2}{\pi}(p - 1) \arcsin(\rho),
\]
which is unbounded for any \( \rho > 0 \). Figure 1 presents the distance \( \| K_n - W_n \|_2^2 / p \) versus the increasing \( \rho \), from which we can see that \( \| \Sigma_1 \| \leq C \) is necessary for bounding the difference between \( F_{K_n} \) and \( F_{W_n} \).

Note that \( A_1, \ldots, A_n \) are i.i.d. random vectors with covariance matrix \( \text{cov}(A_i) = \Sigma_2 \) and
\[
E K_n = \Sigma_1 = 2 \Sigma_2 + \Sigma_3.
\]

Proposition 2.2 shows that one part of Kendall’s rank correlation matrix has similar fluctuations to the usual sample covariance matrix with the population covariance matrix \( 2 \Sigma_2 \) and the other part is concentrated on the deterministic matrix \( \Sigma_3 \). This phenomenon is an analog of Hoeffding decomposition for the classical U-statistics. By implementing the Hoeffding decomposition for the random vector \( A_{ij} \), we then transfer the study of the LSD of \( K_n \) to the study of the LSD of \( W_n \).

Figure 1 Plots of the scaled squared Frobenius norm of \( K_n - W_n \) versus \( \rho \in [0, 0.9] \). Data sample \( x_1, \ldots, x_n \) is generated from a multivariate normal distribution \( N_p(0, \Sigma) \) with \( \Sigma_{ii} = 1 \) and \( \Sigma_{ij} = \rho \). Here, \( (n, p) = (100, 200) \) and the results are based on 100 replications.
For the i.i.d. continuous data sample $x_1, \ldots, x_n \in \mathbb{R}^p$, assume that
\begin{align}
(A) \quad \text{as } p \to \infty,
\frac{1}{p^2} \text{var}(A^T_{12} A_{13}) \to 0 \quad \text{and} \quad \frac{1}{p^2} \text{var}(A^T_{1} BA_{1}) \to 0,
\end{align}

We first provide a general result as follows and then study the case for Gaussian distributions in the next section.

\section{Limiting spectral distribution of $K_n$}

In this section, we present the LSD of Kendall’s rank correlation matrix $K_n$. We first introduce the concept of the Stieltjes transform, which is an important tool in the random matrix theory. Let $\mu$ be a finite measure on $\mathbb{R}$. Its Stieltjes transform $s_\mu(z)$ is defined as
\begin{align}
s_\mu(z) = \int \frac{1}{x-z} \mu(dx), \quad z \in \mathbb{C}^+,
\end{align}
where $\mathbb{C}^+$ denotes the upper complex plane. We can also obtain $\mu$ from $s_\mu(z)$ by the inversion formula. For any two continuity points $a < b$ of $\mu$, we have
\begin{align}
\mu([a, b]) = \lim_{\nu \to 0^+} \frac{1}{\pi} \int_a^b \Im s_\mu(x + i\nu) dx,
\end{align}
where $\Im$ is the imaginary part of a complex number, and $i$ is the imaginary unit.

For Kendall’s rank correlation matrix, Bandeira et al. \cite{6} derived the LSD when the observations $x_1, \ldots, x_n$ are i.i.d. random vectors and the components are also independent with the absolutely continuous density. They show that as $n \to \infty$, $p/n \to c \in (0, \infty)$, the ESD of such Kendall’s rank converges in probability to an affine transformation of the standard Marčenko-Pastur law with the parameter $c$, which has an explicit form whose density function $p_c(x)$ is given by
\begin{align}
p_c(x) = \frac{9}{4\pi c(3x-1)} \sqrt{(c_+ - x)(x - c_-)} + (1 - 1/c) \delta_{\frac{1}{2}}(c > 1), \quad c_- \leq x \leq c_+,
\end{align}
where $c_- = \frac{1}{3} + \frac{2}{3}(1 - \sqrt{c})^2$ and $c_+ = \frac{1}{3} + \frac{2}{3}(1 + \sqrt{c})^2$. The corresponding Stieltjes transform $s(z) \in \mathbb{C}^+$ is the unique solution to the following equation:
\begin{align}
\frac{2}{3} z (z - 1)^2 s(z) + \left(z - 1 + \frac{2}{3} c\right) s(z) + 1 = 0.
\end{align}

To illustrate the challenges of Kendall’s rank correlation matrix in the random matrix theory, we consider the ranking of the data
\begin{align}
x^T_1 &\quad \cdots \quad x^T_1 p \\
\vdots &\quad \ddots \quad \vdots \\
x^T_n &\quad \cdots \quad x^T_n p
\end{align}
\begin{align}
\begin{pmatrix}
x_{11} & \cdots & x_{1p} \\
\vdots & \ddots & \vdots \\
x_{n1} & \cdots & x_{np}
\end{pmatrix}
\implies
\begin{pmatrix}
r_{11} & \cdots & r_{1p} \\
\vdots & \ddots & \vdots \\
r_{n1} & \cdots & r_{np}
\end{pmatrix},
\end{align}
where each column $(r_{1j}, \ldots, r_{nj})^T$ is the rank of the raw data $(x_{1j}, \ldots, x_{nj})^T$. For the i.i.d. sample $x_1, \ldots, x_n \in \mathbb{R}^p$, each column of the ranking matrix follows the uniform distribution on the set of all $n!$ permutations of $\{1, 2, \ldots, n\}$. While the rows of the raw data matrix are independent, the ranking matrix does not have independent rows anymore. In the special case where the columns of the raw data matrix are also independent (e.g., \cite{6, 7, 24, 25}), the columns of the ranking matrix will be independent. Then the raw data matrix with i.i.d. entries has very limited applications. If the columns of the raw data matrix are dependent, e.g., there is a covariance structure among components, both the rows and the columns of the ranking matrix are dependent. From the perspective of the random matrix theory, analyzing such matrices is very challenging.

We first provide a general result as follows and then study the case for Gaussian distributions in the next section.

\begin{theorem}
For the i.i.d. continuous data sample $x_1, \ldots, x_n \in \mathbb{R}^p$, assume that
\begin{align}
(A) \quad \text{as } p \to \infty,
\frac{1}{p^2} \text{var}(A^T_{12} A_{13}) \to 0 \quad \text{and} \quad \frac{1}{p^2} \text{var}(A^T_{1} BA_{1}) \to 0,
\end{align}
\end{theorem}
where $B$ is any deterministic matrix with the bounded spectral norm;
(B) $\|\Sigma_1\| \leq C$ for some universal constant $C$, and also the solution $x(z) \in \mathbb{C}$ to the following equation exists:

$$\frac{1}{x(z)} = 1 + \lim_{n \to \infty} \frac{2}{n} \text{tr}[(\Sigma_3 + 2x(z)\Sigma_2 - zI_p)^{-1}\Sigma_2];$$  \hfill (3.4)

(C) $p, n \to \infty$ such that $p/n \to c \in (0, \infty)$.

Then in probability, the empirical spectral distribution $F^{K_n}$ converges weakly to a limiting spectral distribution $F$ whose Stieltjes transform $s(z)$ is given by

$$s(z) = \lim_{p \to \infty} \frac{1}{p} \text{tr}[(\Sigma_3 + 2x(z)\Sigma_2 - zI_p)^{-1}].$$  \hfill (3.5)

Recall the dominating matrix $W_n$ in (2.6), which has the same LSD as the following matrix:

$$\frac{2}{n} \sum_{i=1}^{n} A_i A_i^T + \Sigma_3,$$  \hfill (3.6)

where $n^{-1} \sum_{i=1}^{n} A_i A_i^T$ is the type of a sample covariance matrix corresponding to the data sample $\{A_i\}$ with the population covariance matrix $\text{cov}(A_i) = \Sigma_2$. However, the components within $A_i$ are nonlinearly correlated and thus cannot be written in the form of the independent component model such that $A_i = \Sigma_2^{1/2} z_i$. For those weakly dependent data samples, Bai and Zhou [5] proved that under certain conditions, the LSD of the sample covariance matrix still follows the generalized Marchenko-Pastur law. One of the crucial conditions is that the variance of the quadratic forms $A_i^T B A_i$ is relatively small (see [5, Theorem 1.1]), i.e., $\text{var}(A_i^T B A_i) = o(n^2)$, which is actually the second part of Assumption (A). Under this condition, the LSD of $n^{-1} \sum_{i=1}^{n} A_i A_i^T$ remains the same as the generalized Marchenko-Pastur law corresponding to the population covariance matrix $\Sigma_2$.

On the other hand, the LSD of the Hermitian matrix of the type $XTX^T/n + A$ has been studied in [32], where $X$ is assumed to be an $n \times p$ random matrix with i.i.d. standardized entries, $T$ is a diagonal matrix having an LSD, $A$ is a Hermitian matrix and the three matrices are independent. Under certain conditions, Silverstein and Bai [32] proved that the LSD of $XTX^T/n + A$ is a shift of the LSD of $A$. Intuitively, this is because the population version of $XTX^T$ equals $(\text{tr}T)I_n$, which shares the same eigenvectors as the matrix $A$. However, in our case, the population version of $n^{-1} \sum_{i=1}^{n} A_i A_i^T$ in (3.6) equals $\Sigma_3$, whose eigenvectors might be different from the ones of $\Sigma_3$. Therefore, we cannot directly apply the results in [32]. Using the terminology of matrix subordination [20], the limiting Stieltjes transform of $XTX^T + A$ is subordinated to the limiting Stieltjes transform of $A$. Roughly, our result is the same as the model $2\Sigma_2^{1/2}X^T X \Sigma_2^{1/2} + \Sigma_3$. Our Theorem 3.1 shows that both the eigenvalues and eigenvectors of the two matrices, $\Sigma_2$ and $\Sigma_3$, will contribute to the LSD of $K_n$. Thus, it is a brand-new LSD for the covariance/correlation matrix.

Technically, Assumption (B) is a new condition and in Appendix A, we prove the uniqueness of $s(z)$ or $x(z)$ if it exists. Here, we have some discussions on this condition. If $\Sigma_3 = \mathbf{0}$, the equation (3.4) will be

$$\lim_{p \to \infty} \frac{1}{p} \text{tr}[(2x(z)\Sigma_2 - zI_p)^{-1}] = \frac{1 - c - x}{cz},$$

which means that the limiting Stieltjes transform of $\Sigma_2$ exists and we solve the above equation to get $x(z)$. The Stieltjes transform of the LSD is then

$$s(z) = \lim_{p \to \infty} \frac{1}{p} \text{tr}[(2x(z)\Sigma_2 - zI_p)^{-1}] = \frac{1 - c - x}{cz},$$

and then we can get

$$\frac{x}{z} = \frac{1 - c}{z} + c \cdot s(z),$$
where the right-hand side is exactly the limiting Stieltjes transform of $X \Sigma_2 X^T / n$. Thus, our result in the special case $\Sigma_3 = 0$ is consistent with the one in [5]. Now we consider another special case $\Sigma_2 = I_p / 2$. From (3.4) and (3.5), we can get

$$s(z) = \lim_{p \to \infty} \frac{1}{p} \text{tr}[(\Sigma_3 + (x - z)I_p)^{-1}] = \frac{1 - x}{xc},$$

which yields $x = \frac{1}{1 + \frac{1}{\pi} \arcsin \rho}$. This result is consistent with the one in [32] when $T = I$.

In summary, our new LSD extends the results in [5, 32]. For general $\Sigma_2$ and $\Sigma_3$, it is challenging to study the limits of (3.4). One special case is that $\Sigma_2$ and $\Sigma_3$ are simultaneously diagonalizable and the Toeplitz matrix is such an example, which will be studied in the next section.

4 Gaussian ensemble

As mentioned in Section 1, Pearson’s correlation matrix has been thoroughly studied in the random matrix theory. Generally, there is no explicit relation between Kendall’s correlation and Pearson’s correlation (see [22] for more details). A special ensemble is the Gaussian distribution, where Kendall’s correlation has a monotonic correspondence with Pearson’s correlation. This neat relation is presented in the following lemma which is called Grothendieck’s identity.

**Lemma 4.1** (Grothendieck’s identity). Consider a bi-variate normal distribution

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right\},$$

where $\rho \in [-1, 1]$. We have

$$E \{ \text{sign}(z_1) \text{sign}(z_2) \} = 4E \{ I(z_1, z_2 > 0) \} - 1 = \frac{2}{\pi} \arcsin \rho.$$

Assume $x_1, \ldots, x_n \overset{i.i.d.}{\sim} N(0, \Sigma)$, where $\Sigma$ is a correlation matrix. By Lemma 4.1, we can show that

$$\Sigma_1 = \frac{2}{\pi} \arcsin(\Sigma), \quad \Sigma_2 = \frac{2}{\pi} \arcsin(\Sigma/2), \quad \Sigma_3 = \frac{2}{\pi} \arcsin(\Sigma) - \frac{4}{\pi} \arcsin(\Sigma/2). \quad (4.1)$$

Thus, for Gaussian distributions, Kendall’s rank correlation matrix is determined by Pearson’s correlation matrix $\Sigma$.

In this section, we consider the LSD of $K_n$ for Gaussian ensembles which can shed new light on Kendall’s correlation matrix and also its connections with the sample covariance/correlation matrix.

**Proposition 4.2.** Assume $x_1, \ldots, x_n \overset{i.i.d.}{\sim} N(0, \Sigma)$, where $\Sigma$ is a correlation matrix. Under Assumptions (B) and (C) in Theorem 3.1, the conclusion of Theorem 3.1 holds.

**Remark 4.3.** It is noted that although we consider Gaussian ensembles, the results actually cover a wider range of distributions, which is called the non-paranormal distribution [26] due to the monotonic invariance of Kendall’s rank correlation matrix. To be specific, a random vector $Y = (Y_1, \ldots, Y_p)^T \in \mathbb{R}^p$ is said to have a non-paranormal distribution if there exist monotone functions $\{f_j\}_{j=1}^p$ such that $(f_1(Y_1), \ldots, f_p(Y_p)) \sim N(\mu, \Sigma)$.

The proof of Proposition 4.2 is to check Assumption (A) for Gaussian distributions. Especially, for the normal distribution or the non-paranormal distribution, we can calculate the variance of $A_{12}^T A_{13}$ explicitly which is based on the classical results in [13] and control the variance of the quadratic form $A_{12}^T B A_{13}$ using the Poincaré inequality. Hence, Assumption (A) holds for Gaussian distributions and the detailed proof is presented in Appendix A. Next, we consider some examples to illustrate the result.
4.1 Independent case

A very special case is the standard multivariate normal distribution, i.e., $\Sigma = I_p$. By the monotonic invariance of Kendall’s rank correlation matrix, it is equivalent to the independent case considered by [6,7,24,25].

When $\Sigma = I_p$, we know $\Sigma_2 = \Sigma_3 = \frac{1}{3} I_p$. Intuitively, the matrix given in (3.6) reduces to a standard sample covariance matrix corresponding to the population covariance matrix $\Sigma_2 = \frac{2}{3} I_p$ and the deterministic matrix $\Sigma_3 = \frac{1}{3} I_p$. This explains that its LSD is $\frac{2}{3} \text{MP} + \frac{1}{3}$. As an illustration of our main theorems, we demonstrate this result using Theorem 3.1 and Proposition 4.2 in the following.

Starting from (3.4), we have

$$\frac{1}{x(z)} = 1 + \frac{2c}{1 + 2x(z) - 3z},$$

and this equation has a unique solution in $\mathbb{C}^-:

$$x(z) = \frac{1}{4} \{1 - 2c + 3z - \sqrt{(2y - 3z - 1)^2 - 8(3z - 1)}\}.$$ 

Plugging it into (3.5), we obtain

$$s(z) = \frac{1 - \frac{2}{3}c - z + \sqrt{(z - 1 - \frac{2}{3}c)^2 - \frac{16}{3}c}}{\frac{3}{4}x(z - \frac{1}{3})},$$

which is the Stieltjes transform of $\frac{2}{3} \text{MP} + \frac{1}{3}$ as shown in [6].

4.2 MA(1) model

Next, we consider an MA(1) model with the population correlation matrix $\Sigma$ as follows:

$$\Sigma = \Sigma(\rho) = \begin{pmatrix} 1 & \rho & 0 & \cdots & 0 \\ \rho & 1 & \rho & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & \rho & 1 \end{pmatrix},$$

where $\rho \in (-1/2, 1/2)$. The eigenvalues of $\Sigma(\rho)$ are given by

$$\lambda_k(\rho) = 1 + 2\rho \cos \frac{k\pi}{p + 1}, \quad k = 1, \ldots, p$$

and the corresponding eigenvectors are

$$u_k = \sqrt{\frac{2}{p + 1}} \begin{pmatrix} \sin \frac{k\pi}{p + 1} \\ \sin \frac{2k\pi}{p + 1} \\ \vdots \\ \sin \frac{pk\pi}{p + 1} \end{pmatrix}^T.$$

A detailed calculation of the eigenvalues and eigenvectors can be found in [33, Lemma 1]. It is noted that the eigenvectors of $\Sigma(\rho)$ do not depend on the correlation parameter $\rho$. Thus, $\Sigma_2$ and $\Sigma_3$ share the same eigenvectors and we can derive the two limits of Theorem 3.1 as follows.

**Proposition 4.4.** Assume that $x_1, \ldots, x_n \overset{\text{i.i.d.}}{\sim} N(0, \Sigma(\rho))$. Then the Stieltjes transform $s(z)$ of the LSD of $K_n$ satisfies

$$s(z) = \frac{1}{\sqrt{(\frac{2}{3} + \frac{2x(z)}{3} - z)^2 - 4\left(\frac{2}{3} \text{arcsin} \rho + \frac{2(x(z) - 1)}{3} \text{arcsin} \frac{2}{3}\right)^2}}.$$ (4.2)
Here, $s(z) \in \mathbb{C}^+$ and $x(z) \in \mathbb{C}^-$ satisfy
\[
\frac{1}{2c} \left( \frac{1}{x(z)} - 1 \right) = \frac{1 - c(x(z), \rho)(1 + 2x(z) - 3z)}{3} s(z) + c(x(z), \rho),
\]
where
\[
c(x, \rho) = \begin{cases} 
\arcsin \frac{\rho}{2}, & \text{if } \rho \neq 0, \\
\arcsin \rho + 2(x - 1) \arcsin \frac{\rho}{2}, & \text{if } \rho = 0.
\end{cases}
\]

By solving (4.2) and (4.3) in Proposition 4.4, we can derive the Stieltjes transform $s(z)$ of the LSD of $K_n$ when samples are from an MA(1) model.

### 4.3 Toeplitz structure

Last but not least, we consider a more general case where the population correlation matrix $\Sigma$ has a Toeplitz structure
\[
\Sigma = \begin{pmatrix} 
1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\
\rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\rho_{p-2} & \cdots & \rho_1 & 1 & \rho_1 \\
\rho_{p-1} & \cdots & \rho_1 & 1 & 1 
\end{pmatrix},
\]
(4.4)

where the correlations are absolutely summable, i.e.,
\[
\sum_{k=1}^{\infty} |\rho_k| < \infty.
\]

Define the function
\[
f(\theta) = 1 + \sum_{k=1}^{\infty} \rho_k (e^{i\kappa \theta} + e^{-i\kappa \theta}) = 1 + 2 \sum_{k=1}^{\infty} \rho_k \cos(k\theta),
\]
whose Fourier series is exactly $(1, \rho_1, \ldots)$. The Szegö theorem [16] shows that the eigenvalues of $\Sigma$ can be approximated by
\[
f \left( \frac{k\pi}{p+1} \right), \quad k = 1, \ldots, p.
\]

For the limit in (3.4), which involves two Toeplitz matrices, we cannot apply the Szegö theorem directly. However, a Toeplitz matrix can be approximated by a circulant matrix [16, Lemma 11] whose eigenvectors are universal for its entries. By [16, Theorems 11 and 12], under some mild conditions, we have the following limit.

**Proposition 4.5.** For the Toeplitz matrix $\Sigma$ defined in (4.4), we have
\[
\lim_{p \to \infty} \frac{1}{p} \text{tr}[(\Sigma_3 + 2x \Sigma_2 - z I_p)^{-1} \Sigma_2] = \frac{1}{2\pi} \int_0^{2\pi} \frac{f_2(\theta)}{f_1(\theta)} d\theta,
\]
(4.6)

where
\[
f_1(\theta) = \frac{1}{3} + \frac{2x}{3} - z + 2 \sum_{k=1}^{\infty} \left[ \frac{2}{\pi} \arcsin \rho_k + \frac{4(x - 1)}{\pi} \arcsin \frac{\rho_k}{2} \right] \cos(k\theta),
\]
\[
f_2(\theta) = \frac{1}{3} + \frac{4}{\pi} \sum_{k=1}^{\infty} \arcsin \frac{\rho_k}{2} \cos(k\theta).
\]
The absolutely summable condition (4.5) and the bound for arcsin(·) guarantee the existence of the Fourier functions $f_1(·)$ and $f_2(·)$. By solving (3.4) using the limit (4.6) in Proposition 4.5, we can theoretically derive $x(z)$. For the limit (3.5) in Theorem 3.1, the Szegő theorem can yield the result directly. In summary, we can obtain the Stieltjes transform $s(z)$ of the LSD of $K_n$ when samples are from the Toeplitz covariance matrix model as follows.

**Proposition 4.6.** Assume that $x_1,\ldots,x_n \overset{i.i.d.}{\sim} N(0, \Sigma)$, where $\Sigma$ is a Toeplitz matrix (4.4). Then the Stieltjes transform $s(z)$ of the LSD of $K_n$ satisfies

$$s(z) = \lim_{p \to \infty} \frac{1}{p} \text{tr}(\Sigma_3 + 2z\Sigma_2 - zI_p)^{-1} = \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{f_1(\theta)} d\theta,$$

where

$$\frac{1}{x(z)} = 1 + \frac{c}{\pi} \int_0^{2\pi} \frac{f_2(\theta)}{f_1(\theta)} d\theta.$$

5 Simulation

In this section, simulation experiments are conducted to examine the finite sample performance of eigenvalues of Kendall’s sample correlation matrix when the data sample follows different dependence structures. We generate the data sample $x_1,\ldots,x_n \sim N_p(0, \Sigma_0)$, draw the histogram of eigenvalues of Kendall’s sample correlation matrix and compare their theoretical densities. Specifically, we consider four types of the covariance matrix $\Sigma_0$:

(I) independent case: $\Sigma_0 = \Sigma = I_p$;

(II) factor model: $x_t = Af_t + \varepsilon_t$, where $f_t \sim N_k(0, I_k)$, $\varepsilon_t \sim N_p(0, I_p)$, and thus cov($x_t$) = $\Sigma_0 = I_p + A^T A$, where rank($A$) = $k = 3$;

(III) MA(1) model: all the diagonal entries of $\Sigma_0$ are 1, both the upper and the lower subdiagonal entries are $\rho$, and others are zero;

(IV) the general Toeplitz matrix with $\rho_1 = \rho_2 = \rho$ and $\rho_k = 0$ for $k \geq 3$.

Here, the population covariance matrix $\Sigma_0$ and the correlation matrix $\Sigma$ are the same in Models (I), (III) and (IV).

**Model (I).** As for the independent case, we consider three types of sample correlation matrices, Pearson $R_n$, Spearman $S_n$ and Kendall $K_n$. Specifically, for our data sample $X_n = (x_1,\ldots,x_n)_{p \times n}$ and $x_t = (x_{t1},\ldots,x_{tp})_{p \times 1}$, both $S_k = (s_{kl})$ and $R_k = (\rho_{kl})$ are $p \times p$ matrices, where $s_{kl}$ and $\rho_{kl}$ are Spearman’s and Pearson’s correlation of the $k$-th and $\ell$-th row of $X_n$ with

$$s_{kl} = \frac{\sum_{i=1}^{n}(r_{ki} - \bar{r}_k)(r_{li} - \bar{r}_l)}{\sqrt{\sum_{i=1}^{n}(r_{ki} - \bar{r}_k)^2} \sqrt{\sum_{i=1}^{n}(r_{li} - \bar{r}_l)^2}}, \quad \bar{r}_k = \frac{1}{n} \sum_{i=1}^{n} r_{ki} = \frac{n + 1}{2},$$

$$\rho_{kl} = \frac{\sum_{i=1}^{n}(x_{ki} - \bar{x}_k)(x_{li} - \bar{x}_l)}{\sqrt{\sum_{i=1}^{n}(x_{ki} - \bar{x}_k)^2} \sqrt{\sum_{i=1}^{n}(x_{li} - \bar{x}_l)^2}}, \quad \bar{x}_k = \frac{1}{n} \sum_{i=1}^{n} x_{ki},$$

where $r_{ki}$ is the rank of $x_{ki}$ among $(x_{k1},\ldots,x_{kn})$. From [5,18], we know that the LSDs of $S_n$ and $R_n$ both are the standard Marčenko-Pastur law while the LSD of $K_n$ is an affine transformation of the Marčenko-Pastur law [6]. Thus we list the histogram of eigenvalues of all three types of sample correlation matrices under different combinations of $(p, n)$ and compare their corresponding limiting densities in Figure 2. It can be seen from Figure 2 that all the histograms conform to their theoretical limits, which fully supports our theoretical results in the independent case.

**Model (II).** As for the factor model or the spiked model, the population covariance matrix is

$$\text{cov}(x_t) = \Sigma_0 = I_p + A^T A,$$
where \( \text{rank}(A) \leq k \). For the related correlation matrix \( \Sigma \), we have

\[
\Sigma = \text{diag}(\Sigma_0)^{-1/2} \cdot \Sigma_0 \cdot \text{diag}(\Sigma_0)^{-1/2} = \text{diag}(\Sigma_0)^{-1} + \tilde{A}^T \tilde{A},
\]

where \( \tilde{A} = A \text{diag}(\Sigma_0)^{-1/2} \). Noting that

\[
\Sigma_1 = \frac{2}{\pi} \arcsin(\Sigma), \quad \Sigma_2 = \frac{2}{\pi} \arcsin(\Sigma/2), \quad \Sigma_3 = \Sigma_1 - 2\Sigma_2
\]

and \( 2x/\pi \leq 2\arcsin(x)/\pi \leq x \) for any \( x \in [0,1] \), we have

\[
\frac{1}{p} \| \Sigma_1 - I_p \|_2^2 \leq \frac{1}{p} \left\| \frac{2}{\pi} \arcsin(\tilde{A}^T \tilde{A}) \right\|_2^2 \leq \frac{1}{p} \left\| \tilde{A}^T \tilde{A} \right\|_2^2 \leq \frac{1}{p} \left\| A^T A \right\|_2^2
\]

and

\[
\frac{1}{p} \left\| \Sigma_2 - \frac{1}{3} I_p \right\|_2^2 \leq \frac{1}{p} \left\| \frac{2}{\pi} \arcsin(\tilde{A}^T \tilde{A}/2) \right\|_2^2 \leq \frac{1}{4p} \left\| A^T A \right\|_2^2.
\]

Thus, when \( \frac{1}{p} \left\| A^T A \right\|_2^2 \to 0 \), the LSD is still an affine transformation of the Marčenko-Pastur law [6]. If the term \( \frac{1}{p} \left\| A^T A \right\|_2^2 / p \) is large, the result violates the affine transformation of the Marčenko-Pastur law. To demonstrate these results, we consider two covariance matrices

\[
\Sigma_0 = I_p + \frac{1}{p} Z^T Z \quad \text{and} \quad \Sigma_0 = I_p + \frac{1}{\sqrt{p}} Z^T Z,
\]

where \( Z = (Z_{ij})_{k \times p} \) and \( Z_{ij} \overset{\text{i.i.d.}}{\sim} N(0,1) \). Figure 3 shows the results which are consistent with our analysis.
Model (III). As for the MA(1) model, we focus on the spectral behavior of $K_n$. The red curves are density functions of the affine transformation of the Marčenko-Pastur law.

As for the MA(1) model, we focus on the spectral behavior of $K_n$ since little is known about $S_n$ and $R_n$ in the dependent case. Similarly, we generate the data sample $x_1, \ldots, x_n \sim N(0, \Sigma)$, where $\Sigma$ follows the MA(1) model with $\rho = 0.5$. The LSD is derived by using Proposition 4.4 and the inversion formula (3.1). Then the histograms of eigenvalues of $K_n$ under different combinations of $(p, n)$ are compared with their corresponding limiting densities in Figure 4. It can be seen from Figure 4 that the LSDs under the MA(1) model are different from the independent case. All the empirical histograms conform to our theoretical limits, which proves the accuracy of our theory.

Model (IV). As an illustration for the general Toeplitz matrix, we consider a band Toeplitz matrix with two parameters, i.e.,

$$\rho_1 = \rho_2 = \rho, \quad \rho_k = 0, \quad k = 3, \ldots$$

Noting that

$$\Sigma_2 = \frac{2}{\pi} \arcsin(\Sigma/2), \quad \Sigma_3 = \frac{2}{\pi} \arcsin(\Sigma) - \frac{4}{\pi} \arcsin(\Sigma/2),$$

we have

$$\Sigma_3 - \frac{1}{3} I_p = a \left( \Sigma_2 - \frac{1}{3} I_p \right), \quad a = \frac{\arcsin(\rho)}{\arcsin(\rho/2)} - 2,$$

and then

$$\Sigma_3 = a \Sigma_2 + \frac{1 - a}{3} I_p.$$

Assumption (A) is

$$\lim_{p \to \infty} \frac{1}{p} \text{tr} \left[ \left( 2x + a \right) \Sigma_2 - \left( z - \frac{1 - a}{3} \right) I_p \right]^{-1} \Sigma_2 = \frac{1 - x}{2cx},$$
Figure 4 (Color online) Histograms of eigenvalues of Kendall’s sample correlation matrices for the data sample \(x_1, \ldots, x_n \sim N_p(0, \Sigma)\), where \(\Sigma\) follows the MA(1) model with \(\rho=0.5\) for \((p, n) = (200, 400), (p, n) = (300, 400), (p, n) = (300, 200)\) and \((p, n) = (400, 200)\). The red curves are density functions of their corresponding limiting spectral distributions. The black dashed lines are densities in the independent case for reference.

Figure 5 (Color online) Histograms of eigenvalues of Kendall’s sample correlation matrices for the data sample \(x_1, \ldots, x_n \sim N_p(0, \Sigma)\), where \(\Sigma\) follows the general Toeplitz model with \(\rho=0.25\) for \((p, n) = (200, 400), (p, n) = (300, 400), (p, n) = (300, 200)\) and \((p, n) = (400, 200)\). The red curves are density functions of their corresponding limiting spectral distributions.
and Proposition 4.5 yields

\begin{equation}
\frac{1-x}{2cx} = \frac{1}{2\pi} \int_0^{2\pi} \frac{f_2(\theta)}{(2x + a)f_2(\theta) - (z - \frac{1-x}{3})} d\theta,
\end{equation}

where

\[ f_2(\theta) = \frac{1}{3} + \frac{4}{\pi} \arcsin \frac{\rho}{2} [\cos(\theta) + \cos(2\theta)]. \]

Solving (5.1) to get \( x(z) \), we see that the Stieltjes transform of the LSD (3.5) is

\[
s(z) = \lim_{p \to \infty} \frac{1}{p} \text{tr}\left( (2x + a)\Sigma_2 - \left(z - \frac{1-a}{3}\right) I_p \right)^{-1}
\]

\[
= -\frac{1}{z_1} \lim_{p \to \infty} \frac{1}{p} \text{tr}\left( (2x + a)\Sigma_2 - z_1 I_p \right)^{-1} ((2x + a)\Sigma_2 - z_1 I_p - (2x + a)\Sigma_2)
\]

\[
= \frac{1}{z_1} \left( 1 - \frac{(2x + a)(1-x)}{2cx} \right) = \frac{(2x + a)(1-x) - 2cx}{2cz_1x},
\]

where \( z_1 = z - \frac{1-a}{3} \). Figure 5 shows the results with \( \rho = 0.25 \) and again, we can see that the empirical histograms conform to our theoretical result.

Acknowledgements Zeng Li’s work was supported by National Natural Science Foundation of China (Grant Nos. 12031005 and 12101292). Cheng Wang’s work was supported by National Natural Science Foundation of China (Grant No. 12031005) and Natural Science Foundation of Shanghai (Grant No. 21ZR1432900). Qinwen Wang’s work was supported by National Natural Science Foundation of China (Grant No. 12171099). The authors contributed equally to this work and are listed in alphabetical order. The authors thank the reviewers for their insightful comments.

References

1 Bai Z D, Jiang D D, Yao J-F, et al. Corrections to LRT on large-dimensional covariance matrix by RMT. Ann Statist, 2009, 37: 3822–3840
2 Bai Z D, Silverstein J W. CLT for linear spectral statistics of large-dimensional sample covariance matrices. Ann Probab, 2004, 32: 553–605
3 Bai Z D, Silverstein J W. Spectral Analysis of Large Dimensional Random Matrices. New York: Springer, 2010
4 Bai Z D, Yin Y Q. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. Ann Probab, 1993, 21: 1275–1294
5 Bai Z D, Zhou W. Large sample covariance matrices without independence structures in columns. Statist Sinica, 2008, 425–442
6 Bandeira A S, Lodhia A, Rigollet P. Marčenko-Pastur law for Kendall’s tau. Electron Commun Probab, 2017, 22: 1–7
7 Bao Z G. Tracy-Widom limit for Kendall’s tau. Ann Statist, 2019, 47: 3504–3532
8 Bao Z G. Tracy-Widom limit for Spearman’s rho. Http://mazgbao.people.ust.hk/TW%20law%20of%20Spearman28.pdf, 2019
9 Bao Z G, Lin L-C, Pan G M, et al. Spectral statistics of large dimensional Spearman’s rank correlation matrix and its application. Ann Statist, 2015, 43: 2588–2623
10 Bao Z G, Pan G M, Zhou W. Tracy-Widom law for the extreme eigenvalues of sample correlation matrices. Electron J Probab, 2012, 17: 1–32
11 Childs D R. Reduction of the multivariate normal integral to characteristic form. Biometrika, 1967, 54: 293–300
12 El Karoui N. Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. Ann Appl Probab, 2009, 19: 2362–2405
13 Eschfe F. On a method of determining correlation from the ranks of the variates. Scand Actuar J, 1924, 1924: 201–219
14 Féral D, Péché S. The largest eigenvalues of sample covariance matrices for a spiked population: Diagonal case. J Math Phys, 2009, 50: 073302
15 Gao J T, Han X, Pan G M, et al. High dimensional correlation matrices: The central limit theorem and its applications. J R Stat Soc Ser B Stat Methodol, 2017, 79: 677–693
16 Gray R M. Toeplitz and circulant matrices: A review. Found Trends Commun Informa Theory, 2006, 2: 155–240
Appendix A contains all the supporting lemmas and proofs.

Appendix A.1 Proof of Proposition 2.2

The following results show that $M_2$ and $M_3$ are concentrated on their population means, respectively.

**Lemma A.1.** Under the assumption of Proposition 2.2,
\[
\frac{1}{p} \mathbb{E} \|M_2\|_2^2 \leq \frac{4p^2}{3np(n-1)} + \frac{8}{np} \text{tr}(\Sigma_2^2) \to 0.
\]

**Lemma A.2.** Under the assumption of Proposition 2.2,
\[
\frac{1}{p} \mathbb{E} \|M_3 - \Sigma_3\|_2^2 \leq \frac{2p^2}{3np(n-1)} + \frac{32}{np} \text{tr} \{\Sigma_1(\Sigma_1 + \Sigma_2)\} \to 0.
\]

Equipped with these two results, we are now ready to prove Proposition 2.2. By [3, Corollary A.41],
\[
L^3(F_{K_n}, F_{W_n}) \leq \frac{1}{p} \|K_n - W_n\|_2 \leq \frac{6}{p} \|M_2\|_2^2 + \frac{3}{p} \|M_3 - \Sigma_3\|_2^2,
\]
which yields
\[
\mathbb{E} L^3(F_{K_n}, F_{W_n}) \to 0.
\]

The proof is completed.

It remains to prove the two auxiliary Lemmas A.1 and A.2.
Proof of Lemma A.1. Writing the kernel function

\[ h(i, j) = (A_i - A_j)(A_{ij} - A_i + A_j)^T, \]

we have

\[ M_2 = \frac{2}{n(n-1)} \sum_{1 < i < j \leq n} (A_i - A_j)(A_{ij} - A_i + A_j)^T \]

For the kernel function, we have the following properties:

- For the mean parts,
  \[ h(i, j) = h(j, i), \quad E[h(1, 2)] = 0, \quad E[h(1, 2) | x_1] = -E[A_2^T A_{12}^T | x_1] - \Sigma_2. \]

- For the Frobenius norm of the kernel function, we have
  \[ E_{tr}(h(1, 2)^T h(1, 2)) = E(||A_{12} - A_1 + A_2||^2_2 ||A_1 - A_2||^2_2). \]

Since

\[ ||A_{12} - A_1 + A_2||^2_\infty = ||\text{sign}(A_1 - A_2) - (A_1 - A_2)||^2_\infty \leq 1 \]

and \( ||A_i||^2_\infty \leq 1 \), we have

\[ ||A_{12} - A_1 + A_2||^2_2 \leq p \quad \text{and} \quad ||A_1 - A_2||^2_2 \leq 4p, \]

which yields

\[ E_{tr}(h(1, 2)^T h(1, 2)) \leq 4p^2. \]

- For the Frobenius norm of the conditional mean, we have
  \begin{align*}
  E_{tr}(h(1, 2)^T h(1, 3)) &= \text{tr}\{E(A_2 A_{12}^T + \Sigma_2)(A_3 A_{13}^T + \Sigma_2)\}
  = \text{cov}(A_{13}^T A_{12}, A_2^T A_3) - \text{tr}(\Sigma_2^2)
  \leq E[A_{13}^T A_{12}||A_2^T A_3|| \leq p(\text{var}(A_3^T A_2))^{1/2}
  = p(\text{tr}(\Sigma_2^2))^{1/2} \leq p^{3/2} ||\Sigma_2|| \leq p^{3/2} ||\Sigma_1||,
  \end{align*}

which yields

\[ E_{tr}(h(1, 2)^T h(1, 3)) \leq C p^{3/2}. \]

Putting together the pieces, we conclude that

\[ \frac{1}{p} E(M_2^T M_2) = \frac{1}{p} E\left\{ \frac{2}{n(n-1)} \sum_{1 < i < j \leq n} h(i, j) \right\}^T \left\{ \frac{2}{n(n-1)} \sum_{1 < k < l \leq n} h(k, l) \right\}
  = \frac{2}{np(n-1)} E_{tr}(h(1, 2)^T h(1, 2)) + \frac{4p(n-2)}{n(n-1)} E_{tr}(h(1, 2)^T h(1, 3))
  \leq \frac{8p^2}{np(n-1)} + \frac{4C p^{3/2}}{np} \to 0. \]

The proof is completed.
Proof of Lemma A.2. Recalling

\[ M_3 = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \epsilon_{ij} \epsilon_{ij}^T, \]

we have

\[
\begin{align*}
\operatorname{Etr}\{(M_3 - \Sigma_3)(M_3 - \Sigma_3)^T\} & = \frac{4}{n^2(n-1)^2} \sum_{i<j,k<l} \operatorname{E}\{\epsilon_{ik}^T(\epsilon_{ij} - \Sigma_3)\epsilon_{kl}\} \\
& = \frac{2}{n(n-1)} \{\operatorname{E}(\epsilon_{12}^T(\epsilon_{12} - \Sigma_3)\epsilon_{12}) + \frac{4(n-2)}{n(n-1)} \operatorname{E}(\epsilon_{12}^T(\epsilon_{13} - \Sigma_3)\epsilon_{12})\} \\
& = \frac{2}{n(n-1)} \{\operatorname{E}(\epsilon_{12}^T\epsilon_{12})^2 - \operatorname{tr}(\Sigma_3^2)\} + \frac{4(n-2)}{n(n-1)} \{\operatorname{E}(\epsilon_{12}^T\epsilon_{13})^2 - \operatorname{tr}(\Sigma_3^2)\}
\end{align*}
\]

For the first term, we have

\[ \epsilon_{12}^T\epsilon_{12} = \|A_{12} - A_1 + A_2\|_2^2 \leq p\|A_{12} - A_1 + A_2\|_\infty^2 \leq p, \]

which yields \(\operatorname{E}(\epsilon_{12}^T\epsilon_{12})^2 \leq p^2\). For the second term, we have

\[
(\epsilon_{12}^T\epsilon_{13})^2 \leq \{(A_{12} - A_1 + A_2)^T(A_{13} - A_1 + A_3)\}^2 \\
= \{(A_{12} - A_1)^T(A_{13} - A_1) + A_1^T(A_{13} - A_1) + A_2^T(A_{13} - A_1)\}^2 \\
\leq 4\{(A_{12} - A_1)^T(A_{13} - A_1)\}^2 + 4\{(A_1^T(A_{13} - A_1)\}^2 + 4\{(A_2^T(A_{13} - A_1)\}^2 + 4(A_2^T A_3)^2
\]

and

\[
\frac{1}{4} \operatorname{E}(\epsilon_{12}^T\epsilon_{13})^2 \leq \operatorname{E}\{(A_{12} - A_1)^T(A_{13} - A_1)\}^2 \leq \operatorname{E}\{(A_1^T A_{13} - A_1)\}^2 + \operatorname{E}\{(A_2^T A_{13})\}^2 \\
= \operatorname{E}\{(A_{12} - A_1)^T A_{13}\}^2 - \operatorname{E}\{(A_{12} - A_1)^T A_1\}^2 + 2\operatorname{tr}(\Sigma_2(\Sigma_1 - \Sigma_2)) + \operatorname{tr}(\Sigma_2^2) \\
\leq \operatorname{E}\{(A_{12} - A_1)^T A_{13}\}^2 + 2\operatorname{tr}(\Sigma_1\Sigma_2) \\
= \operatorname{var}(A_{12}^T A_{13}) - \operatorname{var}(A_{12}^T A_1) + 2\operatorname{tr}(\Sigma_1\Sigma_2) \\
\leq \operatorname{var}(A_{12}^T A_{13}) + 2pC^2.
\]

Finally, we have

\[
\frac{1}{p} \operatorname{Etr}\{(M_3 - \Sigma_3)(M_3 - \Sigma_3)^T\} \leq \frac{2p^2}{np(n-1)} + \frac{16}{np} \operatorname{var}(A_{12}^T A_{13}) + \frac{32}{np} pC^2 \to 0.
\]

The proof is completed. \(\square\)

Appendix A.2 Proof of Theorem 3.1

According to Proposition 2.2, it suffices to study the LSD of the following matrix \(M_n\):

\[ M_n = \frac{2}{n} \sum_{i=1}^{n} A_i A_i^T + \Sigma_3. \quad (A.1) \]

Let \(s_n(z)\) be the Stieltjes transform of \(F_{M_n}\). Then the convergence of \(F_{M_n}\) can be determined in three steps.

**Step 1.** Almost sure convergence of \(s_n(z) = E s_n(z)\).
Define
\[ M_{n,k} = \frac{2}{n} \sum_{i \neq k}^{n} A_i A_i^\top + \Sigma_3. \]

Let \( E_0(\cdot) \) be the expectation and \( E_k(\cdot) \) be the conditional expectation given \( A_1, \ldots, A_k \). From the martingale decomposition and the identity
\[
(M_n - z I_p)^{-1} A_k = \frac{(M_{n,k} - z I_p)^{-1} A_k}{1 + 2n^{-1} A_k^\top (M_{n,k} - z I_p)^{-1} A_k},
\]
we have
\[
s_n(z) - E s_n(z) = \frac{1}{p} \sum_{k=1}^{n} (E_k - E_{k-1}) [\text{tr}(M_n - z I_p)^{-1} - \text{tr}(M_{n,k} - z I_p)^{-1}] \]
\[
= -\frac{1}{p} \sum_{k=1}^{n} (E_k - E_{k-1}) \frac{2n^{-1} A_k^\top (M_{n,k} - z I_p)^{-2} A_k}{1 + 2n^{-1} A_k^\top (M_{n,k} - z I_p)^{-1} A_k} \]
\[
= \frac{1}{p} \sum_{k=1}^{n} (E_k - E_{k-1}) r_k.
\]

Since
\[
|r_k| \leq \frac{2n^{-1} A_k^\top (M_{n,k} - z I_p)^{-2} A_k}{|\Im(1 + 2n^{-1} A_k^\top (M_{n,k} - z I_p)^{-1} A_k)|} \leq \frac{1}{\Im(z)},
\]
\[ \{(E_k - E_{k-1}) r_k\} \] forms a bounded martingale difference sequence. Hence for any \( \ell > 1 \),
\[
E|s_n(z) - E s_n(z)|^\ell \leq K p^{-\ell/2} \left( \sum_{k=1}^{n} |(E_k - E_{k-1}) r_k|^2 \right)^{\ell/2} \leq K p^{-\ell/2} \Im(z)^{-\ell},
\]
which implies \( s_n(z) - E s_n(z) \to 0 \), almost surely.

**Step 2.** Convergence of \( E s_n(z) \).

Define
\[
x_n = x_n(z) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{1 + 2n^{-1} E[\text{tr}\{(M_{n,k} - z I_p)^{-1} \Sigma_2\}]}.
\]

Starting from the identity
\[
(2x_n \Sigma_2 + \Sigma_3 - z I_p)^{-1} - (M_n - z I_p)^{-1} \]
\[
= (2x_n \Sigma_2 + \Sigma_3 - z I_p)^{-1} \left( \frac{2}{n} \sum_{k=1}^{n} A_k A_k - 2x_n \Sigma_2 \right) (M_n - z I_p)^{-1},
\]
and then taking trace on both sides, by (A.2), we have
\[
\frac{1}{p} \text{tr}(2x_n \Sigma_2 + \Sigma_3 - z I_p)^{-1} - s_n(z)
\]
\[
= \frac{2}{np} \sum_{k=1}^{n} A_k^\top (M_{n,k} - z I_p)^{-1} (2x_n \Sigma_2 + \Sigma_3 - z I_p)^{-1} A_k \]
\[
- \frac{2x_n}{p} \text{tr}\{(2x_n \Sigma_2 + \Sigma_3 - z I_p)^{-1} \Sigma_2 (M_n - z I_p)^{-1} \}
\]
\[
\overset{\text{def}}{=} \frac{2}{n} \sum_{k=1}^{n} \frac{d_k}{1 + 2n^{-1} E[\text{tr}\{(M_{n,k} - z I_p)^{-1} \Sigma_2\}]}.
\]
where
\[ d_k = \frac{1 + 2n^{-1}E[\text{tr}((M_{n,k} - zI_p)^{-1}\Sigma_2)]]}{1 + 2n^{-1}A_k^T(M_{n,k} - zI_p)^{-1}A_k} \cdot \frac{1}{p} A_k^T(M_{n,k} - zI_p)^{-1}(2x_n\Sigma_2 + \Sigma_3 - zI_p)^{-1}A_k - \frac{1}{p} \text{tr}((M_n - zI_p)^{-1}(2x_n\Sigma_2 + \Sigma_3 - zI_p)^{-1}\Sigma_2). \]

We decompose
\[ d_k \overset{\text{def}}{=} d_{k1} + d_{k2} + d_{k3}, \]
where
\[ d_{k1} = \frac{1}{p} \text{tr}((M_{n,k} - zI_p)^{-1}(2x_n\Sigma_2 + \Sigma_3 - zI_p)^{-1}\Sigma_2) \]
\[ - \frac{1}{p} \text{tr}((M_n - zI_p)^{-1}(2x_n\Sigma_2 + \Sigma_3 - zI_p)^{-1}\Sigma_2), \]
\[ d_{k2} = \frac{1}{p} A_k^T(M_{n,k} - zI_p)^{-1}(2x_n\Sigma_2 + \Sigma_3 - zI_p)^{-1}A_k \]
\[ - \frac{1}{p} \text{tr}((M_{n,k} - zI_p)^{-1}(2x_n\Sigma_2 + \Sigma_3 - zI_p)^{-1}\Sigma_2), \]
\[ d_{k3} = -\frac{2A_k^T(M_{n,k} - zI_p)^{-1}(2x_n\Sigma_2 + \Sigma_3 - zI_p)^{-1}A_k}{pn(1 + 2n^{-1}A_k^T(M_{n,k} - zI_p)^{-1}A_k)} \]
\[ \times \{ A_k^T(M_{n,k} - zI_p)^{-1}A_k - E[\text{tr}((M_{n,k} - zI_p)^{-1}\Sigma_2)]\}. \]

For the term \( d_{k1} \), we have
\[ |d_{k1}| = \left| \frac{2}{pn} \text{tr}((M_n - zI_p)^{-1}A_kA_k^T(M_{n,k} - zI_p)^{-1}(2x_n\Sigma_2 + \Sigma_3 - zI_p)^{-1}\Sigma_2) \right| \]
\[ = \left| \frac{2n^{-1} - p]{\text{tr}((M_{n,k} - zI_p)^{-1}(2x_n\Sigma_2 + \Sigma_3 - zI_p)^{-1}\Sigma_2})}{1 + 2n^{-1}A_k^T(M_{n,k} - zI_p)^{-1}A_k} \right| \]
\[ \leq Kp^{-1}\Theta(z)^{-2}, \quad (A.6) \]

and then its contribution to \((A.5)\) can be bounded as
\[ \left| \frac{2}{n} \sum_{k=1}^n \frac{d_{k1}}{1 + 2n^{-1}E[\text{tr}((M_{n,k} - zI_p)^{-1}\Sigma_2)]]} \right| \leq \frac{2}{n} \sum_{k=1}^n \left| \frac{|d_{k1}|}{1 + 2n^{-1}E[\text{tr}((M_{n,k} - zI_p)^{-1}\Sigma_2)]]} \right| \]
\[ \leq Kp^{-1}\Theta(z)^{-2} \rightarrow 0. \]

For the term \( d_{k2} \), we have
\[ E(d_{k2}) = 0. \quad (A.7) \]

For the term \( d_{k3} \), we have the first part bounded by
\[ \left| -\frac{2A_k^T(M_{n,k} - zI_p)^{-1}(2x_n\Sigma_2 + \Sigma_3 - zI_p)^{-1}A_k}{pn(1 + 2n^{-1}A_k^T(M_{n,k} - zI_p)^{-1}A_k)} \right| \leq Kp^{-1}n^{-1}\Theta(z)^{-2}\|A_k\|^2. \]

So we have
\[ |E(d_{k3})|^2 \leq Kp^{-2}n^{-2}E\|A_k\|^4E[A_k^T(M_{n,k} - zI_p)^{-1}A_k - E[\text{tr}((M_{n,k} - zI_p)^{-1}\Sigma_2)]]^2]. \]

According to Lemmas A.4 and A.5, we have
\[ E\|A_k\|^4 = \text{var}(A_k^T A_k) + \{E(A_k^T A_k)\}^2 \leq \text{tr}(\Sigma_k^2) - \text{tr}(\Sigma_k^2) + \{\text{tr}(\Sigma_k)\}^2 = O(p^2), \]
\[ E[A_k^T(M_{n,k} - zI_p)^{-1}A_k - E[\text{tr}((M_{n,k} - zI_p)^{-1}\Sigma_2)]]^2 \]
which gives
\[ |E(d_{k3})|^2 \leq Kp^{-1}. \]  
\hspace{1cm} (A.8)

Combining (A.5)–(A.8), we have
\[ E_{sn}(z) \to \frac{1}{p} \text{tr}(2x\Sigma_2 + \Sigma_3 - zI_p)^{-1} \overset{\text{def}}{=} s(z), \]  
\hspace{1cm} (A.9)

where \( z \) is the limit of \( x_n \).

Next, we give the equation that \( x \) satisfies, starting from the quantity
\[ n^{-1}E[\text{tr}((M_{n,k} - zI_p)^{-1}\Sigma_2)] \]

in the denominator in (A.4). By the similar arguments to (A.5), we can replace the term \((M_{n,k} - zI_p)^{-1}\) by that of \((2x_n\Sigma_2 + \Sigma_3 - zI_p)^{-1}\), which leads to the equation
\[ x = \frac{1}{1 + 2\lim_{n \to \infty} n^{-1}\text{tr}(2x\Sigma_2 + \Sigma_3 - zI_p)^{-1}\Sigma_2}. \]  
\hspace{1cm} (A.10)

(A.9) and (A.10) are exactly the equations (3.5) and (3.4) established in Theorem 3.1.

**Step 3.** The uniqueness of the solution \( s(z) \).

We only have to show that the solution \( x(z) \) to (3.4), if it exists, is unique in \( \mathbb{C}^- \). Now suppose that we have two solutions \( x_1 = x_1(z) \) and \( x_2 = x_2(z) \in \mathbb{C}^- \) to (3.4) for a common \( z \in \mathbb{C}^+ \). Then we can obtain
\[ \frac{1}{x_1} - \frac{1}{x_2} = \lim_{n \to \infty} \frac{2}{n} \text{tr}[(\Sigma_3 + 2x_1\Sigma_2 - zI_p)^{-1}(2x_2\Sigma_2 - 2x_1\Sigma_2)(\Sigma_3 + 2x_2\Sigma_2 - zI_p)^{-1}\Sigma_2]. \]

If \( x_1 \neq x_2 \), then
\[ 1 = \lim_{n \to \infty} \frac{4}{n} \text{tr}[x_1\Sigma_2^{-1/2}(\Sigma_3 + 2x_1\Sigma_2 - zI_p)^{-1}\Sigma_2^{-1/2}: x_2\Sigma_2^{-1/2}(\Sigma_3 + 2x_2\Sigma_2 - zI_p)^{-1}\Sigma_2^{-1/2}] 
\[ = \lim_{n \to \infty} \frac{4}{n} \text{tr}[x_1(\Sigma_2^{-1/2}(\Sigma_3 - zI_p))\Sigma_2^{-1/2} + 2x_1I_p)^{-1}: x_2(\Sigma_2^{-1/2}(\Sigma_3 - zI_p))\Sigma_2^{-1/2} + 2x_2I_p)^{-1}] 
\[ = \lim_{n \to \infty} \frac{4}{n} \text{tr}[x_1(Q(z) + 2x_1I_p)^{-1}: x_2(Q(z) + 2x_2I_p)^{-1}], \]

where \( Q(z) \) is defined as
\[ Q(z) = \Sigma_2^{-1/2}(\Sigma_3 - zI_p)\Sigma_2^{-1/2}. \]

By the Cauchy-Schwarz inequality, we have
\[ 1 \leq \left\{ \lim_{n \to \infty} \frac{4|x_1|^2}{n} \text{tr}[(Q(z) + 2x_1I_p)(Q(z) + 2x_1I_p)]^{-1} \right. 
\[ \times \lim_{n \to \infty} \frac{4|x_2|^2}{n} \text{tr}[(Q(z) + 2x_2I_p)(Q(z) + 2x_2I_p)]^{-1}\right\}^{1/2}. \]  
\hspace{1cm} (A.11)

On the other hand, denote the eigen-decomposition of \( Q(z) \) by
\[ Q(z) = \sum_{k=1}^{p} \lambda_k v_k v_k^T. \]

Then we have
\[ \lambda_k = v_k^T \Sigma_2^{-1/2}(\Sigma_3 - zI_p)\Sigma_2^{-1/2}v_k = v_k^T \Sigma_2^{-1/2}\Sigma_3\Sigma_2^{-1/2}v_k - zv_k^T \Sigma_2^{-1}v_k, \]
which yields
\[ \Im(\lambda_k) = -\Im(z)\overline{v}_k^T\Sigma_2^{-1}v_k < 0. \]

Then taking the imaginary part in (3.4), we have
\[ \frac{\Im(\bar{x})}{|x|^2} = \lim_{n \to \infty} \frac{2}{n} \Im(\text{tr}(Q(z) + 2xI_p)^{-1}) \]
\[ = \lim_{n \to \infty} \frac{2}{n} \sum_{k=1}^p \frac{\Im(\lambda_k) + 2\Im(\bar{x})}{|\lambda_k + 2x|^2} \]
\[ > \lim_{n \to \infty} \frac{4}{n} \sum_{k=1}^p \frac{\Im(\bar{x})}{|\lambda_k + 2x|^2}. \]

Since \( \Im(\bar{x}) > 0 \), the above inequality yields that
\[ \frac{1}{|x|^2} > \lim_{n \to \infty} \frac{4}{n} \sum_{k=1}^p \frac{1}{|\lambda_k + 2x|^2} = \lim_{n \to \infty} \frac{4}{n} \text{tr}(Q(z) + 2xI_p)^{-1}(Q(z) + 2xI_p)^{-1}, \]
which is a contradiction to (A.11). This contradiction proves that \( x_1 = x_2 \) and hence (3.4) has at most one solution in \( \mathbb{C}^- \). The proof of this theorem is then completed.

**Appendix A.3 Proof of Proposition 4.2**

The proof is checking Assumption (A) for the normal distribution which is summarized in Lemmas A.4 and A.5. Before proceeding, we need the following variance result for Kendall’s correlation.

**Lemma A.3** (See [13]). Considering a multivariate normal distribution
\[
\begin{pmatrix}
  z_1 \\
  z_2 \\
  z_3 \\
  z_4
\end{pmatrix}
\sim
N
\left(
\begin{pmatrix}
  0 \\
  0 \\
  0 \\
  0
\end{pmatrix},
\begin{pmatrix}
  1/2 & \rho/2 & 1/2 \\
  \rho/2 & 1 & \rho/2 \\
  1/2 & \rho & 1/2 \\
  \rho & 1/2 & 1
\end{pmatrix}
\right),
\]
where \( \rho \in (-1, 1) \), we have
\[
E\left\{\prod_{j=1}^4 \text{sign}(z_j)\right\} = \left( \frac{2}{\pi} \arcsin \rho \right)^2 - \left( \frac{2}{\pi} \arcsin (\rho/2) \right)^2 + \frac{1}{9}. \tag{A.12}
\]

**Lemma A.4.** Assuming \( x_1, \ldots, x_n \overset{i.i.d.}{\sim} N(0, \Sigma) \), we have
\[
\text{var}(A_{12}^T A_{13}) = \text{tr}(\Sigma_1^2) - \text{tr}(\Sigma_2^2). \tag{A.13}
\]

**Proof.** For the covariance part, we have
\[ A_{ij} = \text{sign}(x_i - x_j) \overset{d}{=} \text{sign}(x), \]
and Grothendieck’s identity shows that for any bi-variate normal vector \((z_1, z_2)^T\),
\[ E[\text{sign}(z_1)\text{sign}(z_2)] = \frac{2}{\pi} \arcsin[\text{corr}(z_1, z_2)]. \]

Thus,
\[ \text{cov}(A_{ij}) = \text{cov}[\text{sign}(x)] = \frac{2}{\pi} \arcsin(\Sigma) = \Sigma_1. \]
For $\mathbf{A}_i$,
\[
\text{cov}(\mathbf{A}_{ij}, \mathbf{A}_i) = E(\mathbf{A}_{ij} \mathbf{A}_i^T) = E(\mathbf{A}_i \mathbf{A}_i^T) = \text{cov}(\mathbf{A}_i)
\]
and
\[
\text{cov}(\mathbf{A}_i) = E[\text{sign}(\mathbf{x}_1 - \mathbf{x}_2)\text{sign}(\mathbf{x}_1 - \mathbf{x}_3)^T] = E\left[\text{sign}\left(\frac{\mathbf{x}_1 - \mathbf{x}_2}{\sqrt{2}}\right)\text{sign}\left(\frac{\mathbf{x}_1 - \mathbf{x}_3}{\sqrt{2}}\right)^T\right].
\]

For the enlarged random vector, we have
\[
\frac{1}{\sqrt{2}}\begin{pmatrix} \mathbf{x}_1 - \mathbf{x}_2 \\ \mathbf{x}_1 - \mathbf{x}_3 \end{pmatrix} \sim N\left(0, \begin{pmatrix} \Sigma & \Sigma/2 \\ \Sigma/2 & \Sigma \end{pmatrix}\right),
\]
and thus,
\[
\text{cov}(\mathbf{A}_{ij}, \mathbf{A}_i) = \text{cov}(\mathbf{A}_i) = \frac{2}{\pi} \arcsin(\Sigma/2).
\]

Next, we derive the explicit result for $\text{var}(\mathbf{A}_{12}^T \mathbf{A}_{13})$. Since
\[
\text{var}(\mathbf{A}_{12}^T \mathbf{A}_{13}) = E(\mathbf{A}_{12}^T \mathbf{A}_{13})^2 - \text{tr}^2(\Sigma_2)
\]
\[
= E\left\{\sum_{i=1}^{p} \text{sign}(\mathbf{x}_{1i} - \mathbf{x}_{2i})\text{sign}(\mathbf{x}_{1i} - \mathbf{x}_{3i})\right\}^2 - \text{tr}^2(\Sigma_2)
\]
\[
= E\left\{\sum_{i,j=1}^{p} \text{sign}(\mathbf{x}_{1i} - \mathbf{x}_{2i})\text{sign}(\mathbf{x}_{1i} - \mathbf{x}_{3i})\text{sign}(\mathbf{x}_{1j} - \mathbf{x}_{2j})\text{sign}(\mathbf{x}_{1j} - \mathbf{x}_{3j})\right\} - \frac{1}{9}p^2,
\]
and for any $(i,j)$,
\[
\frac{1}{\sqrt{2}}\begin{pmatrix} \mathbf{x}_{1i} - \mathbf{x}_{2i} \\ \mathbf{x}_{1i} - \mathbf{x}_{3i} \\ \mathbf{x}_{1j} - \mathbf{x}_{2j} \\ \mathbf{x}_{1j} - \mathbf{x}_{3j} \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}\right),
\]
by Lemma A.3, we have
\[
\text{var}(\mathbf{A}_{12}^T \mathbf{A}_{13}) = \sum_{i,j=1}^{p} \left\{\frac{2}{\pi} \arcsin(\Sigma_{ij})\right\}^2 - \left\{\frac{2}{\pi} \arcsin(\Sigma_{ij}/2)\right\}^2 + \frac{1}{9} - \frac{1}{9}p^2
\]
\[
= \text{tr}(\Sigma_2^2) - \text{tr}(\Sigma_2^2).
\]
The proof is completed. \qed

**Lemma A.5.** Let $\mathbf{A} = 2\Phi(\mathbf{x}) - 1$, where $\mathbf{x} \sim N_p(0, \Sigma)$. Then for any non-random $p \times p$ matrix $\mathbf{B}$, we have
\[
\text{var}(\mathbf{A}^T \mathbf{B} \mathbf{A}) \leq 3\|\Sigma\|\text{tr}(\mathbf{B} \Sigma_2 \mathbf{B}^T).
\]

**Proof.** For $\mathbf{x} = (x_1, \ldots, x_p)^T$, we define a function
\[
g(\mathbf{x}) = (2\Phi(\mathbf{x}) - 1)^T \mathbf{B} (2\Phi(\mathbf{x}) - 1)
\]
\[
= \sum_{i=1}^{p} b_{ii}(2\Phi(x_i) - 1)^2 + 2 \sum_{i<j} b_{ij} (2\Phi(x_i) - 1)(2\Phi(x_j) - 1).
\]
Direct calculations can show that
\[
\nabla g(\mathbf{x}) = 4\text{diag}(f(x_1), \ldots, f(x_p))\mathbf{B}\Phi(\mathbf{x}),
\]
where $f(x) = \Phi(x) - \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.
where
\[ f(x) = \exp(-x^2/2)/\sqrt{2\pi}. \]

When \( x \sim N(0, \Sigma) \), by the Gaussian Poincaré inequality, we have
\[
\text{var}(g(x)) = \text{var}(A^T B A) \leq \text{Etr}\{\nabla g(x)^T \Sigma \nabla g(x)\} \leq \frac{16}{2\pi} ||\Sigma|| \cdot \text{E} \{\text{tr}(A^T B^T BA)\} \leq 3 ||\Sigma|| \cdot \text{tr}(B \Sigma_2 B^T).
\]

The proof is completed. \( \square \)

**Appendix A.4 Proof of Proposition 4.4**

We consider a more general matrix
\[
\Sigma = \begin{pmatrix}
  a & b & b & \cdots & b \\
  b & a & b & \cdots & b \\
  b & b & a & \cdots & b \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  b & b & b & \cdots & a
\end{pmatrix},
\]
whose eigenvalues are
\[
\lambda_k = a + 2b \cos \frac{k\pi}{p+1}, \quad k = 1, \ldots, p,
\]
and corresponding eigenvectors are
\[
u_k = \sqrt{\frac{2}{p+1}} \left( \sin \frac{k\pi}{p+1}, \sin \frac{2k\pi}{p+1}, \ldots, \sin \frac{pk\pi}{p+1} \right)^T.
\]

Then
\[
\frac{1}{p} \text{tr}(\Sigma)^{-1} = \frac{1}{p} \sum_{k=1}^{p} \frac{1}{a + 2b \cos \frac{k\pi}{p+1}} \rightarrow \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{a + 2b \cos \theta} d\theta = \frac{1}{\sqrt{a^2 - 4b^2}}
\]
where the limit is due to the Szegö theorem [16] and the integral can be calculated through the residue theorem from complex analysis.

For \( \Sigma = \Sigma(\rho) \), we have
\[
\Sigma_{2,i}^{i,i} = \frac{1}{3}, \quad \Sigma_{2,i}^{i,i+1} = \Sigma_{2,i}^{i-1,i} = \frac{2}{\pi} \arcsin \frac{\rho}{2}
\]
and
\[
\Sigma_{3,i}^{i,i} = \frac{1}{3}, \quad \Sigma_{3,i}^{i,i+1} = \Sigma_{3,i}^{i-1,i} = \frac{2}{\pi} \arcsin \rho - \frac{4}{\pi} \arcsin \frac{\rho}{2}.
\]

Thus,
\[
(\Sigma_3 + 2x \Sigma_2 - z I_p)^{i,i} = \frac{1}{3} + \frac{2x}{3} - z,
\]
\[
(\Sigma_3 + 2x \Sigma_2 - z I_p)^{i,i+1} = (\Sigma_3 + 2x \Sigma_2 - z I_p)^{i-1,i} = \frac{2}{\pi} \arcsin \rho + \frac{4(x-1)}{\pi} \arcsin \frac{\rho}{2}.
\]

This yields
\[
\lim_{p \to \infty} \frac{1}{p} \text{tr}(\Sigma_3 + 2x \Sigma_2 - z I_p)^{-1} = \frac{1}{\sqrt{(\frac{1}{3} + \frac{2x}{3} - z)^2 - 4(\frac{2}{\pi} \arcsin \rho + \frac{4(x-1)}{\pi} \arcsin \frac{\rho}{2})^2}}
\]
\[
\frac{1}{p} \text{tr}[(\Sigma_3 + 2x \Sigma_2 - z I_p)^{-1} \Sigma_2] \\
= \frac{1}{p} \sum_{k=1}^{p} \frac{\frac{1}{3} + 2x - z + 2(\frac{2}{x} \arcsin \frac{\rho}{2} \cos \frac{k\pi}{p+1})}{\sqrt{\left(\frac{1}{3} + \frac{2x}{3} - z\right)^2 - 4\left(\frac{2}{x} \arcsin \frac{\rho}{2} + \frac{4(x-1)}{x} \arcsin \frac{\rho}{2}\right)^2}}.
\]

The proof is completed.