Bernoulli **23**(1), 2017, 23–57
DOI: 10.3150/15-BEJ702

Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution

FANG HAN[1] and HAN LIU[2]

[1]*Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA. E-mail: fhan@jhu.edu*
[2]*Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA. E-mail: hanliu@princeton.edu*

Correlation matrices play a key role in many multivariate methods (e.g., graphical model estimation and factor analysis). The current state-of-the-art in estimating large correlation matrices focuses on the use of Pearson's sample correlation matrix. Although Pearson's sample correlation matrix enjoys various good properties under Gaussian models, it is not an effective estimator when facing heavy-tailed distributions. As a robust alternative, Han and Liu [*J. Am. Stat. Assoc.* **109** (2015) 275–287] advocated the use of a transformed version of the Kendall's tau sample correlation matrix in estimating high dimensional latent generalized correlation matrix under the transelliptical distribution family (or elliptical copula). The transelliptical family assumes that after unspecified marginal monotone transformations, the data follow an elliptical distribution. In this paper, we study the theoretical properties of the Kendall's tau sample correlation matrix and its transformed version proposed in Han and Liu [*J. Am. Stat. Assoc.* **109** (2015) 275–287] for estimating the population Kendall's tau correlation matrix and the latent Pearson's correlation matrix under both spectral and restricted spectral norms. With regard to the spectral norm, we highlight the role of "effective rank" in quantifying the rate of convergence. With regard to the restricted spectral norm, we for the first time present a "sign sub-Gaussian condition" which is sufficient to guarantee that the rank-based correlation matrix estimator attains the fast rate of convergence. In both cases, we do not need any moment condition.

Keywords: double asymptotics; elliptical copula; Kendall's tau correlation matrix; rate of convergence; transelliptical model

1. Introduction

Covariance and correlation matrices play a central role in multivariate analysis. An efficient estimation of covariance/correlation matrix is a major step in conducting many methods, including principal component analysis (PCA), scale-invariant PCA, graphical model estimation, discriminant analysis, and factor analysis. Large covariance/correlation matrix estimation receives a lot of attention in high dimensional statistics. This is partially because the sample covariance/correlation matrix is an inconsistent estimator when $d/n \nrightarrow 0$ (d and n represent the dimensionality and sample size).

Given n observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of a d-dimensional random vector $\mathbf{X} \in \mathbb{R}^d$ with the population covariance matrix $\mathbf{\Omega}$, let $\widehat{\mathbf{S}}$ be the Pearson's sample covariance matrix calculated based on

1350-7265

$\mathbf{x}_1, \ldots, \mathbf{x}_n$. For theoretical analysis, we adopt a similar double asymptotic framework as in Bickel and Levina [4], where we write d to be the abbreviation of d_n, which changes with n. Under this double asymptotic framework, where both the dimension d and sample size n can increase to infinity, Johnstone [23], Baik and Silverstein [1] and Jung and Marron [24] pointed out settings such that, even when $\mathbf{X}$ follows a Gaussian distribution with identity covariance matrix, $\widehat{\mathbf{S}}$ is an inconsistent estimator of $\mathbf{\Sigma}$ under spectral norm. In other words, letting $\|\cdot\|_2$ denote the spectral norm of a matrix, typically for $(n, d) \to \infty$, we have

$$\|\widehat{\mathbf{S}} - \mathbf{\Omega}\|_2 \nrightarrow 0.$$

This observation motivates different versions of sparse covariance/correlation matrix estimation methods. See, for example, banding method (Bickel and Levina [4]), tapering method (Cai et al. [9], Cai and Zhou [10]), and thresholding method (Bickel and Levina [5]). However, although the regularization methods exploited are different, they all use the Pearson's sample covariance/correlation matrix as a pilot estimator, and accordingly the performance of the estimators relies on existence of higher order moments of the data. For example, letting $\|\cdot\|_{\max}$ and $\|\cdot\|_{2,s}$ denote the element-wise supremum norm and restricted spectral norm (detailed definitions provided later), in proving

$$\|\widehat{\mathbf{S}} - \mathbf{\Omega}\|_{\max} = \mathrm{O}_P\left(\sqrt{\frac{\log d}{n}}\right) \quad \text{or} \quad \|\widehat{\mathbf{S}} - \mathbf{\Omega}\|_{2,s} = \mathrm{O}_P\left(\sqrt{\frac{s\log(d/s)}{n}}\right) \tag{1.1}$$

(here, d and s are the abbreviation of d_n and s_n and $\mathrm{O}_P(\cdot)$ is defined to represent the stochastic order with regard to n), it is commonly assumed that, for $d = 1, 2, \ldots, \mathbf{X} = (X_1, \ldots, X_d)^T$ satisfies the following sub-Gaussian condition:

$$\begin{aligned} &\text{(marginal sub-Gaussian)} \quad \mathbb{E}\exp(tX_j) \le \exp\left(\frac{\sigma^2 t^2}{2}\right) \qquad \text{for all } j \in \{1, \ldots, d\} \quad \text{or} \\ &\text{(multivariate sub-Gaussian)} \quad \mathbb{E}\exp\bigl(t\mathbf{v}^T\mathbf{X}\bigr) \le \exp\left(\frac{\sigma^2 t^2}{2}\right) \qquad \text{for all } \mathbf{v} \in \mathbb{S}^{d-1}, \end{aligned} \tag{1.2}$$

for some absolute constant $\sigma^2 > 0$. Here, $\mathbb{S}^{d-1}$ is the d-dimensional unit sphere in $\mathbb{R}^d$.

The moment conditions in (1.2) are not satisfied for many distributions. To elaborate how strong this condition is, we consider the student's t distribution. Assuming that T follows a student's t distribution with degree of freedom ν, it is known (Hogg and Craig [20]) that

$$\mathbb{E}T^{2k} = \infty \qquad \text{for } k \ge \nu/2.$$

Recently, Han and Liu [17] advocated to use the transelliptical distribution for modeling and analyzing complex and noisy data. They exploited a transformed version of the Kendall's tau sample correlation matrix $\widehat{\mathbf{\Sigma}}$ to estimate the latent Pearson's correlation matrix $\mathbf{\Sigma}$. The transelliptical family assumes that, after a set of unknown marginal transformations, the data follow an elliptical distribution. This family is closely related to the elliptical copula and contains many well-known distributions, including multivariate Gaussian, rank-deficient Gaussian, multivariate-t, Cauchy, Kotz, logistic, etc. Under the transelliptical distribution, without any moment constraint, they showed that a transformed Kendall's tau sample correlation matrix $\widehat{\mathbf{\Sigma}}$ approximates the latent Pearson's correlation matrix $\mathbf{\Sigma}$ in a parametric rate:

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max} = \mathrm{O}_P\left(\sqrt{\frac{\log d}{n}}\right), \tag{1.3}$$

which attains the minimax rate of convergence.

Although (1.3) is inspiring, in terms of theoretical analysis of many multivariate methods, the rates of convergence under spectral norm and restricted spectral norm are more desired. For example, Bickel and Levina [5] and Yuan and Zhang [37] showed that the performances of principal component analysis and a computationally tractable sparse PCA method are determined by the rates of convergence for the plug-in matrix estimators under spectral and restricted spectral norms. A trivial extension of (1.3) gives us that

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 = \mathrm{O}_P\left(d\sqrt{\frac{\log d}{n}}\right) \quad \text{and} \quad \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{2,s} = \mathrm{O}_P\left(s\sqrt{\frac{\log d}{n}}\right),$$

which are both not tight compared to the parametric rates (for more details, check Lounici [30] and Bunea and Xiao [7] for results under the spectral norm, and Vu and Lei [34] for results under the restricted spectral norm).

In this paper, we push the results in Han and Liu [17] forward, providing improved results of the transformed Kendall's tau correlation matrix under both spectral and restricted spectral norms. We consider the statistical properties of the Kendall's tau sample correlation matrix $\widehat{\mathbf{T}}$ in estimating the Kendall's tau correlation matrix $\mathbf{T}$, and the transformed version $\widehat{\mathbf{\Sigma}}$ in estimating $\mathbf{\Sigma}$.

First, we considering estimating the Kendall's tau correlation matrix $\mathbf{T}$ itself. Estimating Kendall's tau is of its self-interest. For example, Embrechts et al. [12] claimed that in many cases in modeling dependence Pearson's correlation coefficient "might prove very misleading" and advocated to use the Kendall's tau correlation coefficient as the "perhaps best alternatives to the linear correlation coefficient as a measure of dependence for nonelliptical distributions." In estimating $\mathbf{T}$, we show that, without any condition, for any continuous random vector $\mathbf{X}$,

$$\|\widehat{\mathbf{T}} - \mathbf{T}\|_2 = \mathrm{O}_P\left(\|\mathbf{T}\|_2\sqrt{\frac{r_e(\mathbf{T})\log d}{n}}\right),$$

where $r_e(\mathbf{T}) := \mathrm{Tr}(\mathbf{T})/\|\mathbf{T}\|_2$ is called effective rank. Moreover, we provide a new term called "sign sub-Gaussian condition," under which we have

$$\|\widehat{\mathbf{T}} - \mathbf{T}\|_{2,s} = \mathrm{O}_P\left(\|\mathbf{T}\|_2\sqrt{\frac{s\log d}{n}}\right).$$

Secondly, under the transelliptical family, we consider estimating the Pearson's correlation matrix $\mathbf{\Sigma}$ of the latent elliptical distribution using the transformed Kendall's tau sample correlation matrix $\widehat{\mathbf{\Sigma}} = [\sin(\frac{\pi}{2}\widehat{\mathbf{T}}_{jk})]$. Without any moment condition, we show that, as long as $\mathbf{X}$ belongs to the transelliptical family,

$$\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\|_2=\mathrm{O}_P\left(\|\mathbf{\Sigma}\|_2\left\{\sqrt{\frac{r_e(\mathbf{\Sigma})\log d}{n}}+\frac{r_e(\mathbf{\Sigma})\log d}{n}\right\}\right),$$

which attains the nearly optimal rate of convergence obtained in Lounici [30] and Bunea and Xiao [7]. Moreover, provided that the sign sub-Gaussian condition is satisfied, we have

$$\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\|_{2,s}=\mathrm{O}_P\left(\|\mathbf{\Sigma}\|_2\sqrt{\frac{s\log d}{n}}+\frac{s\log d}{n}\right),$$

which attains the nearly optimal rate of convergence obtained in Vu and Lei [34].

1.1. Discussion with related works

Our work is related to a vast literature in large covariance matrix estimation, with different settings of sparsity assumptions (Cai et al. [8,9], Cai and Zhou [10], Vu and Lei [34]), or without any sparsity assumption (Bunea and Xiao [7], Lounici [30]). In particular, this work is closely related to Lounici [30] and Bunea and Xiao [7] with regard to the theoretical analysis of the spectral norm convergence, and the work of Vu and Lei [34] with regard to the theoretical analysis of the restricted spectral norm convergence.

However, there are various new contributions made in this paper given the aforementioned results. We emphasize the advantage of rank-based statistics over moment-based statistics. One new message delivered in this paper is, via resorting to the rank-based statistics, the statistical efficiency attained by the aforementioned methods under some stringent moment constraints, can be attained under some more flexible models. Moreover, we believe that the technical developments built in this paper, including the analysis of U-statistics, the concentration of matrix-value functions, and the verification of the sign sub-Gaussian condition for several particular models, are distinct from the existing literature and of self-interest.

Our work is also closely related to an expanding literature in extending copula models to the high dimensional settings. These include the use of the nonparanormal (Gaussian copula) and the transelliptical (elliptical copula) distribution families. Methodologically, the Spearman's rho is recommended in the analysis of the nonparanormal family for conducting graphical model estimation (Liu et al. [27], Xue and Zou [36]), classification (Han et al. [18]), and PCA (Han and Liu [16]). The Kendall's tau is recommended in the analysis of the transelliptical family for conducting graphical model estimation (Liu et al. [28]) and PCA (Han and Liu [17]).

Our work is motivated from the aforementioned results. But, different from the existing ones, we give a more general study on the convergence of the Kendall's tau matrix itself, and provide more insights into the rank-based statistics. We characterize three types of convergence with regard to the Kendal's tau matrix $\widehat{\mathbf{T}}$ and its transformed version $\widehat{\mathbf{\Sigma}}$: The element-wise supremum norm ($\ell_{\max}$), the spectral norm (ℓ_2), and the restricted spectral norm ($\ell_{2,s}$). In comparison, the existing results only exploited the $\ell_{\max}$ convergence result, which we find is not sufficient in showing the statistical efficiency of many rank-based methods. It is also worth noting that the new theories developed here with regard to the ℓ_2 and $\ell_{2,s}$ convergence have broad implications.

They can be easily applied to the study of factor model, sparse PCA, robust regression and many other methods, and can lead to more refined statistical analysis.

In an independent work, Wegkamp and Zhao [35] proposed to use the same transformed Kendall's tau correlation coefficient estimator to analyze the elliptical copula factor model and proved a similar spectral norm convergence result as in Theorem 3.1 of this paper. The proofs are different and these two papers are independent work.

1.2. Notation system

Let $\mathbf{M} = [\mathbf{M}_{ij}] \in \mathbb{R}^{d\times d}$ and $\mathbf{v} = (v_1, \dots, v_d)^T \in \mathbb{R}^d$. We denote $\mathbf{v}_I$ to be the subvector of $\mathbf{v}$ whose entries are indexed by a set I. We also denote $\mathbf{M}_{I,J}$ to be the submatrix of $\mathbf{M}$ whose rows are indexed by I and columns are indexed by J. Let $\mathbf{M}_{I*}$ and $\mathbf{M}_{*J}$ be the submatrix of $\mathbf{M}$ with rows indexed by I, and the submatrix of $\mathbf{M}$ with columns indexed by J. Let $\mathrm{supp}(\mathbf{v}) := \{j\colon v_j \neq 0\}$. For $0 < q < \infty$, we define the ℓ_0, ℓ_q, and ℓ_∞ vector (pseudo-)norms as

$$\|\mathbf{v}\|_0 := \mathrm{card}\big(\mathrm{supp}(\mathbf{v})\big), \qquad \|\mathbf{v}\|_q := \left(\sum_{i=1}^d |v_i|^q\right)^{1/q} \quad \text{and} \quad \|\mathbf{v}\|_\infty := \max_{1\le i\le d} |v_i|.$$

Let $\lambda_j(\mathbf{M})$ be the jth largest eigenvalue of $\mathbf{M}$ and $\mathbf{\Theta}_j(\mathbf{M})$ be a corresponding eigenvector. In particular, we let $\lambda_{\max}(\mathbf{M}) := \lambda_1(\mathbf{M})$. We define $\mathbb{S}^{d-1} := \{\mathbf{v} \in \mathbb{R}^d\colon \|\mathbf{v}\|_2 = 1\}$ to be the d-dimensional unit sphere. We define the matrix element-wise supremum norm ($\ell_{\max}$ norm), spectral norm (ℓ_2 norm), and restricted spectral norm ($\ell_{2,s}$ norm) as

$$\|\mathbf{M}\|_{\max} := \max\big\{|\mathbf{M}_{ij}|\big\}, \qquad \|\mathbf{M}\|_2 := \sup_{\mathbf{v}\in\mathbb{S}^{d-1}} \|\mathbf{M}\mathbf{v}\|_2 \quad \text{and} \quad \|\mathbf{M}\|_{2,s} := \sup_{\mathbf{v}\in\mathbb{S}^{d-1}\cap\|\mathbf{v}\|_0\le s} \|\mathbf{M}\mathbf{v}\|_2.$$

We define $\mathrm{diag}(\mathbf{M})$ to be a diagonal matrix with $[\mathrm{diag}(\mathbf{M})]_{jj} = \mathbf{M}_{jj}$ for $j = 1, \dots, d$. We also denote $\mathrm{vec}(\mathbf{M}) := (\mathbf{M}_{*1}^T, \dots, \mathbf{M}_{*d}^T)^T$. For any two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$, we denote $\langle \mathbf{a}, \mathbf{b}\rangle := \mathbf{a}^T\mathbf{b}$ and $\mathrm{sign}(\mathbf{a}) := (\mathrm{sign}(a_1), \dots, \mathrm{sign}(a_d))^T$, where $\mathrm{sign}(x) = x/|x|$ with the convention $0/0 = 0$.

1.3. Paper organization

The rest of this paper is organized as follows. In the next section, we briefly overview the transelliptical distribution family and the main concentration results for the transformed Kendall's tau sample correlation matrix proposed by Han and Liu [17]. In Section 3, we analyze the convergence rates of Kendall's tau sample correlation matrix and its transformed version with regard to the spectral norm. In Section 4, we analyze the convergence rates of Kendall's tau sample correlation matrix and its transformed version with regard to the restricted spectral norm. The technical proofs of these results are provided in Section 5. More discussions and conclusions are provided in Section 6.

2. Preliminaries and background overview

In this section, we briefly review the transelliptical distribution and the corresponding latent generalized correlation matrix estimator proposed by Han and Liu [17].

2.1. Transelliptical distribution family

The concept of transelliptical distribution builds upon the elliptical distribution. Accordingly, we first provide a definition of the elliptical distribution, using the stochastic representation as in Fang et al. [14]. In the sequel, for any two random vectors $\mathbf{X}$ and $\mathbf{Y}$, we denote $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$ if they are identically distributed.

Definition 2.1 (Fang et al. [14]). *A random vector $\mathbf{Z} = (Z_1, \ldots, Z_d)^T$ follows an elliptical distribution if and only if $\mathbf{Z}$ has a stochastic representation: $\mathbf{Z} \stackrel{d}{=} \boldsymbol{\mu} + \xi \mathbf{A}\mathbf{U}$. Here $\boldsymbol{\mu} \in \mathbb{R}^d$, $q := \operatorname{rank}(\mathbf{A})$, $\mathbf{A} \in \mathbb{R}^{d\times q}$, $\xi \geq 0$ is a random variable independent of $\mathbf{U}$, $\mathbf{U} \in \mathbb{S}^{q-1}$ is uniformly distributed on the unit sphere in $\mathbb{R}^q$. In this setting, letting $\boldsymbol{\Sigma} := \mathbf{A}\mathbf{A}^T$, we denote $\mathbf{Z} \sim EC_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi)$. Here, $\boldsymbol{\Sigma}$ is called the scatter matrix.*

The elliptical family can be viewed as a semiparametric generalization of the Gaussian family, maintaining the symmetric property of the Gaussian distribution but allowing heavy tails and richer structures. Moreover, it is a natural model for many multivariate methods such as principal component analysis (Boente et al. [6]). The transelliptical distribution family further relaxes the symmetric assumption of the elliptical distribution by assuming that, after unspecified strictly increasing marginal transformations, the data are elliptically distributed. A formal definition of the transelliptical distribution is as follows.

Definition 2.2 (Han and Liu [17]). *A random vector $\mathbf{X} = (X_1, \ldots, X_d)^T$ follows a transelliptical distribution, denoted by $\mathbf{X} \sim TE_d(\boldsymbol{\Sigma}, \xi; f_1, \ldots, f_d)$, if there exist univariate strictly increasing functions $f_1, \ldots, f_d$ such that*

$$\bigl(f_1(X_1), \ldots, f_d(X_d)\bigr)^T \sim EC_d(\mathbf{0}, \boldsymbol{\Sigma}, \xi) \qquad \textit{where } \operatorname{diag}(\boldsymbol{\Sigma}) = \mathbf{I}_d \textit{ and } \mathbb{P}(\xi = 0) = 0.$$

Here $\mathbf{I}_d \in \mathbb{R}^{d\times d}$ is the d-dimensional identity matrix and $\boldsymbol{\Sigma}$ is called the latent generalized correlation matrix.

We note that the transelliptical distribution is closely related to the nonparanormal distribution (Liu et al. [27,29], Xue and Zou [36], Han and Liu [16], Han et al. [18]) and meta-elliptical distribution (Fang et al. [13]). The nonparanormal distribution assumes that after unspecified strictly increasing marginal transformations the data are Gaussian distributed. It is easy to see that the transelliptical family contains the nonparanormal family. On the other hand, it is subtle to elaborate the difference between the transelliptical and meta-elliptical. In short, the transelliptical family contains meta-elliptical family. Compared to the meta-elliptical, the transelliptical family does not require the random vectors to have densities and brings new insight into both theoretical analysis and model interpretability. We refer to Liu et al. [28] for more detailed discussion on the comparison between the transelliptical family, nonparanormal and meta-elliptical families.

2.2. Latent generalized correlation matrix estimation

Following Han and Liu [17], we are interested in estimating the latent generalized correlation matrix $\mathbf{\Sigma}$, i.e., the correlation matrix of the latent elliptically distributed random vector $f(\mathbf{X}) := (f_1(X_1), \ldots, f_d(X_d))^T$. By treating both the generating variable ξ and the marginal transformation functions $f = \{f_j\}_{j=1}^d$ as nuisance parameters, Han and Liu [17] proposed to use a transformed Kendall's tau sample correlation matrix to estimate the latent generalized correlation matrix $\mathbf{\Sigma}$. More specifically, letting $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be n independent and identically distributed observations of a random vector $\mathbf{X} \in TE_d(\mathbf{\Sigma}, \xi; f_1, \ldots, f_d)$, the Kendall's tau correlation coefficient between the variables X_j and X_k is defined as

$$\widehat{\tau}_{jk} := \frac{2}{n(n-1)} \sum_{i<i'} \operatorname{sign}\bigl((\mathbf{x}_i - \mathbf{x}_{i'})_j (\mathbf{x}_i - \mathbf{x}_{i'})_k\bigr).$$

Its population quantity can be written as

$$\tau_{jk} := \mathbb{P}\bigl((X_j - \widetilde{X}_j)(X_k - \widetilde{X}_k) > 0\bigr) - \mathbb{P}\bigl((X_j - \widetilde{X}_j)(X_k - \widetilde{X}_k) < 0\bigr), \tag{2.1}$$

where $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \ldots, \widetilde{X}_d)^T$ is an independent copy of $\mathbf{X}$. We denote

$$\mathbf{T} := [\tau_{jk}] \quad \text{and} \quad \widehat{\mathbf{T}} := [\widehat{\tau}_{jk}]$$

to be the Kendall's tau correlation matrix and Kendall's tau sample correlation matrix.

For the transelliptical family, it is known that $\mathbf{\Sigma}_{jk} = \sin(\frac{\pi}{2}\tau_{jk})$ (check, e.g., Theorem 3.2 in Han and Liu [17]). A latent generalized correlation matrix estimator $\widehat{\mathbf{\Sigma}} := [\widehat{\mathbf{\Sigma}}_{jk}]$, called the transformed Kendall's tau sample correlation matrix, is accordingly defined by

$$\widehat{\mathbf{\Sigma}}_{jk} = \sin\left(\frac{\pi}{2}\widehat{\tau}_{jk}\right). \tag{2.2}$$

Han and Liu [17] showed that, without any moment constraint,

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max} = O_P\left(\sqrt{\frac{\log d}{n}}\right),$$

and accordingly by simple algebra we have

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 = O_P\left(d\sqrt{\frac{\log d}{n}}\right) \quad \text{and} \quad \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{2,s} = O_P\left(s\sqrt{\frac{\log d}{n}}\right). \tag{2.3}$$

The rates of convergence in (2.3) are far from optimal (check Lounici [30], Bunea and Xiao [7], and Vu and Lei [34] for the parametric rates). In the next two sections, we will push the results in Han and Liu [17] forward, showing that better rates of convergence can be built in estimating the Kendall's tau correlation matrix and the latent generalized correlation matrix.

3. Rate of convergence under spectral norm

In this section, we provide the rate of convergence of the Kendall's tau sample correlation matrix $\widehat{\mathbf{T}}$ to $\mathbf{T}$, as well as the transformed Kendall's tau sample correlation matrix $\widehat{\mathbf{\Sigma}}$ to $\mathbf{\Sigma}$, under the spectral norm. The next theorem shows that, without any moment constraint or assumption on the data distribution (as long as it is continuous), the rate of convergence of $\widehat{\mathbf{T}}$ to $\mathbf{T}$ under the spectral norm is $\|\mathbf{T}\|_2\sqrt{r_e(\mathbf{T})\log d/n}$, where for any positive semidefinite matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$,

$$r_e(\mathbf{M}) := \frac{\mathrm{Tr}(\mathbf{M})}{\|\mathbf{M}\|_2}$$

is called the effective rank of $\mathbf{M}$ and must be less than or equal to the dimension d. For notational simplicity, in the sequel we assume that the sample size n is even. When n is odd, we can always use $n-1$ data points without affecting the obtained rate of convergence.

Theorem 3.1. *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be n observations of a d-dimensional continuous random vector $\mathbf{X}$. Then when $r_e(\mathbf{T})\log d/n \to 0$, for sufficiently large n and any $0 < \alpha < 1$, with probability larger than $1-2\alpha$, we have*

$$\|\widehat{\mathbf{T}} - \mathbf{T}\|_2 \le 4\|\mathbf{T}\|_2 \sqrt{\frac{\{r_e(\mathbf{T})+1\}\log(d/\alpha)}{3n}}. \tag{3.1}$$

Theorem 3.1 shows that, when $r_e(\mathbf{T})\log d/n \to 0$, we have

$$\|\widehat{\mathbf{T}} - \mathbf{T}\|_2 = \mathrm{O}_P\left(\|\mathbf{T}\|_2\sqrt{\frac{r_e(\mathbf{T})\log d}{n}}\right).$$

This rate of convergence we proved is the same parametric rate as obtained in Vershynin [33], Lounici [30], and Bunea and Xiao [7] when there is not any additional structure.

In the next theorem, we show that, under the modeling assumption that $\mathbf{X}$ is transelliptically distributed, which is of particular interest in real applications as shown in Han and Liu [17], we have that a transformed version of the Kendall's tau sample correlation matrix can estimate the latent generalized correlation matrix in a nearly optimal rate.

Theorem 3.2. *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be n observations of $\mathbf{X} \sim TE_d(\mathbf{\Sigma}, \xi; f_1, \ldots, f_d)$. Let $\widehat{\mathbf{\Sigma}}$ be the transformed Kendall's tau sample correlation matrix defined in (2.2). We have, when $r_e(\mathbf{\Sigma})\log d/n \to 0$, for n large enough and $0 < \alpha < 1$, with probability larger than $1-2\alpha-\alpha^2$,*

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 \le \pi^2\|\mathbf{\Sigma}\|_2\left(2\sqrt{\frac{\{r_e(\mathbf{\Sigma})+1\}\log(d/\alpha)}{3n}} + \frac{r_e(\mathbf{\Sigma})\log(d/\alpha)}{n}\right). \tag{3.2}$$

Theorem 3.2 indicates that, when $r_e(\mathbf{\Sigma})\log d/n \to 0$, we have

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 = \mathrm{O}_P\left(\|\mathbf{\Sigma}\|_2\sqrt{\frac{r_e(\mathbf{\Sigma})\log d}{n}}\right).$$

By the discussion of Theorem 2 in Lounici [30], the obtained rate of convergence is minimax optimal up to a logarithmic factor with respect to a suitable parameter space. However, compared to the conditions in Lounici [30], and Bunea and Xiao [7], which require strong multivariate sub-Gaussian modeling assumption on $\mathbf{X}$ (which implies the existence of moments of arbitrary order), $\widehat{\boldsymbol{\Sigma}}$ attains this parametric rate in estimating the latent generalized correlation matrix without any moment constraints.

Remark 3.3. The $\log d$ term presented in the rate of convergence of $\widehat{\mathbf{T}}$ and $\widehat{\boldsymbol{\Sigma}}$ is an artifact of the proof, and also appears in the statistical analysis of the sample covariance matrix under the sub-Gaussian model (see, e.g., Proposition 3 in Lounici [30] and Theorem 2.2 in Bunea and Xiao [7]). If we would like to highlight the role of the effective rank, $r_e(\mathbf{T})$ and $r_e(\boldsymbol{\Sigma})$, to our knowledge there is no work that can avoid the $\log d$ term. On the other hand, in estimating $\mathbf{T}$ using $\widehat{\mathbf{T}}$, a $\mathrm{O}_P(\sqrt{d/n})$ rate of convergence can be attained under the condition of Theorem 4.11 provided in the next section. In estimating $\boldsymbol{\Sigma}$ using $\widehat{\boldsymbol{\Sigma}}$, a $\mathrm{O}_P(\sqrt{d/n})$ rate of convergence is also attainable under the condition of Theorem 4.11 when $d(\log d)^2 = \mathrm{O}(n)$.

4. Rate of convergence under restricted spectral norm

In this section, we analyze the rates of convergence of the Kendall's tau sample correlation matrix and its transformed version under the restricted spectral norm. The main target is to improve the rate $\mathrm{O}_P(s\sqrt{\log d/n})$ shown in (2.3) to the rate $\mathrm{O}_P(\sqrt{s\log(d/s)/n})$. Such a rate has been shown to be minimax optimal under the Gaussian model (via combining Theorem 2.1 and Lemma 3.2.1 in Vu and Lei [34]). Obtaining such an improved rate is technically challenging since the data could be very heavy-tailed and the transformed Kendall's tau sample correlation matrix has a much more complex structure than the Pearson's covariance/correlation matrix.

In the following, we lay out a venue to analyze the statistical efficiency of $\widehat{\mathbf{T}}$ and $\widehat{\boldsymbol{\Sigma}}$ under the restricted spectral norm. In particular, we characterize a subset of the transelliptical distributions for which $\widehat{\mathbf{T}}$ and $\widehat{\boldsymbol{\Sigma}}$ can approximate $\mathbf{T}$ and $\boldsymbol{\Sigma}$ in an improved rate. More specifically, we provide a "sign sub-Gaussian" condition which is sufficient for $\widehat{\mathbf{T}}$ and $\widehat{\boldsymbol{\Sigma}}$ to attain the nearly optimal rate. This condition is related to the sub-Gaussian assumption in Vu and Lei [34], Lounici [30], and Bunea and Xiao [7] (see Assumption 2.2 in Vu and Lei [34], e.g.). Before proceeding to the formal definition of this condition, we first define an operator $\psi : \mathbb{R} \to \mathbb{R}$ as follows.

Definition 4.1. *For any random variable $Y \in \mathbb{R}$, the operator $\psi : \mathbb{R} \to \mathbb{R}$ is defined as*

$$\psi(Y; \alpha, t_0) := \inf\bigl\{c > 0 \colon \mathbb{E}\exp\bigl\{t\bigl(Y^\alpha - \mathbb{E}Y^\alpha\bigr)\bigr\} \le \exp\bigl(ct^2\bigr), \textit{for } |t| < t_0\bigr\}. \tag{4.1}$$

The operator $\psi(\cdot)$ can be used to quantify the tail behaviors of random variables. We recall that a zero-mean random variable $X \in \mathbb{R}$ is said to be sub-Gaussian if there exists a constant c such that $\mathbb{E}\exp(tX) \le \exp(ct^2)$ for all $t \in \mathbb{R}$. A zero-mean random variable $Y \in \mathbb{R}$ with $\psi(Y; 1, \infty)$ bounded is well known to be sub-Gaussian, which implies a tail probability

$$\mathbb{P}\bigl(|Y - \mathbb{E}Y| > t\bigr) < 2\exp\bigl(-t^2/(4c)\bigr),$$

where c is the constant defined in equation (4.1). Moreover, $\psi(Y;\alpha,t_0)$ is related to the Orlicz ψ_2-norm. A formal definition of the Orlicz norm is provided as follows.

Definition 4.2. *For any random variable $Y \in \mathbb{R}$, its Orlicz ψ_2-norm is defined as*

$$\|Y\|_{\psi_2} := \inf\bigl\{c > 0\colon \mathbb{E}\exp\bigl(|Y/c|^2\bigr) \le 2\bigr\}.$$

It is well known that a random variable Y has $\psi(Y;1,\infty)$ to be bounded if and only if $\|Y\|_{\psi_2}$ in Definition 4.2 is bounded (van de Geer and Lederer [32]). We refer to Lemma A.1 in the Appendix for a more detailed description on this property.

Another relevant norm to $\psi(\cdot)$ is the sub-Gaussian norm $\|\cdot\|_{\phi_2}$ used in, for example, Vershynin [33]. A former definition of the sub-Gaussian norm is as follows.

Definition 4.3. *For any random variable $X \in \mathbb{R}$, its sub-Gaussian norm is defined as*

$$\|X\|_{\phi_2} := \sup_{k\ge 1} k^{-1/2}\bigl(\mathbb{E}|X|^k\bigr)^{1/k}.$$

The sub-Gaussian norm is also highly related to the sub-Gaussian random variables. In particular, we have if $\mathbb{E}X = 0$, then $\mathbb{E}\exp(tX) \le \exp(Ct^2\|X\|_{\phi_2}^2)$.

Using the operator $\psi(\cdot)$, we now proceed to define the sign sub-Gaussian condition. For mathematical rigorousness, the formal definition is posed on $\{\mathcal{F}^d, d = 1, 2, \ldots\}$, where $\mathcal{F}^d$ represents a set of probability measures on $\mathbb{R}^d$. Here for any vector $\mathbf{v} = (v_1, \ldots, v_d) \in \mathbb{R}^d$, we remind that $\operatorname{sign}(\mathbf{v}) := (\operatorname{sign}(v_1), \ldots, \operatorname{sign}(v_d))^T$. In the following, a random vector $\mathbf{X}$ is said to be in a set of probability measures $\mathcal{F}'$ if its distribution is in $\mathcal{F}'$.

Definition 4.4 (Sign sub-Gaussian condition). *For $d = 1, 2, \ldots,$ let $\mathcal{F}^d$ be a set of probability measures on $\mathbb{R}^d$ such that infinitely many sets $\mathcal{F}^d$ are nonempty and $\mathcal{F} := \bigcup_{d=1}^{\infty}\mathcal{F}^d$. $\mathcal{F}$ is said to satisfy the sign sub-Gaussian condition if and only if for any $\mathbf{X}$ in $\mathcal{F}$, we have*

$$\sup_{\mathbf{v}\in\mathbb{S}^{d-1}} \psi\bigl(\bigl\langle \operatorname{sign}(\mathbf{X}-\widetilde{\mathbf{X}}), \mathbf{v}\bigr\rangle; 2, t_0\bigr) \le K\|\mathbf{T}\|_2^2, \tag{4.2}$$

where $\widetilde{\mathbf{X}}$ is an independent copy of $\mathbf{X}$, K is an absolute constant, and t_0 is another absolute positive number such that $t_0\|\mathbf{T}\|_2$ is lower bounded by an absolute positive constant. We remind that here $\mathbf{T}$ can be written as

$$\mathbf{T} := \mathbb{E}\operatorname{sign}(\mathbf{X}-\widetilde{\mathbf{X}}) \cdot \bigl(\operatorname{sign}(\mathbf{X}-\widetilde{\mathbf{X}})\bigr)^T.$$

To gain more insights about the sign sub-Gaussian condition, we point out two sets of probability measures of interest that satisfy the sign sub-Gaussian condition.

Proposition 4.5. *Suppose the set of probability measures $\mathcal{F}$ satisfies that for any random vector $\mathbf{X}$ in $\mathcal{F}$ and $\widetilde{\mathbf{X}}$ being an independent copy of $\mathbf{X}$, we have*

$$\sup_{\mathbf{v}\in\mathbb{S}^{d-1}} \bigl\|\bigl\langle \operatorname{sign}(\mathbf{X}-\widetilde{\mathbf{X}}), \mathbf{v}\bigr\rangle^2 - \mathbf{v}^T\mathbf{T}\mathbf{v}\bigr\|_{\psi_2} \le L_1\|\mathbf{T}\|_2, \tag{4.3}$$

where L_1 is a fixed constant. Then $\mathcal{F}$ satisfies the sign sub-Gaussian condition by setting $t_0 = \infty$ and $K = 5L_1^2/2$ in equation (4.2).

Proposition 4.6. *Suppose the set of probability measure $\mathcal{F}$ satisfies that for any random vector $\mathbf{X}$ in $\mathcal{F}$ and $\widetilde{\mathbf{X}}$ being an independent copy of $\mathbf{X}$, we have there exists an absolute constant L_2 such that*

$$\left\| \mathbf{v}^T \operatorname{sign}(\mathbf{X} - \widetilde{\mathbf{X}}) \right\|_{\phi_2}^2 \le \frac{L_2 \|\mathbf{T}\|_2}{2} \qquad \text{for all } \mathbf{v} \in \mathbb{S}^{d-1}. \tag{4.4}$$

Then $\mathcal{F}$ satisfies the sign sub-Gaussian condition with $t_0 = c\|\mathbf{T}\|_2^{-1}$ and $K = C$ in equation (4.2), *where c and C are two fixed absolute constants.*

In the following, for clarity of presentation, we abuse notation a little and write that $\mathbf{X}$ satisfies the sign sub-Gaussian condition if there exists a set of probability measures $\mathcal{F}$ satisfying the sign sub-Gaussian condition such that for $d = 1, 2, \dots, \mathbf{X} \in \mathbb{R}^d$ is in $\mathcal{F}$.

Proposition 4.6 builds a bridge between the sign sub-Gaussian condition and Assumption 1 in Bunea and Xiao [7] and Lounici [30]. More specifically, saying that $\mathbf{X}$ satisfies equation (4.4) is equivalent to saying that $\operatorname{sign}(\mathbf{X} - \widetilde{\mathbf{X}})$ satisfies the multivariate sub-Gaussian condition defined in Bunea and Xiao [7]. Therefore, Proposition 4.6 can be treated as an explanation of why we call the condition in equation (4.2) "sign sub-Gaussian." However, by Lemma 5.14 in Vershynin [33], the sign sub-Gaussian condition is weaker than that of equation (4.4), that is, a set of probability measures satisfying the sign sub-Gaussian condition does not necessarily satisfy the condition in Proposition 4.6.

The sign sub-Gaussian condition is intuitive due to its relation to the Orlicz and sub-Gaussian norms. However, it is extremely difficult to verify whether a given set of distributions satisfies this condition. The main difficulty lies in the fact that we must sharply characterize the tail behavior of the summation of a sequence of possibly correlated discrete Bernoulli random variables, which is much harder than analyzing the summation of Gaussian random variables as usually done in the literature.

In the following, we provide several examples of sets of distributions that satisfy the sign sub-Gaussian condition. The next theorem shows that the transelliptically distributed random vector $\mathbf{X} \sim TE_d(\mathbf{\Sigma}, \xi; f_1, \dots, f_d)$ such that $\mathbf{\Sigma} = \mathbf{I}_d$ (i.e., the underlying is a spherical distribution) for $d = 1, 2, \dots$ satisfies the sign sub-Gaussian condition. The proof of Theorem 4.7 is in Section 5.4.

Theorem 4.7. *Suppose that, for $d = 1, 2, \dots, \mathbf{X} \sim TE_d(\mathbf{I}_d, \xi; f_1, \dots, f_d)$ is transelliptically distributed with a latent spherical distribution. Then $\mathbf{X}$ satisfies the sign sub-Gaussian condition.*

In the next theorem, we provide a stronger version of Theorem 4.7. We call a square matrix compound symmetric if the off-diagonal values of the matrix are equal. The next theorem shows that the transelliptically distributed $\mathbf{X} \sim TE_d(\mathbf{\Sigma}, \xi; f_1, \dots, f_d)$, with $\mathbf{\Sigma}$ a compound symmetric matrix, satisfies equation (4.4) and, therefore, satisfies the sign sub-Gaussian condition.

Theorem 4.8. *Suppose that for $d = 1, 2, \dots, \mathbf{X} \sim TE_d(\mathbf{\Sigma}, \xi; f_1, \dots, f_d)$ is transelliptically distributed such that $\mathbf{\Sigma}$ is a compound symmetric matrix (i.e., $\mathbf{\Sigma}_{jk} = \rho$ for all $j \neq k$). Then if*

$0 \le \rho := \mathbf{\Sigma}_{12} \le C_0 < 1$ *for some absolute positive constant* C_0, *we have that* $\mathbf{X}$ *satisfies the sign sub-Gaussian condition.*

Although Theorem 4.7 can be directly proved using the result in Theorem 4.8, the proof of Theorem 4.7 contains utterly different techniques which are more transparent and illustrate the main challenges of analyzing binary sequences even in the uncorrelated setting. Therefore, we still list this theorem separately and provide a separate proof in Section 5.4. Theorem 4.8 leads to the following corollary, which characterizes a subfamily of the transelliptical distributions satisfying the sign sub-Gaussian condition.

Corollary 4.9. *Suppose that for* $d = 1, 2, \dots, \mathbf{X} \sim TE_d(\mathbf{\Sigma}, \xi; f_1, \dots, f_d)$ *is transelliptically distributed with* $\mathbf{\Sigma}$ *a block diagonal compound symmetric matrix, that is,*

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & \dots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{\Sigma}_q \end{pmatrix}, \tag{4.5}$$

where $\mathbf{\Sigma}_k \in \mathbb{R}^{d_k \times d_k}$ *for* $k = 1, \dots, q$ *is compound symmetric matrix with* $\rho_k := [\mathbf{\Sigma}_k]_{12} \ge 0$. *We have, if* q *is upper bounded by an absolute positive constant and* $0 \le \rho_k \le C_1 < 1$ *for some absolute positive constant* C_1, $\mathbf{X}$ *satisfies the sign sub-Gaussian condition.*

We call the matrix in the form of equation (4.5) block diagonal compound symmetric matrix. Corollary 4.9 implies that transelliptically distributed random vectors with a latent block diagonal compound symmetric latent generalized correlation matrix satisfy the sign sub-Gaussian condition.

Remark 4.10. The sub-Gaussian condition is an artifact of the proof. Right now, we are not aware of any transelliptical distribution that does not satisfy this condition. More investigation on the necessity of this condition is challenging due to the discontinuity issue of the sign transformation and will be left for future investigation.

Using the sign sub-Gaussian condition, we have the following main result, which shows that as long as the sign sub-Gaussian condition holds, improved rates of convergence for both $\widehat{\mathbf{T}}$ and $\widehat{\mathbf{\Sigma}}$ under the restricted spectral norm can be attained.

Theorem 4.11. *For* $d = 1, 2, \dots$, *let* $\mathbf{x}_1, \dots, \mathbf{x}_n$ *be* n *observations of* $\mathbf{X} \in \mathbb{R}^d$, *for which the sign sub-Gaussian condition holds. We have, when* $s \log(d/s)/n \to 0$, *with probability larger than* $1 - 2\alpha$,

$$\|\widehat{\mathbf{T}} - \mathbf{T}\|_{2,s} \le 4(2K)^{1/2}\|\mathbf{T}\|_2 \sqrt{\frac{s(3 + \log(d/s)) + \log(1/\alpha)}{n}}. \tag{4.6}$$

Moreover, when we further have $\mathbf{X} \sim TE_d(\boldsymbol{\Sigma}, \xi; f_1, \ldots, f_d)$, *with probability larger* $1 - 2\alpha - \alpha^2$,

$$\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{2,s} \leq \pi^2 \biggl(2(2K)^{1/2} \|\boldsymbol{\Sigma}\|_2 \sqrt{\frac{s(3 + \log(d/s)) + \log(1/\alpha)}{n}} + \frac{s \log(d/\alpha)}{n} \biggr). \tag{4.7}$$

The results presented in Theorem 4.11 show that under various settings the rate of convergence for $\widehat{\boldsymbol{\Sigma}}$ under the restricted spectral norm is $\mathrm{O}_P(\sqrt{s\log(d/s)/n})$, which is the parametric and minimax optimal rate shown in Vu and Lei [34] within the Gaussian family. However, the Kendall's tau sample correlation matrix and its transformed version attains this rate with all the moment constraints waived.

5. Technical proofs

We provide the technical proofs of the theorems shown in Sections 3 and 4.

5.1. Proof of Theorem 3.1

Proof. Reminding that $\mathbf{x}_i := (x_{i1}, \ldots, x_{id})^T$, for $i \neq i'$, let

$$\mathbf{S}_{i,i'} := \bigl(\mathrm{sign}(x_{i,1} - x_{i',1}), \ldots, \mathrm{sign}(x_{i,d} - x_{i',d})\bigr)^T.$$

We denote $\widehat{\boldsymbol{\Delta}}_{i,i'}$ to be $n(n-1)$ random matrices with

$$\widehat{\boldsymbol{\Delta}}_{i,i'} := \frac{1}{n(n-1)} \bigl(\mathbf{S}_{i,i'}\mathbf{S}_{i,i'}^T - \mathbf{T}\bigr).$$

By simple calculation, we have $\widehat{\mathbf{T}} - \mathbf{T} = \sum_{i,i'} \widehat{\boldsymbol{\Delta}}_{i,i'}$ and $\widehat{\mathbf{T}} - \mathbf{T}$ is a U-statistic.

In the following we extend the standard decoupling trick from Hoeffding [19] from the U-statistic of random variables to the matrix setting. The extension relies on the matrix version of the Laplace transform method. For any square matrix $\mathbf{M} \in \mathbb{R}^d$, we define

$$\exp(\mathbf{M}) := \mathbf{I}_d + \sum_{k=1}^{\infty} \frac{\mathbf{M}^k}{k!},$$

where $k!$ represents the factorial product of k. Using Proposition 3.1 in Tropp [31], we have

$$\mathbb{P}\bigl[\lambda_{\max}(\widehat{\mathbf{T}} - \mathbf{T}) \geq t\bigr] \leq \inf_{\theta > 0} \mathrm{e}^{-\theta t} \mathbb{E}\bigl[\mathrm{Tr}\, \mathrm{e}^{\theta(\widehat{\mathbf{T}} - \mathbf{T})}\bigr], \tag{5.1}$$

and we bound $\mathbb{E}[\mathrm{Tr}\, \mathrm{e}^{\theta(\widehat{\mathbf{T}} - \mathbf{T})}]$ as follows.

The trace exponential function

$$\mathrm{Tr} \exp : \mathbf{A} \to \mathrm{Tr}\, \mathrm{e}^{\mathbf{A}}$$

is a convex mapping from the space of self-adjoint matrix to $\mathbb{R}^+$ (see Section 2.4 of Tropp [31] and reference therein). Let $m = n/2$. For any permutation σ of $1, \dots, n$, let $(i_1, \dots, i_n) := \sigma(1, \dots, n)$. For $r = 1, \dots, m$, we define $\mathbf{S}_r^\sigma$ and $\widehat{\boldsymbol{\Delta}}_r^\sigma$ to be

$$\mathbf{S}_r^\sigma := \mathbf{S}_{i_{2r}, i_{2r-1}} \quad \text{and} \quad \widehat{\boldsymbol{\Delta}}_r^\sigma := \frac{1}{m}\big(\mathbf{S}_r^\sigma \big[\mathbf{S}_r^\sigma\big]^T - \mathbf{T}\big).$$

Moreover, for $i = 1, \dots, m$, let

$$\mathbf{S}_i := \mathbf{S}_{2i, 2i-1} \quad \text{and} \quad \widehat{\boldsymbol{\Delta}}_i := \frac{1}{m}\big(\mathbf{S}_i \mathbf{S}_i^T - \mathbf{T}\big).$$

The convexity of the trace exponential function implies that

$$\begin{aligned}
\operatorname{Tr} \mathrm{e}^{\theta(\widehat{\mathbf{T}} - \mathbf{T})} &= \operatorname{Tr} \mathrm{e}^{\theta \sum_{i,i'} \widehat{\boldsymbol{\Delta}}_{i,i'}} \\
&= \operatorname{Tr} \exp\left\{ \frac{1}{\operatorname{card}(S_n)} \sum_{\sigma \in S_n} \theta \sum_{r=1}^m \widehat{\boldsymbol{\Delta}}_r^\sigma \right\} \\
&\le \frac{1}{\operatorname{card}(S_n)} \sum_{\sigma \in S_n} \operatorname{Tr} \mathrm{e}^{\theta \sum_{r=1}^m \widehat{\boldsymbol{\Delta}}_r^\sigma},
\end{aligned} \tag{5.2}$$

where S_n is the permutation group of $\{1, \dots, n\}$. Taking expectation on both sides of equation (5.2) gives that

$$\mathbb{E} \operatorname{Tr} \mathrm{e}^{\theta(\widehat{\mathbf{T}} - \mathbf{T})} \le \mathbb{E} \operatorname{Tr} \mathrm{e}^{\theta \sum_{i=1}^m \widehat{\boldsymbol{\Delta}}_i}. \tag{5.3}$$

According to the definition, $\widehat{\boldsymbol{\Delta}}_1, \dots, \widehat{\boldsymbol{\Delta}}_m$ are m independent and identically distributed random matrices, and this finishes the decoupling step.

Combing equations (5.1) and (5.3), we have

$$\mathbb{P}\big[\lambda_{\max}(\widehat{\mathbf{T}} - \mathbf{T}) \ge t\big] \le \inf_{\theta > 0} \mathrm{e}^{-\theta t} \mathbb{E} \operatorname{Tr} \mathrm{e}^{\theta \sum_{i=1}^m \widehat{\boldsymbol{\Delta}}_i}. \tag{5.4}$$

Recall that $\mathbb{E}\widehat{\boldsymbol{\Delta}}_i = 0$. Following the proof of Theorem 6.1 in Tropp [31], if we can show that there are some nonnegative numbers R_1 and R_2 such that

$$\lambda_{\max}(\widehat{\boldsymbol{\Delta}}_i) \le R_1, \qquad \left\| \sum_{i=1}^m \mathbb{E} \widehat{\boldsymbol{\Delta}}_i^2 \right\|_2 \le R_2,$$

then the right-hand side of equation (5.4) can be bounded by

$$\inf_{\theta > 0} \mathrm{e}^{-\theta t} \mathbb{E} \operatorname{Tr} \mathrm{e}^{\theta \sum_{i=1}^m \widehat{\boldsymbol{\Delta}}_i} \le d \exp\left\{ -\frac{t^2/2}{R_2 + R_1 t/3} \right\}.$$

We first show that $R_1 = \frac{2d}{m}$. Because $\|\widehat{\boldsymbol{\Delta}}_i\|_{\max} \le 2/m$, by simple calculation, we have

$$\lambda_{\max}(\widehat{\boldsymbol{\Delta}}_i) \le \|\widehat{\boldsymbol{\Delta}}_i\|_1 \le d \cdot \|\widehat{\boldsymbol{\Delta}}_i\|_{\max} \le \frac{2d}{m}.$$

We then calculate R_2. For this, we have, because $\mathbf{X}$ is continuous,

$$\sum_{i=1}^m \mathbb{E}\widehat{\boldsymbol{\Delta}}_i^2 = \frac{1}{m}\mathbb{E}\big(\mathbf{S}_1\mathbf{S}_1^T - \mathbf{T}\big)^2 = \frac{1}{m}\big(\mathbb{E}\big(d\mathbf{S}_1\mathbf{S}_1^T\big) - \mathbf{T}^2\big) = \frac{1}{m}\big(d\mathbf{T} - \mathbf{T}^2\big).$$

Accordingly,

$$\left\|\sum_{i=1}^m \mathbb{E}\widehat{\boldsymbol{\Delta}}_i^2\right\|_2 \le \frac{1}{m}\big(d\|\mathbf{T}\|_2 + \|\mathbf{T}\|_2^2\big),$$

so we set $R_2 = \frac{1}{m}(d\|\mathbf{T}\|_2 + \|\mathbf{T}\|_2^2)$.

Thus, using Theorem 6.1 in Tropp [31], for any

$$t \le R_2/R_1 = \frac{d\|\mathbf{T}\|_2 + \|\mathbf{T}\|_2^2}{2d},$$

we have

$$\mathbb{P}\big\{\lambda_{\max}(\widehat{\mathbf{T}} - \mathbf{T}) \ge t\big\} \le d \cdot \exp\bigg(-\frac{3nt^2}{16(d\|\mathbf{T}\|_2 + \|\mathbf{T}\|_2^2)}\bigg).$$

A similar argument holds for $\lambda_{\max}(-\widehat{\mathbf{T}} + \mathbf{T})$. Accordingly, we have

$$\mathbb{P}\big\{\|\widehat{\mathbf{T}} - \mathbf{T}\|_2 \ge t\big\} \le 2d \cdot \exp\bigg(-\frac{3nt^2}{16(d\|\mathbf{T}\|_2 + \|\mathbf{T}\|_2^2)}\bigg).$$

Finally, when

$$n \ge \frac{64d^2\log(d/\alpha)}{3(d\|\mathbf{T}\|_2 + \|\mathbf{T}\|_2^2)},$$

we have

$$\sqrt{\frac{16(d\|\mathbf{T}\|_2 + \|\mathbf{T}\|_2^2)\log(d/\alpha)}{3n}} \le \frac{d\|\mathbf{T}\|_2 + \|\mathbf{T}\|_2^2}{2d}.$$

This completes the proof. $\square$

5.2. Proof of Theorem 3.2

To prove Theorem 3.2, we first need the following lemma, which connects $\sqrt{1 - \boldsymbol{\Sigma}_{jk}^2}$ to a Gaussian distributed random vector $(X, Y)^T \in \mathbb{R}^2$ and plays a key role in bounding $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_2$ by $\|\widehat{\mathbf{T}} - \mathbf{T}\|_2$.

Lemma 5.1. *Provided that*

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\mathbf{0}, \begin{bmatrix} 1 & \sigma \\ \sigma & 1 \end{bmatrix}\right),$$

we have

$$\mathbb{E}|XY| = \mathbb{E}XY\mathbb{E}\,\mathrm{sign}(XY) + \frac{2}{\pi}\sqrt{1-\sigma^2}.$$

Proof. We recall that $\sigma := \sin(\frac{\pi}{2}\tau)$ with τ the Kendall's tau correlation coefficient of X, Y. Without loss of generality, assume that $\sigma > 0, \tau > 0$ (otherwise show for $-Y$ instead of Y). Define

$$\beta_+ = \mathbb{E}|XY|I(XY > 0), \qquad \beta_- = \mathbb{E}|XY|I(XY < 0),$$

where $I(\cdot)$ is the indicator function. We then have

$$\mathbb{E}|XY| = \beta_+ + \beta_-, \qquad \mathbb{E}XY = \sigma = \beta_+ - \beta_-. \tag{5.5}$$

To compute β_+, using the fact that

$$X \stackrel{d}{=} \sqrt{\frac{1+\sigma}{2}}Z_1 + \sqrt{\frac{1-\sigma}{2}}Z_2, \qquad Y \stackrel{d}{=} \sqrt{\frac{1+\sigma}{2}}Z_1 - \sqrt{\frac{1-\sigma}{2}}Z_2,$$

where $Z_1, Z_2 \sim N_1(0, 1)$ are independently and identically distributed.

Let $F_{X,Y}$ and F_{Z_1,Z_2} be the joint distribution functions of $(X, Y)^T$ and $(Z_1, Z_2)^T$. We have

$$\begin{aligned}
\beta_+ &= \int_{xy>0} |xy|\,\mathrm{d}F_{X,Y}(x, y) \\
&= \int_{xy>0} \frac{(x+y)^2 - (x-y)^2}{4}\,\mathrm{d}F_{X,Y}(x, y) \\
&= \int_{z_1^2 > ((1-\sigma)/(1+\sigma))z_2^2} \left(\frac{1+\sigma}{2}z_1^2 - \frac{1-\sigma}{2}z_2^2\right)\mathrm{d}F_{Z_1,Z_2}(z_1, z_2) \\
&= \int_0^{+\infty}\int_{-\alpha}^{\alpha} 2\left\{\frac{1+\sigma}{2}r^2\cos^2(\theta) - \frac{1-\sigma}{2}r^2\sin^2(\theta)\right\}\cdot\frac{1}{2\pi}\mathrm{e}^{-r^2/2}r\,\mathrm{d}\theta\,\mathrm{d}r,
\end{aligned}$$

where $\alpha := \arcsin(\sqrt{\frac{1+\sigma}{2}})$. By simple calculation, we have

$$\int_0^{\infty} r^3\mathrm{e}^{-r^2/2}\,\mathrm{d}r = \frac{1}{2}\int_0^{\infty} ue^{-u/2}\,\mathrm{d}u = 2.$$

Accordingly, we can proceed the proof and show that

$$\begin{aligned}
\beta_+ &= \int_0^{+\infty}\int_{-\alpha}^{\alpha}\big(\cos(2\theta) + \sigma\big)\cdot r^3\frac{1}{2\pi}\mathrm{e}^{-r^2/2}\,\mathrm{d}\theta\,\mathrm{d}r \\
&= \frac{1}{\pi}\big(\sin(2\alpha) + 2\alpha\sigma\big).
\end{aligned} \tag{5.6}$$

Since $\sin(2\alpha) = \sqrt{1-\sigma^2} = \cos(\pi\tau/2)$ and $\alpha \geq \arcsin(\sqrt{1/2}) \geq \pi/4$, we have that $2\alpha = \frac{\pi}{2}(1+\tau)$, and then equation (5.6) continues to give

$$\beta_+ = \frac{\sigma}{2}(1+\tau) + \frac{1}{\pi}\sqrt{1-\sigma^2}.$$

Combined with equation (5.5) gives the equality claimed. □

Using Theorem 3.1 and Lemma 5.1, we proceed to prove Theorem 3.2.

Proof of Theorem 3.2. Using Taylor expansion, for any $j \neq k$, we have

$$\sin\left(\frac{\pi}{2}\widehat{\tau}_{jk}\right) - \sin\left(\frac{\pi}{2}\tau_{jk}\right) = \cos\left(\frac{\pi}{2}\tau_{jk}\right)\frac{\pi}{2}(\widehat{\tau}_{jk} - \tau_{jk}) - \frac{1}{2}\sin(\theta_{jk})\left(\frac{\pi}{2}\right)^2(\widehat{\tau}_{jk} - \tau_{jk})^2,$$

where θ_{jk} lies between τ_{jk} and $\widehat{\tau}_{jk}$. Thus,

$$\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma} = \mathbf{E}_1 + \mathbf{E}_2,$$

where $\mathbf{E}_1, \mathbf{E}_2 \in \mathbb{R}^{d\times d}$ satisfy that for $j \neq k$,

$$[\mathbf{E}_1]_{jk} = \cos\left(\frac{\pi}{2}\tau_{jk}\right)\frac{\pi}{2}(\widehat{\tau}_{jk} - \tau_{jk}),$$

$$[\mathbf{E}_2]_{jk} = -\frac{1}{2}\sin(\theta_{jk})\left(\frac{\pi}{2}\right)^2(\widehat{\tau}_{jk} - \tau_{jk})^2,$$

and the diagonal entries of both E_1 and E_2 are all zero.

Using the results of U-statistics shown in Hoeffding [19], we have that for any $j \neq k$ and $t > 0$,

$$\mathbb{P}\big(|\widehat{\tau}_{jk} - \tau_{jk}| > t\big) < 2\mathrm{e}^{-nt^2/4}.$$

For some constant α, let the event Ω_2 be defined as

$$\Omega_2 := \left\{\exists 1 \leq j \neq k \leq d, \big|[\mathbf{E}_2]_{jk}\big| > \pi^2 \cdot \frac{\log(d/\alpha)}{n}\right\}.$$

Since $|[\mathbf{E}_2]_{jk}| \leq \frac{\pi^2}{8}(\widehat{\tau}_{jk} - \tau_{jk})^2$, by union bound, we have

$$\mathbb{P}(\Omega_2) \leq \frac{d^2}{2} \cdot 2\mathrm{e}^{-2\log(d/\alpha)} = \alpha^2.$$

Conditioning on Ω_2^C, for any $\mathbf{v} \in \mathbb{S}^{d-1}$, we have

$$\big|\mathbf{v}^T\mathbf{E}_2\mathbf{v}\big| \leq \sqrt{\sum_{j,k\in J}[\mathbf{E}_2]_{jk}^2} \cdot \|\mathbf{v}\|_2^2 \leq \sqrt{d^2\left(\pi^2 \cdot \frac{\log(d/\alpha)}{n}\right)^2} = \pi^2 \cdot \frac{d\log(d/\alpha)}{n}. \tag{5.7}$$

We then analyze the term $\mathbf{E}_1$. Let $\mathbf{W} = [\mathbf{W}_{jk}] \in \mathbb{R}^{d\times d}$ with $\mathbf{W}_{jk} = \frac{\pi}{2}\cos(\frac{\pi}{2}\tau_{jk})$ and $\widehat{\mathbf{T}} = [\widehat{\mathbf{T}}_{jk}]$ be the Kendall's tau sample correlation matrix with $\widehat{\mathbf{T}}_{jk} = \widehat{\tau}_{jk}$. We can write

$$\mathbf{E}_1 = \mathbf{W} \circ (\widehat{\mathbf{T}} - \mathbf{T}),$$

where $\circ$ represents the Hadamard product. Given the spectral norm bound of $\widehat{\mathbf{T}} - \mathbf{T}$ shown in Theorem 3.1, we now focus on controlling $\mathbf{E}_1$. Let $\mathbf{Y} := (Y_1, \ldots, Y_d)^T \sim N_d(\mathbf{0}, \mathbf{\Sigma})$ follow a Gaussian distribution with mean zero and covariance matrix $\mathbf{\Sigma}$. Using the equality in Lemma 5.1, we have, for any $j \neq k$,

$$\mathbb{E}|Y_j Y_k| = \tau_{jk}\mathbf{\Sigma}_{jk} + \frac{2}{\pi}\sqrt{1 - \mathbf{\Sigma}_{jk}^2}.$$

Reminding that

$$\cos\left(\frac{\pi}{2}\tau_{jk}\right) = \sqrt{1 - \sin^2\left(\frac{\pi}{2}\tau_{jk}\right)} = \sqrt{1 - \mathbf{\Sigma}_{jk}^2},$$

we have

$$\mathbf{W}_{jk} = \frac{\pi}{2}\cos\left(\frac{\pi}{2}\tau_{jk}\right) = \frac{\pi^2}{4}\big(\mathbb{E}|Y_j Y_k| - \tau_{jk}\mathbf{\Sigma}_{jk}\big).$$

Then let $\mathbf{Y}' := (Y_1', \ldots, Y_d')^T \in \mathbb{R}^d$ be an independent copy of $\mathbf{Y}$. We have, for any $\mathbf{v} \in \mathbb{S}^{d-1}$ and symmetric matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$,

$$\begin{aligned}
\big|\mathbf{v}^T\mathbf{M} \circ \mathbf{W}\mathbf{v}\big| &= \left|\sum_{j,k=1}^{d} v_j v_k \mathbf{M}_{jk}\mathbf{W}_{jk}\right| \\
&= \left|\mathbb{E}\frac{\pi^2}{4}\sum_{j,k} v_j v_k \mathbf{M}_{jk}\big(|Y_j Y_k| - Y_j Y_k \operatorname{sign}\big(Y_j' Y_k'\big)\big)\right| \\
&\le \frac{\pi^2}{4}\mathbb{E}\left(\left|\sum_{j,k} v_j v_k \mathbf{M}_{jk}|Y_j Y_k|\right| + \left|\sum_{j,k} v_j v_k \mathbf{M}_{jk} Y_j Y_k \operatorname{sign}\big(Y_j' Y_k'\big)\right|\right) \\
&\le \frac{\pi^2}{4}\|\mathbf{M}\|_2 \cdot \mathbb{E}\left(2\sum_j v_j^2 Y_j^2\right) \\
&= \frac{\pi^2}{4}\|\mathbf{M}\|_2 \cdot \left(2\sum_j v_j^2\right) \\
&= \frac{\pi^2}{2}\|\mathbf{M}\|_2.
\end{aligned} \tag{5.8}$$

Here, the second inequality is due to the fact that for any $\mathbf{M} \in \mathbb{R}^{d\times d}$ and $\mathbf{v} \in \mathbb{R}^d$, $|\mathbf{v}^T\mathbf{M}\mathbf{v}| \le \|\mathbf{M}\|_2\|\mathbf{v}\|_2$ and the third equality is due to the fact that $\mathbb{E}Y_j^2 = \mathbf{\Sigma}_{jj} = 1$ for any $j \in \{1, \ldots, d\}$.

Accordingly, we have

$$\|\mathbf{E}_1\|_2 = \left\|\mathbf{W} \circ (\widehat{\mathbf{T}} - \mathbf{T})\right\|_2 \le \frac{\pi^2}{2} \|\widehat{\mathbf{T}} - \mathbf{T}\|_2. \tag{5.9}$$

The bound in Theorem 3.2, with $\mathbf{\Sigma}$ being replaced by $\mathbf{T}$, follows from the fact that

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 = \|\mathbf{E}_1 + \mathbf{E}_2\|_2 \le \|\mathbf{E}_1\|_2 + \|\mathbf{E}_2\|_2$$

and by combining equations (3.1), (5.7) and (5.9). Finally, we prove that $\|\mathbf{T}\|_2 \le \|\mathbf{\Sigma}\|_2$. We have $\mathbf{T}_{jk} = \frac{2}{\pi}\arcsin(\mathbf{\Sigma}_{jk})$. Using the Taylor expansion and the fact that $|\mathbf{\Sigma}_{jk}| \le 1$ for any $(j,k) \in \{1,\dots,d\}$, we have

$$\mathbf{T} = \frac{2}{\pi} \sum_{m=0}^{\infty} \frac{(2m)!}{4^m (m!)^2 (2m+1)} \underbrace{(\mathbf{\Sigma} \circ \cdots \circ \mathbf{\Sigma})}_{2m+1}.$$

By Schur's theorem (see, e.g., page 95 in Johnson [22]), we have for any two positive semidefinite matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\|\mathbf{A} \circ \mathbf{B}\|_2 \le \left(\max_j \mathbf{A}_{jj}\right) \|\mathbf{B}\|_2.$$

Accordingly, using the fact that $\mathbf{\Sigma}_{jj} = 1$ for all $1 \le j \le d$, we have

$$\Big\|\underbrace{(\mathbf{\Sigma} \circ \cdots \circ \mathbf{\Sigma})}_{2m+1}\Big\|_2 \le \|\mathbf{\Sigma}\|_2,$$

implying that

$$\begin{aligned} \|\mathbf{T}\|_2 &\le \|\mathbf{\Sigma}\|_2 \cdot \frac{2}{\pi} \sum_{m=0}^{\infty} \frac{(2m)!}{4^m (m!)^2 (2m+1)} \\ &= \|\mathbf{\Sigma}\|_2 \cdot \frac{2}{\pi} \arcsin 1 = \|\mathbf{\Sigma}\|_2. \end{aligned} \tag{5.10}$$

Accordingly, we can replace $\mathbf{T}$ with $\mathbf{\Sigma}$ in the upper bound and have the desired result. □

5.3. Proofs of Propositions 4.5 and 4.6

Proposition 4.5 is a direct consequence of Lemma A.1. To prove Proposition 4.6, we first introduce the subexponential norm. For any random variable $X \in \mathbb{R}$, $\|X\|_{\phi_1}$ is defined as follows:

$$\|X\|_{\phi_1} := \sup_{k \ge 1} \frac{1}{k} \left(\mathbb{E}|X|^k\right)^{1/k}.$$

Let $\mathbf{S} := \operatorname{sign}(\mathbf{X} - \widetilde{\mathbf{X}})$. Because $\mathbf{v}^T\mathbf{S}$ is sub-Gaussian and $\mathbb{E}\mathbf{v}^T\mathbf{S} = 0$, using Lemma 5.14 in Vershynin [33], we get

$$\begin{aligned}\big\|\big(\mathbf{v}^T\mathbf{S}\big)^2 - \mathbb{E}\big(\mathbf{v}^T\mathbf{S}\big)^2\big\|_{\phi_1} &\le \big\|\big(\mathbf{v}^T\mathbf{S}\big)^2\big\|_{\phi_1} + \big\|\mathbf{v}^T\mathbf{T}\mathbf{v}\big\|_{\phi_1} \\ &\le 2\big\|\mathbf{v}^T\mathbf{S}\big\|_{\phi_2}^2 + \mathbf{v}^T\mathbf{T}\mathbf{v} \\ &\le (L_2+1)\|\mathbf{T}\|_2.\end{aligned}$$

Since $(\mathbf{v}^T\mathbf{S})^2 - \mathbb{E}(\mathbf{v}^T\mathbf{S})^2$ is a zero-mean random variable and $\mathbf{v}^T\mathbf{S}$ is sub-Gaussian, using Lemma 5.15 in Vershynin [33], there exist two fixed constants C', c' such that if $|t| \le c'/\|(\mathbf{v}^T\mathbf{S})^2 - \mathbb{E}(\mathbf{v}^T\mathbf{S})^2\|_{\phi_1}$, we have

$$\mathbb{E}\exp\big(t\big(\big(\mathbf{v}^T\mathbf{S}\big)^2 - \mathbb{E}\big(\mathbf{v}^T\mathbf{S}\big)^2\big)\big) \le \exp\big(C't^2\big\|\big(\mathbf{v}^T\mathbf{S}\big)^2 - \mathbb{E}\big(\mathbf{v}^T\mathbf{S}\big)^2\big\|_{\phi_1}^2\big).$$

Accordingly, by choosing $t_0 = c'(L_2+1)^{-1}\|\mathbf{T}\|_2^{-1}$ and $K = C'(L_2+1)^2$ in equation (4.2), noticing that $t_0\|\mathbf{T}\|_2 = c'(L_2+1)^{-1}$, the sign sub-Gaussian condition is satisfied.

5.4. Proof of Theorem 4.7

In this section, we provide the proof of Theorem 4.7. In detail, we show that for any transelliptically distributed random vector $\mathbf{X}$ such that $f(\mathbf{X}) \sim EC_d(\mathbf{0}, \mathbf{I}_d, \xi)$, we have that $\mathbf{X}$ satisfies the condition in equation (4.2).

Proof. Because for any strictly increasing function $g : \mathbb{R} \to \mathbb{R}$ and $x, y \in \mathbb{R}$, $\operatorname{sign}(g(x) - g(y)) = \operatorname{sign}(x - y)$, $\operatorname{sign}(\xi x) = \operatorname{sign}(x)$ (a.s.) for any ξ with $\mathbb{P}(\xi > 0) = 1$, and the fact that the elliptical family is closed to the independent sums (Lindskog et al. [26]), we only need to consider the random vector $\mathbf{X} \sim N_d(\mathbf{0}, \mathbf{I}_d)$. For $\mathbf{X} = (X_1, \dots, X_d)^T \sim N_d(\mathbf{0}, \mathbf{I}_d)$ and $\widetilde{\mathbf{X}}$ as an independent copy of $\mathbf{X}$, we have $\mathbf{X} - \widetilde{\mathbf{X}} \sim N_d(\mathbf{0}, 2\mathbf{I}_d)$. Reminding that the off-diagonal entries of $\mathbf{I}_d$ are all zero, defining $\mathbf{X}^0 = (X_1^0, \dots, X_d^0)^T = \mathbf{X} - \widetilde{\mathbf{X}}$ and

$$g\big(\mathbf{X}^0, \mathbf{v}\big) := \sum_{j,k} v_j v_k \operatorname{sign}\big(X_j^0 X_k^0\big),$$

we have

$$\big\{\mathbf{v}^T \operatorname{sign}(\mathbf{X} - \widetilde{\mathbf{X}})\big\}^2 - \mathbb{E}\big\{\mathbf{v}^T \operatorname{sign}(\mathbf{X} - \widetilde{\mathbf{X}})\big\}^2 = g\big(\mathbf{X}^0, \mathbf{v}\big) - \mathbb{E}g\big(\mathbf{X}^0, \mathbf{v}\big).$$

Accordingly, to bound $\psi(\langle \operatorname{sign}(\mathbf{X} - \widetilde{\mathbf{X}}), \mathbf{v}\rangle; 2)$, we only need to focus on $g(\mathbf{X}^0, \mathbf{v})$. Letting $\mathbf{S} := (S_1, \dots, S_d)^T$ with $S_j := \operatorname{sign}(Y_j^0)$ for $j = 1, \dots, d$. Using the property of Gaussian distribution, $S_1, \dots, S_d$ are independent Bernoulli random variables in $\{-1, 1\}$ almost surely. We then have

$$g\big(\mathbf{Y}^0, \mathbf{v}\big) - \mathbb{E}g\big(\mathbf{Y}^0, \mathbf{v}\big) = \sum_{j,k} v_j v_k \operatorname{sign}\big(Y_j^0 Y_k^0\big) - 1 = \big(\mathbf{v}^T\mathbf{S}\big)^2 - 1.$$

Here, the first equality is due to the fact that $\|\mathbf{v}\|_2 = \sum_{j=1}^d v_j^2 = 1$.

We then proceed to analyze the property of $(\mathbf{v}^T\mathbf{S})^2 - 1$. By the Hubbard–Stratonovich transform (Hubbard [21]), for any $\eta \in \mathbb{R}$,

$$\exp(\eta^2) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi}} \mathrm{e}^{-y^2/4+y\eta}\,\mathrm{d}y. \tag{5.11}$$

Using equation (5.11), we have that, for any $t > 0$,

$$\begin{aligned}
\mathbb{E}\exp\bigl[t\bigl\{(\mathbf{v}^T\mathbf{S})^2 - 1\bigr\}\bigr] &= \mathrm{e}^{-t}\mathbb{E}\mathrm{e}^{t(\mathbf{v}^T\mathbf{S})^2} \\
&= \frac{\mathrm{e}^{-t}}{\sqrt{4\pi t}} \int_{-\infty}^{+\infty} \mathrm{e}^{-y^2/4t}\mathbb{E}\mathrm{e}^{y\sum_{j=1}^d v_j S_j}\,\mathrm{d}y \\
&= \frac{\mathrm{e}^{-t}}{\sqrt{4\pi t}} \int_{-\infty}^{+\infty} \mathrm{e}^{-y^2/4t} \prod_{j=1}^d \frac{1}{2}\bigl(\mathrm{e}^{yv_j} + \mathrm{e}^{-yv_j}\bigr)\,\mathrm{d}y.
\end{aligned}$$

For any number $z \in \mathbb{N}$, we define $z!$ to represent the factorial product of z. Because for any $a \in \mathbb{R}$, by Taylor expansion, we have

$$\bigl\{\exp(a) + \exp(-a)\bigr\}/2 = \sum_{k=0}^{\infty} a^{2k}/(2k)! \quad \text{and} \quad \exp\bigl(a^2/2\bigr) = \sum_{k=0}^{\infty} a^{2k}/\bigl(2^k \cdot k!\bigr).$$

Because $(2k)! > 2^k \cdot k!$, we have

$$\bigl\{\exp(a) + \exp(-a)\bigr\}/2 \le \exp\bigl(a^2/2\bigr).$$

Accordingly, we have for any $0 < t < 1/4$,

$$\begin{aligned}
\mathbb{E}\exp\bigl[t\bigl\{(\mathbf{v}^T\mathbf{S})^2 - 1\bigr\}\bigr] &= \frac{\mathrm{e}^{-t}}{\sqrt{4\pi t}} \int_{-\infty}^{+\infty} \mathrm{e}^{-y^2/4t} \prod_{j=1}^d \frac{1}{2}\bigl(\mathrm{e}^{yv_j} + \mathrm{e}^{-yv_j}\bigr)\,\mathrm{d}y \\
&\le \frac{\mathrm{e}^{-t}}{\sqrt{4\pi t}} \int_{-\infty}^{+\infty} \mathrm{e}^{-y^2/4t}\mathrm{e}^{\sum_{j=1}^d (1/2)y^2 v_j^2}\,\mathrm{d}y \\
&= \frac{\mathrm{e}^{-t}}{\sqrt{4\pi t}} \int_{-\infty}^{+\infty} \mathrm{e}^{-y^2/4t+(1/2)y^2}\,\mathrm{d}y \\
&= \frac{\mathrm{e}^{-t}}{\sqrt{1-2t}}.
\end{aligned}$$

By Taylor expansion of $\log(1-x)$, we have that

$$\frac{1}{\sqrt{1-2t}} = \exp\Biggl\{\frac{1}{2}\sum_{k=1}^{\infty}\frac{(2t)^k}{k}\Biggr\},$$

which implies that for all $0 < t < 1/4$,

$$\frac{e^{-t}}{\sqrt{1-2t}} = \exp\left(t^2 + \frac{1}{2}\sum_{k=3}^{\infty}\frac{(2t)^k}{k}\right) \le \exp(2t^2).$$

This concludes that for $0 < t < 1/4$,

$$\mathbb{E}\exp\left[t\left\{\left(\mathbf{v}^T\mathbf{S}\right)^2 - 1\right\}\right] \le \exp(2t^2). \tag{5.12}$$

Due to that $(\mathbf{v}^T\mathbf{S})^2 \ge 0$, we can apply Theorem 2.6 in Chung and Lu [11] to control the term $\mathbb{E}\exp[t\{1-(\mathbf{v}^T\mathbf{S})^2\}]$. In detail, suppose that the random variable Y satisfying $\mathbb{E}Y = 0$, $Y \le a_0$, and $\mathbb{E}Y^2 = b_0$ for some absolute constants a_0 and b_0. Then for any $0 < t < 2/a_0$, using the proof of Theorem 2.8 in Chung and Lu [11], we have

$$\mathbb{E}e^{tY} \le \exp\left\{3b_0/2 \cdot t^2\right\}. \tag{5.13}$$

For $Y = 1 - (\mathbf{v}^T\mathbf{S})^2$, we have

$$a_0 = 1 \quad \text{and} \quad b_0 = \mathbb{E}\left(\mathbf{v}^T\mathbf{S}\right)^4 - 1 = 2 - 2\sum_{j=1}^{d} v_j^4 < 2. \tag{5.14}$$

Here, we remind that $\mathbb{E}(\mathbf{v}^T\mathbf{S})^2 = \sum_j v_j^2 = 1$. Combining equations (5.13) and (5.14) implies that for any $t > 0$,

$$\mathbb{E}\exp\left[t\left\{1 - \left(\mathbf{v}^T\mathbf{S}\right)^2\right\}\right] \le \exp\left\{3t^2\right\}. \tag{5.15}$$

Combining equations (5.12) and (5.15), we see that equation (4.2) holds with $K = 3/4$ and $t_0 = 1/4$ (reminding that here $\|\mathbf{T}\|_2 = 1$). □

5.5. Proof of Theorem 4.8 and Corollary 4.9

In this section, we prove Theorem 4.8 and Corollary 4.9. Using the same argument as in the proof of Theorem 4.7, we only need to focus on those random vectors that are Gaussian distributed.

Proof of Theorem 4.8. Assume that $\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$ is a compound symmetric matrix such that

$$\mathbf{\Sigma}_{jj} = 1 \quad \text{and} \quad \mathbf{\Sigma}_{jk} = \rho \qquad \text{for } j \ne k.$$

By the discussion on page 11 of Vershynin [33], to prove equation (4.4) holds, we only need to prove that for $0 \le \rho \le C_0$ where C_0 is some absolute constant, $\mathbf{X} = (X_1, \dots, X_d)^T \sim N_d(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{v} \in \mathbb{S}^{d-1}$, we have

$$\exp\left(t\mathbf{v}^T \operatorname{sign}(\mathbf{X} - \widetilde{\mathbf{X}})\right) \le \exp\left(c\|\mathbf{T}\|_2 t^2\right),$$

for some fixed constant c. This result can be proved as follows. Let $\eta_0, \eta_1, \ldots, \eta_d$ be i.i.d. standard Gaussian random variables, then $\mathbf{Z} := \mathbf{X} - \widetilde{\mathbf{X}}$ can be expressed as $\mathbf{Z} \stackrel{d}{=} (Z_1', \ldots, Z_d')^T$, where

$$
\begin{aligned}
Z_1' &= \sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_1, \\
Z_2' &= \sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_2, \\
&\ldots \\
Z_d' &= \sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_d.
\end{aligned}
$$

Accordingly, we have

$$
\begin{aligned}
\mathbb{E}\exp\bigl(t\mathbf{v}^T \operatorname{sign}(\mathbf{X} - \widetilde{\mathbf{X}})\bigr) &= \mathbb{E}\Biggl(\exp\Biggl(t\sum_{j=1}^{d} v_j \operatorname{sign}(\sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_j)\Biggr)\Biggr) \\
&= \mathbb{E}\Biggl(\mathbb{E}\Biggl(\exp\Biggl(t\sum_{j=1}^{d} v_j \operatorname{sign}(\sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_j)\Biggr)\Bigg|\eta_0\Biggr)\Biggr)
\end{aligned}
$$

Moreover, we have

$$
\sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_j | \eta_0 \sim N_1(\sqrt{2\rho}\eta_0, 2-2\rho). \tag{5.16}
$$

Letting $\mu := \sqrt{2\rho}\eta_0$ and $\sigma := \sqrt{2-2\rho}$, equation (5.16) implies that

$$
\mathbb{P}(\sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_j > 0 | \eta_0) = \Phi\left(\frac{\mu}{\sigma}\right),
$$

where $\Phi(\cdot)$ is the CDF of the standard Gaussian. This further implies that

$$
\operatorname{sign}(\sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_j) | \eta_0 \sim \operatorname{Bern}\left(\Phi\left(\frac{\mu}{\sigma}\right)\right),
$$

where we denote $Y \sim \operatorname{Bern}(p)$ if $\mathbb{P}(Y = 1) = p$ and $\mathbb{P}(Y = -1) = 1 - p$. Accordingly, letting $\alpha := \Phi(\mu/\sigma)$, we have

$$
\mathbb{E}\bigl(\exp\bigl(tv_j \operatorname{sign}(\sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_j)\bigr)|\eta_0\bigr) = (1-\alpha)\mathrm{e}^{-v_j t} + \alpha \mathrm{e}^{v_j t}.
$$

Letting $\beta := \alpha - 1/2$, we have

$$
\mathbb{E}\bigl(\exp\bigl(tv_j \operatorname{sign}(\sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_j)\bigr)|\eta_0\bigr) = \tfrac{1}{2}\mathrm{e}^{-v_j t} + \tfrac{1}{2}\mathrm{e}^{v_j t} + \beta\bigl(\mathrm{e}^{v_j t} - \mathrm{e}^{-v_j t}\bigr).
$$

Using that fact that $\frac{1}{2}\mathrm{e}^{a} + \frac{1}{2}\mathrm{e}^{-a} \le \mathrm{e}^{a^2/2}$, we have

$$
\mathbb{E}\bigl(\exp\bigl(tv_j \operatorname{sign}(\sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_j)\bigr)|\eta_0\bigr) \le \exp\bigl(v_j^2 t^2/2\bigr) + \beta\bigl(\mathrm{e}^{v_j t} - \mathrm{e}^{-v_j t}\bigr).
$$

Because conditioning on η_0, $\operatorname{sign}(\sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_j)$, $j = 1, \dots, d$, are independent of each other, we have

$$\begin{aligned}
&\mathbb{E}\left(\exp\left(t\sum_{j=1}^{d} v_j \operatorname{sign}(\sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_j)\right)\middle|\eta_0\right)\\
&\quad\le \prod_{j=1}^{d}\bigl\{\exp\bigl(v_j^2t^2/2\bigr) + \beta\bigl(\mathrm{e}^{v_jt} - \mathrm{e}^{-v_jt}\bigr)\bigr\}\\
&\quad= \mathrm{e}^{t^2/2}\left(1 + \sum_{k=1}^{d}\beta^k \sum_{j_1<j_2<\cdots<j_k}\ \prod_{j\in\{j_1,\dots,j_k\}} \frac{\mathrm{e}^{v_jt} - \mathrm{e}^{-v_jt}}{\mathrm{e}^{v_j^2t^2/2}}\right).
\end{aligned}$$

Moreover, for any centered Gaussian distribution $Y \sim N_1(0, \kappa)$ and $t \in \mathbb{R}$, we have

$$\begin{aligned}
\mathbb{P}\bigl(\Phi(Y) > 1/2 + t\bigr) &= \mathbb{P}\bigl(Y > \Phi^{-1}(1/2 + t)\bigr)\\
&= \mathbb{P}\bigl(Y > -\Phi^{-1}(1/2 - t)\bigr)\\
&= \mathbb{P}\bigl(Y < \Phi^{-1}(1/2 - t)\bigr)\\
&= \mathbb{P}\bigl(\Phi(Y) < 1/2 - t\bigr).
\end{aligned}$$

Combined with the fact that $\Phi(Y) \in [0, 1]$, we have

$$\mathbb{E}\bigl(\Phi(Y) - 1/2\bigr)^k = 0 \qquad \text{when } k \text{ is odd.}$$

This implies that when k is odd,

$$\mathbb{E}\beta^k = 0 = \mathbb{E}\bigl(\Phi\bigl(\sqrt{\rho/(1-\rho)}\eta_0\bigr) - \tfrac{1}{2}\bigr)^k = 0.$$

Accordingly, denoting $\varepsilon = \mathbb{E}\exp(t\sum_{j=1}^{d} v_j \operatorname{sign}(\sqrt{2\rho}\eta_0 + \sqrt{2-2\rho}\eta_j))$, we have

$$\varepsilon \le \mathrm{e}^{t^2/2}\left(1 + \sum_{k \text{ is even}} \mathbb{E}\beta^k \sum_{j_1<j_2<\cdots<j_k}\ \prod_{j\in\{j_1,\dots,j_k\}} \frac{\mathrm{e}^{v_jt} - \mathrm{e}^{-v_jt}}{\mathrm{e}^{v_j^2t^2/2}}\right).$$

Using the fact that

$$\begin{aligned}
\bigl|\mathrm{e}^{a} - \mathrm{e}^{-a}\bigr| &= \left|\sum_{j=1}^{\infty}\frac{a^j}{j!} - \sum_{j=1}^{\infty}\frac{(-a)^j}{j!}\right|\\
&= 2\left|\sum_{m=0}^{\infty}\frac{a^{2m+1}}{(2m+1)!}\right|\\
&= 2|a|\cdot\left|\sum_{m=0}^{\infty}\frac{a^{2m}}{(2m+1)!}\right|
\end{aligned}$$

$$\leq 2|a|\exp\bigl(a^2/2\bigr),$$

we further have

$$\begin{aligned}\varepsilon &\leq e^{t^2/2}\biggl(1+\sum_{k \text{ is even}}\mathbb{E}\beta^k\sum_{j_1<j_2<\cdots<j_k}\prod_{j\in\{j_1,\ldots,j_k\}}2|v_j t|\biggr)\\ &= e^{t^2/2}\biggl(1+\sum_{k \text{ is even}}\mathbb{E}\beta^k\bigl(2|t|\bigr)^k\sum_{j_1<j_2<\cdots<j_k}|v_{j_1}\cdots v_{j_k}|\biggr).\end{aligned}$$

By Maclaurin's inequality, for any $x_1,\ldots,x_d \geq 0$, we have

$$\frac{x_1+\cdots+x_n}{n} \geq \biggl(\frac{\sum_{1\leq i<j\leq n}x_i x_j}{\binom{n}{2}}\biggr)^{1/2} \geq \cdots \geq (x_1\cdots x_n)^{1/n}.$$

Accordingly,

$$\begin{aligned}&e^{t^2/2}\biggl(1+\sum_{k \text{ is even}}\mathbb{E}\beta^k\bigl(2|t|\bigr)^k\sum_{j_1<j_2<\cdots<j_k}|v_{j_1}\cdots v_{j_k}|\biggr)\\ &\quad\leq e^{t^2/2}\biggl(1+\sum_{k \text{ is even}}\mathbb{E}\beta^k\bigl(2|t|\bigr)^k\biggl\{\binom{n}{2}\cdot\bigl(\|\mathbf{v}\|_1/d\bigr)^k\biggr\}\biggr)\\ &\quad\leq e^{t^2/2}\biggl(1+\sum_{k \text{ is even}}\mathbb{E}\beta^k\bigl(2|t|\bigr)^k d^{k/2}(e/k)^k\biggr).\end{aligned}\tag{5.17}$$

The last inequality is due to the fact that $\|\mathbf{v}\|_1 \leq \sqrt{d}\|\mathbf{v}\|_2 = \sqrt{d}$ and $\binom{n}{2} \leq (ed/k)^k$.

Finally, we analyze $\mathbb{E}\beta^{2m}$ for $m = 1,2,\ldots$. Reminding that

$$\beta := \Phi\biggl(\sqrt{\frac{\rho}{1-\rho}}\eta_0\biggr)-\frac{1}{2},$$

consider the function $f(x): x \to \Phi(\sqrt{\rho/(1-\rho)}x)$, we have

$$\bigl|f'(x)\bigr| = \sqrt{\frac{\rho}{1-\rho}}\cdot\frac{1}{\sqrt{2\pi}}\exp\biggl(-\frac{\rho}{2(1-\rho)}x^2\biggr) \leq \sqrt{\frac{\rho}{2\pi(1-\rho)}}.$$

Accordingly, $f(\cdot)$ is a Lipschitz function with a Lipschitz constant $K_0 := \sqrt{\frac{\rho}{2\pi(1-\rho)}}$. By the concentration of Lipschitz functions of Gaussian (Ledoux [25]), we have

$$\mathbb{P}\bigl(|\beta|>t\bigr) = \mathbb{P}\bigl(\bigl|f(\eta_0)-\mathbb{E}f(\eta_0)\bigr|>t\bigr) \leq 2\exp\bigl(-t^2/\bigl(2K_0^2\bigr)\bigr).$$

This implies that, for $m = 1,2,\ldots,$

$$\mathbb{E}\beta^{2m} = 2m\int_0^\infty t^{2m-1}\mathbb{P}\bigl(|\beta|>t\bigr)\,\mathrm{d}t$$

$$
\begin{aligned}
&\leq 4m \int_0^\infty t^{2m-1} \exp\bigl(-t^2/\bigl(2K_0^2\bigr)\bigr)\,\mathrm{d}t \\
&= 4m(\sqrt{2}K_0)^{2m} \int_0^\infty t^{2m-1} \exp\bigl(-t^2\bigr)\,\mathrm{d}t \\
&= 2m\bigl(2K_0^2\bigr)^m \int_0^\infty t^{m-1} \exp(-t)\,\mathrm{d}t.
\end{aligned}
$$

Using the fact that $\int_0^\infty \exp(-t)\,\mathrm{d}t = 1$ and for any $m \geq 1$,

$$
m \int_0^\infty t^{m-1} \exp(-t)\,\mathrm{d}t = \int_0^\infty \exp(-t)\,\mathrm{d}t^m = \int_0^\infty t^m \exp(-t)\,\mathrm{d}t,
$$

we have for $m \in \mathbb{Z}^+$, $\int_0^\infty t^m \exp(-t)\,\mathrm{d}t = m!$. Accordingly,

$$
\mathbb{E}\beta^{2m} \leq 2m\bigl(2K_0^2\bigr)^m (m-1)! = 2\bigl(2K_0^2\bigr)^m m!.
$$

Plugging the above result into equation (5.17), we have

$$
\begin{aligned}
\varepsilon &\leq \mathrm{e}^{t^2/2}\left(1 + \sum_{m=1}^\infty 2\bigl(2K_0^2\bigr)^m m!(2t)^{2m} d^m \bigl(e/(2m)\bigr)^{2m}\right) \\
&= \mathrm{e}^{t^2/2}\left(1 + \sum_{m=1}^\infty \bigl(K_0^2 d\bigr)^m \cdot m!2(2\sqrt{2}et)^{2m}/(2m)^{2m}\right).
\end{aligned}
$$

Reminding that $\rho \leq C_0$ and $K_0 := \sqrt{\frac{\rho}{2\pi(1-\rho)}} \leq \sqrt{\frac{\rho}{2\pi(1-C_0)}}$, we have

$$
\begin{aligned}
\varepsilon &\leq \mathrm{e}^{t^2/2}\left(1 + \sum_{m=1}^\infty \bigl(K_0^2 d\bigr)^m \cdot m!2(2\sqrt{2}et)^{2m}/(2m)^{2m}\right) \\
&\leq \mathrm{e}^{t^2/2}\left(1 + \sum_{m=1}^\infty m!2\left(2\sqrt{\frac{d\rho}{\pi(1-C_0)}}et\right)^{2m}/(2m)^{2m}\right).
\end{aligned}
$$

Finally, we have for any $m \geq 1$

$$
2m! \cdot m! \leq (2m)^{2m},
$$

implying that

$$
\varepsilon \leq \mathrm{e}^{t^2/2} \cdot \exp\bigl(4d\rho\mathrm{e}^2/\pi \cdot t^2\bigr) = \exp\left\{\left(\frac{1}{2} + \frac{4d\rho\mathrm{e}^2}{\pi(1-C_0)}\right)t^2\right\}, \tag{5.18}
$$

where the term $\frac{1}{2} + \frac{4d\rho\mathrm{e}^2}{\pi(1-C_0)}$ is in the same scale of $\|\mathbf{T}\|_2 = 1 + (d-1) \cdot \frac{2}{\pi}\arcsin(\rho)$. This completes the proof. $\square$

Corollary 4.9 can be proved similar to Theorem 4.8.

Proof of Corollary 4.9. Letting $J_k = \{1 + \sum_{j=1}^{k-1} d_j, \dots, \sum_{j=1}^{k} d_j\}$. By the product structure of the Gaussian distribution, we have

$$\mathbb{E}\exp\big(t\mathbf{v}^T \operatorname{sign}(\mathbf{X} - \widetilde{\mathbf{X}})\big) = \prod_{k=1}^{q} \mathbb{E}\exp\big(t\mathbf{v}_{J_k}^T \operatorname{sign}(\mathbf{X} - \widetilde{\mathbf{X}})_{J_k}\big).$$

Here we note that the bound in equation (5.18) also holds for each $\mathbb{E}\exp(t\mathbf{v}_{J_k}^T \operatorname{sign}(\mathbf{X} - \widetilde{\mathbf{X}})_{J_k})$ by checking equation (5.17). Accordingly,

$$\begin{aligned}\prod_{k=1}^{q} \mathbb{E}\exp\big(t\mathbf{v}_{J_k}^T \operatorname{sign}(\mathbf{X} - \widetilde{\mathbf{X}})_{J_k}\big) &\le \prod_{k=1}^{q} \exp\left\{\left(\frac{1}{2} + \frac{4d_k\rho_k \mathrm{e}^2}{\pi(1-C_1)}\right)t^2\right\} \\ &\le \exp\left\{t^2\left(\frac{q}{2} + \frac{4\mathrm{e}^2 q}{\pi(1-C_1)}\max_k(d_k\rho_k)\right)\right\}.\end{aligned}$$

Because q is upper bounded by a fixed constant, we have $\mathbf{v}^T \operatorname{sign}(\mathbf{X} - \widetilde{\mathbf{X}})$ is sub-Gaussian. This completes the proof. $\square$

5.6. Proof of Theorem 4.11

Proof. We first prove that (4.6) in Theorem 4.11 holds. Letting $\zeta := K\|\mathbf{T}\|_2^2$, we aim to prove that with probability larger than or equal to $1 - 2\alpha$,

$$\sup_{\mathbf{b}\in\mathbb{S}^{s-1}} \sup_{J_s\in\{1,\dots,d\}} \big|\mathbf{b}^T[\widehat{\mathbf{T}} - \mathbf{T}]_{J_s,J_s}\mathbf{b}\big| \le 2(8\zeta)^{1/2}\sqrt{\frac{s(3+\log(d/s)) + \log(1/\alpha)}{n}}. \tag{5.19}$$

For the sphere $\mathbb{S}^{s-1}$ equipped with Euclidean metric, we let $\mathcal{N}_\varepsilon$ be a subset of $\mathbb{S}^{s-1}$ such that for any $\mathbf{v} \in \mathbb{S}^{s-1}$, there exists $\mathbf{u} \in \mathcal{N}_\varepsilon$ subject to $\|\mathbf{u} - \mathbf{v}\|_2 \le \varepsilon$. The cardinal number of $\mathcal{N}_\varepsilon$ has the upper bound

$$\operatorname{card}(\mathcal{N}_\varepsilon) < \left(1 + \frac{2}{\varepsilon}\right)^s.$$

Let $\mathcal{N}_{1/4}$ be a $(1/4)$-net of $\mathbb{S}^{s-1}$. Then the cardinality of $\mathcal{N}_{1/4}$ is bounded by 9^s. Moreover, for any symmetric matrix $\mathbf{M} \in \mathbb{R}^{s\times s}$,

$$\sup_{\mathbf{v}\in\mathbb{S}^{s-1}} \big|\mathbf{v}^T\mathbf{M}\mathbf{v}\big| \le \frac{1}{1-2\varepsilon} \sup_{\mathbf{v}\in\mathcal{N}_\varepsilon} \big|\mathbf{v}^T\mathbf{M}\mathbf{v}\big|.$$

This implies that

$$\sup_{\mathbf{v}\in\mathbb{S}^{s-1}} \big|\mathbf{v}^T\mathbf{M}\mathbf{v}\big| \le 2 \sup_{\mathbf{v}\in\mathcal{N}_{1/4}} \big|\mathbf{v}^T\mathbf{M}\mathbf{v}\big|.$$

Let $\beta > 0$ be a constant defined as

$$\beta := (8\zeta)^{1/2}\sqrt{\frac{s(3+\log(d/s))+\log(1/\alpha)}{n}}.$$

We have

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\mathbf{b}\in\mathbb{S}^{s-1}}\sup_{J_s\subset\{1,\ldots,d\}}\big|\mathbf{b}^T[\widehat{\mathbf{T}}-\mathbf{T}]_{J_s,J_s}\mathbf{b}\big| > 2\beta\Big)\\
&\leq \mathbb{P}\Big(\sup_{\mathbf{b}\in\mathcal{N}_{1/4}}\sup_{J_s\subset\{1,\ldots,d\}}\big|\mathbf{b}^T[\widehat{\mathbf{T}}-\mathbf{T}]_{J_s,J_s}\mathbf{b}\big| > \beta\Big)\\
&\leq 9^s\binom{d}{s}\mathbb{P}\Bigg(\big|\mathbf{b}^T[\widehat{\mathbf{T}}-\mathbf{T}]_{J_s,J_s}\mathbf{b}\big| > (8\zeta)^{1/2}\sqrt{\frac{s(3+\log(d/s))+\log(1/\alpha)}{n}},\\
&\qquad\qquad \text{for fixed } \mathbf{b} \text{ and } J_s\Bigg).
\end{aligned}$$

Thus, if we can show that for any fixed $\mathbf{b}$ and J_s holds

$$\mathbb{P}\big(\big|\mathbf{b}^T[\widehat{\mathbf{T}}-\mathbf{T}]_{J_s,J_s}\mathbf{b}\big| > t\big) \leq 2\mathrm{e}^{-nt^2/(8\zeta)}, \tag{5.20}$$

then using the bound $\binom{d}{s} < \{ed/(s)\}^s$, we have

$$\begin{aligned}
&9^s\binom{d}{s}\mathbb{P}\Bigg(\big|\mathbf{b}^T[\widehat{\mathbf{T}}-\mathbf{T}]_{J_s,J_s}\mathbf{b}\big| > (8\zeta)^{1/2}\sqrt{\frac{s(3+\log(d/s))+\log(1/\alpha)}{n}}, \text{ for fixed } \mathbf{b} \text{ and } J\Bigg)\\
&\leq 2\exp\big\{s(1+\log 9-\log s)+s\log d - s(3+\log d-\log s)-\log(1/\alpha)\big\}\\
&\leq 2\alpha.
\end{aligned}$$

It gives that with probability greater than $1-2\alpha$ the bound in equation (5.19) holds.

Finally, we show that equation (5.20) holds. For any t, we have

$$\begin{aligned}
&\mathbb{E}\exp\big\{t\cdot\mathbf{b}^T[\widehat{\mathbf{T}}-\mathbf{T}]_{J_s,J_s}\mathbf{b}\big\}\\
&=\mathbb{E}\exp\Bigg\{t\cdot\sum_{j\neq k\in J_s} b_j b_k(\widehat{\tau}_{jk}-\tau_{jk})\Bigg\}\\
&=\mathbb{E}\exp\Bigg\{t\cdot\frac{1}{\binom{n}{2}}\sum_{i<i'}\sum_{j\neq k\in J_s} b_j b_k\big(\mathrm{sign}\big((\mathbf{x}_i-\mathbf{x}_{i'})_j(\mathbf{x}_i-\mathbf{x}_{i'})_k\big)-\tau_{jk}\big)\Bigg\}.
\end{aligned}$$

Let S_n represent the permutation group of $\{1,\ldots,n\}$. For any $\sigma \in S_n$, let $(i_1,\ldots,i_n) := \sigma(1,\ldots,n)$ represent a permuted series of $\{1,\ldots,n\}$ and $\mathrm{O}(\sigma) := \{(i_1,i_2),(i_3,i_4),\ldots,(i_{n-1},i_n)\}$. In particular, we denote $\mathrm{O}(\sigma_0) := \{(1,2),(3,4),\ldots,(n-1,n)\}$. By simple calculation,

$$
\begin{aligned}
&\mathbb{E}\exp\Big\{t\cdot\frac{1}{\binom{n}{2}}\sum_{i<i'}\sum_{j\neq k\in J_s} b_j b_k\big(\operatorname{sign}\big((\mathbf{x}_i-\mathbf{x}_{i'})_j(\mathbf{x}_i-\mathbf{x}_{i'})_k\big)-\tau_{jk}\big)\Big\}\\
&\quad=\mathbb{E}\exp\Big\{t\cdot\frac{1}{\operatorname{card}(S_n)}\sum_{\sigma\in S_n}\frac{2}{n}\sum_{(i,i')\in \mathrm{O}(\sigma)}\sum_{j\neq k\in J_s} b_j b_k\big(\operatorname{sign}\big((\mathbf{x}_i-\mathbf{x}_{i'})_j(\mathbf{x}_i-\mathbf{x}_{i'})_k\big)-\tau_{jk}\big)\Big\}\\
&\quad\le\frac{1}{\operatorname{card}(S_n)}\sum_{\sigma\in S_n}\mathbb{E}\exp\Big\{t\cdot\frac{2}{n}\sum_{(i,i')\in \mathrm{O}(\sigma)}\sum_{j\neq k\in J_s} b_j b_k\big(\operatorname{sign}\big((\mathbf{x}_i-\mathbf{x}_{i'})_j(\mathbf{x}_i-\mathbf{x}_{i'})_k\big)-\tau_{jk}\big)\Big\}\\
&\quad=\mathbb{E}\exp\Big\{t\cdot\frac{2}{n}\sum_{(i,i')\in \mathrm{O}(\sigma_0)}\sum_{j\neq k\in J_s} b_j b_k\big(\operatorname{sign}\big((\mathbf{x}_i-\mathbf{x}_{i'})_j(\mathbf{x}_i-\mathbf{x}_{i'})_k\big)-\tau_{jk}\big)\Big\}.
\end{aligned}
\tag{5.21}
$$

The inequality is due to the Jensen's inequality.

Let $m := n/2$ and remind that $\mathbf{X}=(X_1,\dots,X_d)^T\sim TE_d(\boldsymbol{\Sigma},\xi;f_1,\dots,f_d)$. Let $\widetilde{\mathbf{X}}=(\widetilde{X}_1,\dots,\widetilde{X}_d)^T$ be an independent copy of $\mathbf{X}$. By equation (4.2), we have that for any $|t|<t_0$ and $\mathbf{v}\in\mathbb{S}^{d-1}$,

$$
\mathbb{E}\exp\big[t\big\{\big(\mathbf{v}^T\operatorname{sign}(\mathbf{X}-\widetilde{\mathbf{X}})\big)^2-\mathbb{E}\big(\mathbf{v}^T\operatorname{sign}(\mathbf{X}-\widetilde{\mathbf{X}})\big)^2\big\}\big]\le e^{\zeta t^2}.
$$

In particular, letting $\mathbf{v}_{J_s}=\mathbf{b}$ and $\mathbf{v}_{J_s^C}=\mathbf{0}$, we have

$$
\mathbb{E}\exp\Big\{t\sum_{j\neq k\in J_s} b_j b_k\big(\operatorname{sign}\big((\mathbf{X}-\widetilde{\mathbf{X}})_j(\mathbf{X}-\widetilde{\mathbf{X}})_k\big)-\tau_{jk}\big)\Big\}\le e^{\zeta t^2}.
\tag{5.22}
$$

Then we are able to continue equation (5.21) as

$$
\begin{aligned}
&\mathbb{E}\exp\Big\{t\cdot\frac{2}{n}\sum_{(i,i')\in \mathrm{O}(\sigma_0)}\sum_{j\neq k\in J_s} b_j b_k\big(\operatorname{sign}\big((\mathbf{x}_i-\mathbf{x}_{i'})_j(\mathbf{x}_i-\mathbf{x}_{i'})_k\big)-\tau_{jk}\big)\Big\}\\
&\quad=\mathbb{E}\exp\Big\{\frac{t}{m}\sum_{i=1}^{m}\Big\{\sum_{j\neq k\in J_s} b_j b_k\big(\operatorname{sign}\big((\mathbf{x}_{2i}-\mathbf{x}_{2i-1})_j(\mathbf{x}_{2i}-\mathbf{x}_{2i-1})_k\big)-\tau_{jk}\big)\Big\}\Big\}\\
&\quad=\big(\mathbb{E}e^{(t/m)(\operatorname{sign}((\mathbf{X}-\widetilde{\mathbf{X}})_j(\mathbf{X}-\widetilde{\mathbf{X}})_k)-\tau_{jk})}\big)^m\\
&\quad\le e^{\zeta t^2/m},
\end{aligned}
\tag{5.23}
$$

where by equation (4.2), the last inequality holds for any $|t/m|<t_0$. Accordingly, choosing $t=\beta m/(2\zeta)$, by Markov inequality, we have for sufficiently large n,

$$
\mathbb{P}\big(\mathbf{b}^T[\widehat{\mathbf{T}}-\mathbf{T}]_{J_s,J_s}\mathbf{b}>\beta\big)\le e^{-n\beta^2/(8\zeta)}\qquad\text{for all }\beta<2\zeta t_0.
\tag{5.24}
$$

Because $t_0\|\mathbf{T}\|_2>C$ for some generic constant C, we have $2\zeta t_0\ge 2CK^{1/2}\zeta^{1/2}$, and hence as long as $\beta\le 2CK^{1/2}\zeta^{1/2}$, (5.24) holds.

By symmetry, we have the same bound for $\mathbb{P}(\mathbf{b}^T[\widehat{\mathbf{T}} - \mathbf{T}]_{J_s,J_s}\mathbf{b} < -\beta)$ as in equation (5.24). Together they give us equation (5.20). This completes the proof of the first part.

Using (4.6), we can now proceed to the quantify the term

$$\sup_{\mathbf{v}\in\mathbb{S}^{d-1},\|\mathbf{v}\|_0\le s} \left|\mathbf{v}^T(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})\mathbf{v}\right|.$$

We aim to prove that, under the conditions in Theorem 4.11, we have with probability larger than or equal to $1 - 2\alpha - \alpha^2$,

$$\begin{aligned}
&\sup_{\mathbf{b}\in\mathbb{S}^{s-1}} \sup_{J_s\in\{1,\dots,d\}} \left|\mathbf{b}^T[\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}]_{J_s,J_s}\mathbf{b}\right| \\
&\quad \le \pi^2(8\zeta)^{1/2}\sqrt{\frac{s(3+\log(d/s))+\log(1/\alpha)}{n}} + \pi^2\cdot\frac{s\log(d/\alpha)}{n}.
\end{aligned} \tag{5.25}$$

Using a similar argument as in the proof of Theorem 3.2, we let $\mathbf{E}_1, \mathbf{E}_2 \in \mathbb{R}^{d\times d}$, satisfying that for $j \neq k$,

$$\begin{aligned}
[\mathbf{E}_1]_{jk} &= \cos\left(\frac{\pi}{2}\tau_{jk}\right)\frac{\pi}{2}(\widehat{\tau}_{jk} - \tau_{jk}), \\
[\mathbf{E}_2]_{jk} &= -\frac{1}{2}\sin(\theta_{jk})\left(\frac{\pi}{2}\right)^2(\widehat{\tau}_{jk} - \tau_{jk})^2,
\end{aligned}$$

where θ_{jk} lies between τ_{jk} and $\widehat{\tau}_{jk}$. We then have

$$\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma} = \mathbf{E}_1 + \mathbf{E}_2.$$

Let the event Ω_2 be defined as

$$\Omega_2 := \left\{\exists 1 \le j \neq k \le d, \left|[\mathbf{E}_2]_{jk}\right| > \pi^2\frac{\log(d/\alpha)}{n}\right\}.$$

Using the result in the proof of Theorem 3.2, we have $\mathbb{P}(\Omega_2) \le \alpha^2$. Moreover, conditioning on Ω_2, for any $J_s \in \{1,\dots,d\}$ and $\mathbf{b} \in \mathbb{S}^{s-1}$,

$$\begin{aligned}
\left|\mathbf{b}^T[\mathbf{E}_2]_{J_s,J_s}\mathbf{b}\right| &\le \sqrt{\sum_{j,k\in J_s}[\mathbf{E}_2]_{jk}^2}\cdot\|\mathbf{b}\|_2^2 \\
&\le s\cdot\pi^2\cdot\frac{\log(d/\alpha)}{n} \\
&= \pi^2\cdot\frac{s\log(d/\alpha)}{n}.
\end{aligned} \tag{5.26}$$

We then proceed to control the term $|\mathbf{b}^T[\mathbf{E}_1]_{J_s,J_s}\mathbf{b}|$. Using a similar argument as shown in equation (5.8), for $\mathbf{Y} = (Y_1,\dots,Y_d)^T \sim N_d(\mathbf{0},\mathbf{\Sigma})$, any symmetric matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$, $\mathbf{W}$ with

$\mathbf{W}_{jk} = \frac{\pi}{2}\cos(\frac{\pi}{2}\tau_{jk})$ and $\mathbf{v} \in \mathbb{S}^{d-1}$ with $\|\mathbf{v}\|_0 \le q$, we have

$$\begin{aligned}
\left|\mathbf{v}^T\mathbf{M} \circ \mathbf{W}\mathbf{v}\right| &\le \frac{\pi^2}{4}\mathbb{E}\Bigg(\bigg|\sum_{j,k} v_j v_k \mathbf{M}_{jk}|Y_j Y_k|\bigg| + \bigg|\sum_{j,k} v_j v_k \mathbf{M}_{jk} Y_j Y_k \operatorname{sign}\big(Y_j' Y_k'\big)\bigg|\Bigg) \\
&\le \frac{\pi^2}{4} \sup_{\mathbf{b}\in\mathbb{S}^{d-1}, \|\mathbf{b}\|_0 \le q} \left|\mathbf{b}^T\mathbf{M}\mathbf{b}\right| \cdot \mathbb{E}\bigg(2\sum_j v_j^2 Y_j^2\bigg) \\
&= \frac{\pi^2}{4} \sup_{\mathbf{b}\in\mathbb{S}^{d-1}, \|\mathbf{b}\|_0 \le q} \left|\mathbf{b}^T\mathbf{M}\mathbf{b}\right| \cdot \bigg(2\sum_j v_j^2\bigg) \\
&= \frac{\pi^2}{2} \sup_{\mathbf{b}\in\mathbb{S}^{d-1}, \|\mathbf{b}\|_0 \le q} \left|\mathbf{b}^T\mathbf{M}\mathbf{b}\right|.
\end{aligned}$$

Accordingly, we have

$$\sup_{\mathbf{b}\in\mathbb{S}^{s-1}} \sup_{J_s\in\{1,\ldots,d\}} \left|\mathbf{b}^T[\mathbf{E}_1]_{J_s,J_s}\mathbf{b}\right| \le \frac{\pi^2}{2} \sup_{\mathbf{b}\in\mathbb{S}^{s-1}} \sup_{J_s\in\{1,\ldots,d\}} \left|\mathbf{b}^T[\widehat{\mathbf{T}} - \mathbf{T}]_{J_s,J_s}\mathbf{b}\right|.$$

Combined with equations (4.6), (5.26) and (5.10), we have the desired result in (4.7). $\square$

6. Discussions

This paper considers robust estimation of the correlation matrix using the rank-based correlation coefficient estimator Kendall's tau and its transformed version. We showed that the Kendall's tau is an very robust estimator in high dimensions, in terms of that it can achieve the parametric rate of convergence under various norms without any assumption on the data distribution, and in particular, without assuming any moment constraints. We further consider the transelliptical family proposed in Han and Liu [17], showing that a transformed version of the Kendall's tau attains the parametric rate in estimating the latent Pearson's correlation matrix without assuming any moment constraints. Moreover, unlike the Gaussian case, the theoretical analysis performed here motivates new understandings on rank-based estimators as well as new proof techniques. These new understandings and proof techniques are of self-interest.

Han and Liu [15] studied the performance of the latent generalized correlation matrix estimator on dependent data under some mixing conditions and proved that $\widehat{\mathbf{\Sigma}}$ can attain a $s\sqrt{\log d/(n\gamma)}$ rate of convergence under the restricted spectral norm, where $\gamma \le 1$ reflects the impact of non-independence on the estimation accuracy. It is also interesting to consider extending the results in this paper to dependent data under similar mixing conditions and see whether a similar $\sqrt{s\log d/(n\gamma')}$ rate of convergence can be attained. However, it is much more challenging to obtain such results in dependent data. The current theoretical analysis based on U-statistics is not sufficient to achieve this goal.

A problem closely related to the leading eigenvector estimation is principal component detection, which is initiated in the work of Berthet and Rigollet [2,3]. It is interesting to extend the analysis here to this setting and conduct sparse principal component detection under the transelliptical family. It is worth pointing out that Theorems 3.2 and 4.11 in this paper can be exploited in measuring the statistical performance of the corresponding detection of sparse principal components.

Appendix

In this section, we provide a lemma quantifying the relationship between Orlicz ψ_2-norm and the sub-Gaussian condition. Although this result is well known, in order to quantify this relationship in numbers, we include a proof here. We do not claim any original contribution in this section.

Lemma A.1. *For any random variable $Y \in \mathbb{R}$, we say that Y is a sub-Gaussian random variable with factor $c > 0$ if and only if for any $t \in \mathbb{R}$, $\mathbb{E}\exp(tY) \le \exp(ct^2)$. We than have Y is sub-Gaussian if and only if $\|Y\|_{\psi_2}$ is bounded. In particular, we have that if Y is sub-Gaussian with factor c, then $\|Y\|_{\psi_2} \le \sqrt{12c}$. If $\|Y\|_{\psi_2} \le D \le \infty$, then Y is sub-Gaussian with factor $c = 5D^2/2$.*

Proof. If Y is sub-Gaussian, then for any $m > 0$, we have

$$\mathbb{E}\exp\bigl(|Y/m|^2\bigr) = 1 + \int_0^\infty \mathbb{P}\biggl(\frac{Y^2}{m^2} > t\biggr)\mathrm{e}^t\,\mathrm{d}t$$

$$= 1 + \int_0^\infty \mathbb{P}\bigl(|Y| > m\sqrt{t}\bigr)\mathrm{e}^t\,\mathrm{d}t.$$

By Markov inequality, we know that if Y is sub-Gaussian, then for any $t > 0$

$$\mathbb{P}\bigl(|Y| > t\bigr) \le 2\exp\bigl(-t^2/(4c)\bigr).$$

Accordingly, we can proceed the proof

$$\mathbb{E}\exp\bigl(|Y/m|^2\bigr) \le 1 + 2\int_0^\infty \mathrm{e}^{-m^2t/(4c)} \cdot \mathrm{e}^t\,\mathrm{d}t$$

$$= 1 + 2\int_0^\infty \mathrm{e}^{-(m^2/(4c)-1)t}\,\mathrm{d}t$$

$$= 1 + \frac{2}{m^2/(4c) - 1}.$$

Picking $m = \sqrt{12c}$, we have $\mathbb{E}\exp(|Y/m|^2) \le 2$. Accordingly, $\|Y\|_{\psi_2} \le \sqrt{12c}$. On the other hand, if $\|Y\|_{\psi_2} \le \infty$, then there exists some $m < \infty$ such that $\mathbb{E}\exp(|Y/m|^2) \le 2$. Using integration by part, it is easy to check that

$$\exp(a) = 1 + a^2 \int_0^1 (1-y)\mathrm{e}^{ay}\,\mathrm{d}y.$$

This implies that

$$\begin{aligned}\mathbb{E}\exp(tX) &= 1 + \int_0^1 (1-u)\mathbb{E}\big[(tX)^2\exp(utX)\big]\,\mathrm{d}u \\ &\le 1 + t^2\mathbb{E}\big(X^2\exp\big(|tX|\big)\big)\int_0^1 (1-u)\,\mathrm{d}u \\ &\le 1 + \frac{t^2}{2}\mathbb{E}\big(X^2\mathrm{e}^{|tX|}\big).\end{aligned}$$

Using the fact that for any $a, b \in \mathbb{R}$, $|ab| \le \frac{a^2+b^2}{2}$ and $a \le \mathrm{e}^a$, we can further prove that

$$\begin{aligned}\mathbb{E}\exp(tX) &\le 1 + \frac{t^2}{2}\mathbb{E}\big(X^2\mathrm{e}^{|tX|}\big) \\ &\le 1 + m^2t^2\mathrm{e}^{m^2t^2/2}\mathbb{E}\left(\frac{X^2}{2m^2}\mathrm{e}^{X^2/(2m^2)}\right) \\ &\le 1 + m^2t^2\mathrm{e}^{m^2t^2/2}\mathbb{E}\mathrm{e}^{X^2/m^2} \\ &\le \big(1+2m^2t^2\big)\mathrm{e}^{m^2t^2/2} \\ &\le \mathrm{e}^{5m^2t^2/2}.\end{aligned}$$

The last inequality is due to the fact that for any $a \in \mathbb{R}$, $1 + a \le \mathrm{e}^a$. Accordingly, X is sub-Gaussian with the factor $c = 5m^2/2$. □

Acknowledgement

We sincerely thank Marten Wegkamp for his very helpful discussions and generously providing independent credit for our work. We thank the Editor, Associate Editor, and two anonymous referees for their very valuable comments, which significantly improve the quality of our work. We also thank Xiuyuan Cheng, Ramon van Handel, Philippe Rigollet, and Luo Xiao for their many helps. Fang Han's research was supported by NIBIB-EB012547. Han Liu's research was supported by the NSF CAREER Award DMS-1454377, NSF IIS-1546482, NSF IIS-1408910, NSF IIS-1332109, NIH R01-MH102339, NIH R01-GM083084, and NIH R01-HG06841.

References

[1] Baik, J. and Silverstein, J.W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680

[2] Berthet, Q. and Rigollet, P. (2013). Computational lower bounds for sparse PCA. Preprint. Available at arXiv:1304.0828.
[3] Berthet, Q. and Rigollet, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann*. *Statist*. **41** 1780–1815. MR3127849
[4] Bickel, P.J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann*. *Statist*. **36** 199–227. MR2387969
[5] Bickel, P.J. and Levina, E. (2008). Covariance regularization by thresholding. *Ann*. *Statist*. **36** 2577–2604. MR2485008
[6] Boente, G., Barrerab, M.S. and Tylerc, D.E. (2012). A characterization of elliptical distributions and some optimality properties of principal components for functional data. Technical report. Available at http://www.stat.ubc.ca/~matias/Property_FPCA_rev1.pdf.
[7] Bunea, F. and Xiao, L. (2015). On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA. *Bernoulli* **21** 1200–1230. MR3338661
[8] Cai, T., Ma, Z. and Wu, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields* **161** 781–815. MR3334281
[9] Cai, T.T., Zhang, C.-H. and Zhou, H.H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann*. *Statist*. **38** 2118–2144. MR2676885
[10] Cai, T.T. and Zhou, H.H. (2012). Minimax estimation of large covariance matrices under ℓ_1-norm. *Statist*. *Sinica* **22** 1319–1349. MR3027084
[11] Chung, F. and Lu, L. (2006). *Complex Graphs and Networks*. *CBMS Regional Conference Series in Mathematics* **107**. Providence, RI: Amer. Math. Soc. MR2248695
[12] Embrechts, P., Lindskog, F. and McNeil, A. (2003). Modelling dependence with copulas and applications to risk management. *Handbook of Heavy Tailed Distributions in Finance* **8** 329–384.
[13] Fang, H.-B., Fang, K.-T. and Kotz, S. (2002). The meta-elliptical distributions with given marginals. *J*. *Multivariate Anal*. **82** 1–16. MR1918612
[14] Fang, K.T., Kotz, S. and Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions*. *Monographs on Statistics and Applied Probability* **36**. London: Chapman & Hall. MR1071174
[15] Han, F. and Liu, H. (2013). Principal component analysis on non-Gaussian dependent data. *J*. *Mach*. *Learn*. *Res*. *Workshop Conf*. *Proc*. **28** 240–248.
[16] Han, F. and Liu, H. (2014). High dimensional semiparametric scale-invariant principal component analysis. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **36** 2016–2032.
[17] Han, F. and Liu, H. (2014). Scale-invariant sparse PCA on high dimensional meta-elliptical data. *J*. *Am*. *Stat*. *Assoc*. **109** 275–287. MR3180563
[18] Han, F., Zhao, T. and Liu, H. (2013). CODA: High dimensional copula discriminant analysis. *J*. *Mach*. *Learn*. *Res*. **14** 629–671. MR3033343
[19] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J*. *Amer*. *Statist*. *Assoc*. **58** 13–30. MR0144363
[20] Hogg, R.V. and Craig, A. (2012). *Introduction to Mathematical Statistics*, 7th ed. Upper Saddle River: Harlow, Essex.
[21] Hubbard, J. (1959). Calculation of partition functions. *Phys*. *Rev*. *Lett*. **3** 77.
[22] Johnson, C.R., ed. (1990). *Matrix Theory and Applications*. *Proceedings of Symposia in Applied Mathematics* **40**. Providence, RI: Amer. Math. Soc. MR1059481
[23] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann*. *Statist*. **29** 295–327. MR1863961
[24] Jung, S. and Marron, J.S. (2009). PCA consistency in high dimension, low sample size context. *Ann*. *Statist*. **37** 4104–4130. MR2572454
[25] Ledoux, M. (2001). *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Providence, RI: Amer. Math. Soc. MR1849347

[26] Lindskog, F., McNeil, A. and Schmock, U. (2003). Kendall's tau for elliptical distributions. *Credit risk*: *Measurement*, *Evaluation and Management* 149–156.
[27] Liu, H., Han, F., Yuan, M., Lafferty, J. and Wasserman, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *Ann*. *Statist*. **40** 2293–2326. MR3059084
[28] Liu, H., Han, F. and Zhang, C.-H. (2012). Transelliptical graphical models. In *Proceedings of the Twenty-Fifth Annual Conference on Neural Information Processing Systems* 809–817.
[29] Liu, H., Lafferty, J. and Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *J*. *Mach*. *Learn*. *Res*. **10** 2295–2328. MR2563983
[30] Lounici, K. (2014). High-dimensional covariance matrix estimation with missing observations. *Bernoulli* **20** 1029–1058. MR3217437
[31] Tropp, J.A. (2012). User-friendly tail bounds for sums of random matrices. *Found*. *Comput*. *Math*. **12** 389–434. MR2946459
[32] van de Geer, S. and Lederer, J. (2013). The Bernstein–Orlicz norm and deviation inequalities. *Probab. Theory Related Fields* **157** 225–250. MR3101846
[33] Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge: Cambridge Univ. Press. MR2963170
[34] Vu, V. and Lei, J. (2012). Minimax rates of estimation for sparse PCA in high dimensions. *J*. *Mach*. *Learn*. *Res*. *Workshop Conf*. *Proc*. **22** 1278–1286.
[35] Wegkamp, M. and Zhao, Y. (2013). Analysis of elliptical copula correlation factor model with Kendall's tau. Personal communication.
[36] Xue, L. and Zou, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *Ann*. *Statist*. **40** 2541–2571. MR3097612
[37] Yuan, X.-T. and Zhang, T. (2013). Truncated power method for sparse eigenvalue problems. *J*. *Mach*. *Learn*. *Res*. **14** 899–925. MR3063614

Received November 2013 and revised November 2014