The critical threshold level on Kendall’s tau statistic concerning minimax estimation of sparse correlation matrices

Kamil Jurczak*

Ruhr-Universität Bochum

September 2, 2014

Abstract

Let $X_1, \ldots, X_n \in \mathbb{R}^p$ be a sample from an elliptical distribution with correlation matrix $\rho$ and Kendall’s tau correlation matrix $\tau$ such that the distributions of the components $X_{i1}, \ldots, X_{ip}$ have no atoms. Then $\sin\left(\frac{\pi}{2} \hat{\tau}_{ij}\right)$ is a well-behaved estimator for the entry $\rho_{ij}$, where $\hat{\tau}_{ij}$ is Kendall’s tau sample correlation based on $(X_{i1}, X_{j1}), \ldots, (X_{in}, X_{jn})$. We study the family of entrywise threshold estimators $\{\hat{\rho}_\alpha | \alpha > 0\}$, where $\hat{\rho}_\alpha = : \hat{\rho} = (\hat{\rho}_{ij})$ consists of the entries

$$\hat{\rho}_{ij} := \sin\left(\frac{\pi}{2} \hat{\tau}_{ij}\right) 1\left(\left|\sin\left(\frac{\pi}{2} \hat{\tau}_{ij}\right)\right| > \alpha \sqrt{\log p n}\right) \text{ for } i \neq j \text{ and } \hat{\rho}_{ii} = 1.$$

In particular, we raise the question how large the threshold constant $\alpha$ needs to be so that $\hat{\rho}$ attains the minimax rate under the Frobenius norm over all permissible elliptical distributions, which suffice a sparsity condition on the rows of the correlation matrix $\rho$. It is shown that $\hat{\rho}$ achieves the optimal rate $c_{n,p}(\log p n)^{1-q/2}$ if $\alpha > \pi$, where the parameters $c_{n,p}$ and $q$ depend on the class of sparse correlation matrices.

For Gaussian observations we even establish a critical threshold constant, i.e. we identify a constant $\alpha^* > 0$ such that the proposed estimator attains the minimax rate for $\alpha > \alpha^*$ but in general not for $\alpha < \alpha^*$. This critical value $\alpha^*$ is given by $\sqrt{\frac{2}{3\pi}}$.

The main ingredient to provide the critical threshold level is a sharp large deviation result for Kendall’s tau sample correlation if the underlying 2-dimensional normal distribution implies weak correlation between the components. This result is evolved from an asymptotic expansion of the number of permutations with a certain number of inversions.

To the best of the authors knowledge this is the first work concerning critical threshold constants.

1 Introduction

Let $X_1, \ldots, X_n$ be $n$ i.i.d. observations in $\mathbb{R}^p$ with population covariance matrix $\Sigma$ and correlation matrix $\rho$. The estimation of $\Sigma$ and $\rho$ is of great interest in multivariate analysis. In modern statistical problems the dimension $p$ of the observations is typically much larger than the sample size $n$. In this case the sample covariance matrix $\hat{\Sigma} = (n-1)^{-1} \sum_i (X_i - \bar{X})(X_i - \bar{X})^T$, where $\bar{X} = n^{-1} \sum_i X_i$ is the sample mean, is a poor and in general inconsistent estimator.

*Supported by the DFG research unit FG 1735, RO 3766/3-1
E-MAIL: kamil.jurczak@ruhr-uni-bochum.de
for the population covariance matrix (see Johnstone (2001) and Baik and Silverstein (2006)). Therefore the problem of estimating high-dimensional covariance matrices has been investigated under a variety of additional structural assumptions on $\Sigma$. For instance the spiked covariance model, i.e. the population covariance matrix has a representation as the sum of a diagonal and a low-rank matrix, with sparse leading eigenvectors - and variants of it - has been extensively studied - see Johnstone and Lu (2009), Johnstone and Paul (2007), D’Aspremont et al. (2007), Huang and Shen (2008), Amini and Wainwright (2009), Cai, Ma, and Wu (2013), Fan, Yuan, and Mincheva (2013), Berthet and Rigollet (2013), Vu and Lei (2013) and Cai, Ma, and Wu (2014+) among others.

A further common assumption is sparsity on the covariance matrix itself (compare Bickel and Levina (2008)(2), El Karoui (2008) and Levina and Vershynin (2012)). While in some models there are more information about the structure of the sparsity as for band covariance matrices, the case, that the position of the entries with small magnitude is unknown, is also of great importance. The latter assumption can be formulated by assuming each row to lie in a (weak) $l_q$-ball (see Bickel and Levina (2008)(1) and Cai and Zhou (2012)(2)). While Bickel and Levina (2008)(1) proved the consistency of an entrywise threshold estimator based on the sample covariance if $\frac{\log p}{n}$ tends to zero, Cai and Zhou (2012)(2) established the minimax rates of estimation for $\Sigma$ under the assumption that each row is an element of a weak $l_q$-ball and proved that the threshold estimator proposed by Bickel and Levina (2008)(1) attains the minimax rate if the constant $\alpha>0$ in the threshold level $\alpha \frac{\sqrt{\log p}}{n}$ is sufficiently large. Their results hold for sub-Gaussian random vectors $X_1, \ldots, X_n$ and therefore $\alpha$ depends on the largest $\Psi$-Orlicz norm of all 2-dimensional sub-vectors of $X_1$, where $\Psi(z) = \exp(z^2) - 1$. For general classes of distributions on $\mathbb{R}^p$ (with at least two moments) threshold levels of type $\alpha \sqrt{\log p \over n}$ are not suitable to estimate a sparse covariance matrix.

Han and Liu (2013) and Wegkamp and Zhao (2013) recently worked on a related subject. Han and Liu (2013) studied the problem of estimating the generalized latent correlation matrix of a so called transelliptical distribution, which was introduced by the same authors in Han and Liu (2014). They call a distribution transelliptical if under monotone transformations of the marginals the transformed distribution is elliptical. Then the generalized latent correlation matrix is just the correlation matrix of the transformed distribution. Wegkamp and Zhao (2013) investigate the related problem of estimating the copula correlation matrix for elliptical copulas. Han and Liu (2013) study the rate of convergence to the generalized latent correlation matrix for an transformed version of Kendall’s tau sample correlation matrix without any additional structural assumptions on the transelliptical distribution whereas Wegkamp and Zhao (2013) additionally consider copula correlation matrices with spikes. In this case the authors propose an adaptive estimator based on Kendall’s tau correlation matrix. Clearly in both models no moment assumptions are necessary.

In the present paper we assume that $X_1, \ldots, X_n \in \mathbb{R}^p$ is a sample from an elliptical distribution with correlation matrix $\rho$ and Kendall’s tau correlation matrix $\tau$ such that the distributions of the components $X_i$, $i = 1, \ldots, p$, have no atoms. Then $\sin(\frac{\pi}{2} \hat{\tau}_{ij})$ is a well-behaved estimator for the entry $\rho_{ij}$, where $\hat{\tau}_{ij}$ is Kendall’s tau sample correlation based on $(X_{i1}, X_{j1}), \ldots, (X_{in}, X_{jn})$. We study the family of entrywise threshold estimators $\{\hat{\rho}_{\alpha}|\alpha > 0\}$, where $\hat{\rho}_{\alpha} := \hat{\rho} = (\hat{\rho}_{ij})$ consists of the entries

$$\hat{\rho}_{ij} := \sin \left( \frac{\pi}{2} \hat{\tau}_{ij} \right) 1 \left( \left| \sin \left( \frac{\pi}{2} \hat{\tau}_{ij} \right) \right| > \alpha \sqrt{\frac{\log p}{n}} \right) \text{ for } i \neq j \text{ and } \hat{\rho}_{ii} = 1.$$

In particular, we raise the question how large the threshold constant $\alpha$ needs to be so that $\hat{\rho}$ attains the minimax rate under the Frobenius norm over all permissible elliptical distributions,
which suffice a sparsity condition on the rows of the correlation matrix \( \rho \). It is shown that \( \hat{\rho} \) achieves the optimal rate \( c_{n,p}(\frac{\log p}{n})^{1-q/2} \) if \( \alpha > \pi \), where the parameters \( c_{n,p} \) and \( q \) depend on the class of sparse correlation matrices.

For Gaussian observations we even establish a critical threshold constant, i.e. we identify a constant \( \alpha^* > 0 \) such that the proposed estimator attains the minimax rate for \( \alpha > \alpha^* \) but in general not for \( \alpha < \alpha^* \). The critical value \( \alpha^* \) for estimation of \( \rho \) is given by \( \frac{\pi}{3} \) and therefore by choosing \( \alpha \) slightly larger than \( \alpha^* \) the corresponding estimator is not only minimax rate optimal but provides a non-trivial estimate of the true correlation matrix even for moderate sample sizes \( n \). Furthermore we prove that for \( \alpha < \frac{\pi}{3} \) the considered estimator does not even attain the minimax rate over any regarded set of sparse correlation matrices.

Simultaneously, analogous results for the estimation of Kendall’s tau correlation matrix are established. Note that the estimation of Kendall’s tau correlation matrix is also of self-interest (see Embrechts, Lindskog, and McNeil [2003]).

The main ingredient to provide the critical threshold level is a sharp large deviation result for Kendall’s tau sample correlation if the underlying 2-dimensional normal distribution implies weak correlation between the components, which says that even quite far in the tails of the distribution of Kendall’s tau sample correlation applies the Gaussian approximation induced by the central limit theorem for Kendall’s tau sample correlation. This result is evolved from an asymptotic expansion of the number of permutations with a certain number of inversions (see Clark [2000]).

To the best of the author’s knowledge this is the first work concerning critical threshold constants.

1.1 Relation to other literature

This article is related to the works of Wegkamp and Zhao [2013] and Han and Liu [2013] about high-dimensional correlation matrix estimation based on Kendall’s tau sample correlation matrix. In contrast to their works we assume the correlation matrices to be sparse and equip them with the same weak \( l_q \)-ball sparsity condition on the rows as Cai and Zhou [2012](2) do for covariance matrices. We replace the sample covariance matrix in the hard threshold estimator of Bickel and Levina [2008](1) by a transformed version of Kendall’s tau sample correlation matrix. In contrast to Cai and Zhou [2012](2) we are mainly interested in threshold levels for which the proposed estimator attains the minimax rate. In other words the central question of this paper is how much information from the pilot estimate is permitted to retain under the restriction to obtain a rate optimal estimator. This enables us to recognize as much dependence structure in the data as possible without overfitting. Note that the permissible threshold constants for covariance matrix estimation in the inequalities (26) and (27) in Cai and Zhou [2012](2) based on Saulis and Statulevicius [1991] and Bickel and Levina [2008](2) are not given explicitly and therefore it is even vague if any practically applicable threshold estimator attains the minimax rate. Hence it is a natural question to ask how large the threshold constant in fact needs to be to get a minimax rate optimal estimator. We answer this question in the related setting stated above. Note that our estimator - introduced in the next section - does not require knowledge of any parameters of the underlying elliptical distribution.

1.2 Structure of the article

The article is structured as follows. In the next section we clarify the notation and the setting of the model. Moreover we give a brief introduction to elliptical distributions and
Kendall’s tau correlation. In the third section we discuss the minimax rates of estimation for correlation matrices and Kendall’s tau correlation matrices under the Frobenius norm in the underlying model. The fourth section is devoted to the main results of the article. We establish the critical threshold constants for Gaussian observations regarding the minimax rates from section 3. Most proofs of the results from section 3 and 4 are postponed to section 7. In section 5 we present the main new ingredients to obtain the critical threshold level of section 4, especially we provide a sharp large deviation result for Kendall’s tau sample correlation under two-dimensional normal distributions with weak correlation. Finally in section 6 the results of the article are summarized and some related problems are discussed.

Acknowledgments
I sincerely thank Fang Han for pointing out a mistake in an earlier version of this paper. This work was supported by the Deutsches Forschungsgemeinschaft research unit 1735, Ro 3766/3-1.

2 Preleminaries and model specification

2.1 Notation
We write $X \overset{d}{=} Y$ if the random variables $X$ and $Y$ have the same distribution. If $X$ is a discrete real-valued random variable then we write $\text{Im}(X)$ for the set of all $x \in \mathbb{R}$ such that $P(X = x) > 0$. A sample $Y_1, \ldots, Y_n$ of real valued random variables will be often abbreviated by $Y_{1:n}$. Moreover $\phi$ and $\Phi$ denote the probability density and cumulative distribution function of the standard normal distribution.

For a vector $x \in \mathbb{R}^p$ the mapping $i \mapsto [i] = [i]$ is a bijection on $\{1, \ldots, p\}$ such that $|x[1]| \geq \cdots \geq |x[p]|$. Clearly $[\cdot]$ is uniquely determined only if $|x_i| \neq |x_j|$ for all $i \neq j$. The $l_q$-norm of a vector $x \in \mathbb{R}^p$, $q \geq 1$, is denoted by

$$||x||_q := \left( \sum_{i=1}^{p} x_i^q \right)^{\frac{1}{q}}.$$ 

This notation will also be used if $0 < q < 1$. Then, of course, $|| \cdot ||_q$ is not a norm anymore. For $q = 0$ we write $||x||_0$ for the support of the vector $x \in \mathbb{R}^p$, that means $||x||_0$ is the number of nonzero entries of $x$, which is meanwhile a common notation from compressed sensing (see Donoho (2006) and Candès et al. (2010)).

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be an arbitrary function. Then to apply the function element wise on the matrix $A \in \mathbb{R}^{p \times d}$ we write $f[A] := (f(A_{ij}))_{i \in \{1, \ldots, p\}, j \in \{1, \ldots, d\}}$. The Frobenius norm of $A$ is defined by $||A||^2_F := \sum_{i,j} A_{ij}^2$. For a symmetric matrix $A \in \mathbb{R}^{p \times p}$ we denote the $i$-th row of $A$ without the diagonal entry by $A_i \in \mathbb{R}^{p-1}$. We will regularize correlation matrices by the threshold operator $T_\alpha := T_{\alpha,n}$, $\alpha > 0$, where $T_{\alpha,n}$ is defined on the set of all correlation matrices and satisfies for any correlation matrix $A \in \mathbb{R}^{p \times p}$ that

$$T_\alpha(A)_{ij} = \begin{cases} 1 & \text{if } |A_{ij}| > \alpha \left( \frac{\log n}{n} \right)^{1/2}, \\ 0 & \text{otherwise}, \end{cases}$$

$i \neq j$.

$C > 0$ denotes a constant factor in an inequality that does not depend on any variable contained in the inequality, in other words for the fixed value $C > 0$ the corresponding inequality holds uniformly in all variables on the left and right handside of the inequality. If we want to allow
Critical threshold levels

5

C to depend on some parameter we will add this parameter to the subscript of C. In some computations C may differ from line to line. O, o are the usual Landau symbols. Finally we write \([a]\) for the largest integer not greater than \(a \in \mathbb{R}\).

### 2.2 Sparsity condition

As mentioned earlier we want to assume that the correlation matrix and Kendall’s tau correlation matrix satisfy a sparsity condition. There are several possibilities to formulate such a sparsity assumption. One way is to reduce the permissible correlation matrices to a set

\[ G_q(c_{n,p}) := \{ A = (a_{ij})_{1 \leq i,j \leq p} : A = A^T, a_{jj} = 1, A_i \in B_q(c_{n,p}), i = 1, \ldots, p \}, \]

where \(B_q(c_{n,p}), 0 \leq q < 1\) is the \(l_q\)-ball of radius \(c_{n,p} > 0\). Bickel and Levina (2008) used this kind of sparsity condition for covariance matrix estimation. Though, it is more handy to assume that each row \(A_i\) is an element of a weak \(l_q\)-ball \(B_{w,q}(c_{n,p})\), \(0 \leq q < 1\), \(c_{n,p} > 0\) instead, in other words for \(x := A_i\) we have \(|x_{[i]}| \leq c_{n,p}^{1-\frac{1}{q}}\). Weak \(l_q\)-balls were originally introduced by Abramovich et al. (2006) in a different context. Note that the \(l_q\)-ball \(B_q(c_{n,p})\) is contained in the weak \(l_q\)-ball \(B_{w,q}(c_{n,p})\). Nevertheless the complexity of estimating a correlation matrix over \(G_{w,q}(c_{n,p})\) is the same as over \(G_q(c_{n,p})\), where

\[ G_{w,q}(c_{n,p}) := \{ A = (a_{ij})_{1 \leq i,j \leq p} : A = A^T, a_{jj} = 1, |A_i| \in B_{w,q}(c_{n,p}), i = 1, \ldots, p \}. \]

The reader is referred to Cai and Zhou (2012) for analogous statements for covariance matrix estimation. Therefore throughout the paper we will only consider the case that \(\rho, \tau \in G_{w,q}(c_{n,p})\).

### 2.3 Elliptical distributions

Commonly a random vector \(Y \in \mathbb{R}^p\) is called elliptically distributed if its characteristic function \(\varphi_Y\) is for a positive semi-definite matrix \(\bar{\Sigma} \in \mathbb{R}^{p \times p}\) and a function \(\psi\) of the form

\[ \varphi_Y(t) = e^{it^T \mu - \psi(t^T \bar{\Sigma} t)}, t \in \mathbb{R}^p. \]

The following representation for elliptically distributed random vectors will be convenient for our purposes, which may be found for example in Fang, Kotz, and Ng (1990):

**Proposition 1**

A random vector \(X \in \mathbb{R}^p\) has an elliptical distribution iff for a matrix \(A \in \mathbb{R}^{p \times q}\) with \(\text{rank}(A) = q\), a vector \(\mu \in \mathbb{R}^q\), a non-negative random variable \(\xi\) and random vector \(U \in \mathbb{R}^q\), where \(\xi\) and \(U\) are independent and \(U\) is uniformly distributed on the unit sphere \(S^{q-1}\), \(X\) has the same distribution as \(\mu + \xi AU\).

Therefore we write \(X = (X_1, \ldots, X_p)^T \sim EC_p(\mu, \Sigma, \xi)\) if \(Y\) has the same distribution as \(\mu + \xi AU\), where \(A \in \mathbb{R}^{p \times q}\) satisfies \(AA^T = \Sigma\).
2.4 Kendall’s tau

Let \((Y, Z)\) be a two-dimensional random vector. We denote Kendall’s tau correlation between \(Y\) and \(Z\) by

\[
\tau(Y, Z) := \frac{1}{n(n-1)} \sum_{k \neq l} \text{sign}(Y_k - Y_l) \text{sign}(Z_k - Z_l).
\]

Analogously we write \(\rho(X, Y)\) for the correlation between \(X\) and \(Y\). Let \(X_1 = (X_{11}, \ldots, X_{nd})^T\) be a \(p\)-dimensional random vector. We denote the Kendall’s tau correlation matrix of \(X_1\) by \(\tau := (\tau_{ij})\), where \(\tau_{ij} = \tau(X_{ij}, X_{i1})\). Thus \(\tau\) is positive semidefinite since

\[
\tau = \text{Cov}(\text{sign}(X_1 - \tilde{X}_1), \text{sign}(X_1 - \tilde{X}_1)).
\]

Moreover for an i.i.d. sample \(X_1, \ldots, X_n \in \mathbb{R}^p\) we call \(\hat{\tau} = (\hat{\tau}_{ij})\) with

\[
\hat{\tau} := \frac{1}{n(n-1)} \sum_{k \neq l} \text{sign}(X_k - X_l) \text{sign}(X_k - X_l)^T
\]

Kendall’s tau sample correlation matrix. Hence \(\hat{\tau}\) is positive semi-definite. Furthermore if the distributions of the components of \(X_1\) have no atoms, then \(\tau\) (resp. \(\hat{\tau}\)) is (resp. a.s.) a correlation matrix. Note that the distributions of the components of \(X_1\) have no atoms iff \(\text{rank}(\Sigma) = 1\) and the distribution of \(\xi\) has no atoms or \(\text{rank}(\Sigma) \geq 2\) and \(\mathbb{P}(\xi = 0) = 0\). For distributions \(\mathcal{EC}_p(\mu, \Sigma, \xi)\) such that the components have non-atomic distributions Kendall’s tau correlation matrix \(\tau\) is determined by the correlation matrix \(\rho\), particularly we have \(\rho = \sin(\frac{\pi}{2} \tau)\) (see Hult and Lindskog [2002]). Hence, \(\sin(\frac{\pi}{2} \tau)\) is a natural estimator for \(\rho\). In the following we will occasionally use the next two elementary bounds to connect the correlation matrix to Kendall’s tau correlation matrix:

**Lemma 1**

For any matrices \(A, B \in \mathbb{R}^{p \times p}\) holds

\[
||\sin(\frac{\pi}{2} A) - \sin(\frac{\pi}{2} B)||_F^2 \leq \frac{\pi^2}{4} ||A - B||_F^2.
\]

**Proof.** The function \(x \mapsto \sin(\frac{\pi}{2} x)\) is Lipschitz continuous with Lipschitz constant \(L = \frac{\pi}{2}\). Therefore we conclude

\[
||\sin(\frac{\pi}{2} A) - \sin(\frac{\pi}{2} B)||_F^2 = \sum_{i,j} (\sin(\frac{\pi}{2} A_{ij}) - \sin(\frac{\pi}{2} B_{ij}))^2 \\
\leq \sum_{i,j} \frac{\pi^2}{4} (A_{ij} - B_{ij})^2 = \frac{\pi^2}{4} ||A - B||_F^2.
\]
Lemma 2

Let \( \tau \in G_{w,q}(c_{n,p}) \) for \( c_{n,p} > 0 \), then \( \sin[\frac{\tau}{2}] \in G_{w,q}(c_{n,p}) \). On the other hand, for \( \sin[\frac{\tau}{2}] \in G_{w,q}(c_{n,p}) \) holds \( \tau \in G_{w,q}(c_{n,p}) \).

Proof. The first statement follows easily by the fact that the derivative of the sine function is bounded by 1. The second statement is obtained by concavity of the sine function on \([0, \frac{\pi}{2}]\) and convexity of the sine function on \([-\frac{\pi}{2}, 0]\).

2.5 Further model specification and the regarded threshold estimators

For a better overview we summarize the assumptions of the results in Section 3 and 4.

\((A_1)\) \(X_1, \ldots, X_n \overset{i.i.d.}{\sim} EC_p(\mu, \Sigma, \xi)\) such that the distributions of the components \(X_{1i}\) have no atoms.

\((A_1^*)\) \(X_1, \ldots, X_n \overset{i.i.d.}{\sim} N_p(\mu, \Sigma)\) such that \(\Sigma_{ii} > 0\) for all \(i = 1, \ldots, p\).

\((A_2)\) The parameters of the set \(G_{w,q}(c_{n,p})\) satisfy \(0 \leq q < 1\) and \(c_{n,p} > 0\).

\((A_3)\) There exists a constant \(M > 0\) such that \(c_{n,p} \leq Mn^{(1-q)/2}(\log p)^{-3/2}\).

\((A_3^*)\) The set \(G_{w,q}(c_{n,p})\) consists only of very sparse correlation matrices, formally \(c_{n,p} = o((\log p)^{q/2})\).

\((A_4)\) There exists a constant \(m < 0\) such that the radius \(c_{n,p} \geq 2m^q \left(\frac{\log p}{n}\right)^{q/2}\).

\((A_5)\) There exists a constant \(\eta_l > 1\) such that \(p > n^\eta\).

\((A_6)\) There exists a constant \(\eta_u > 1\) such that \(p < n^\eta_u\).

The Assumptions \((A_3)\) – \((A_5)\) are sufficient to ensure that the minimax lower bound is true. For an upper bound on the maximal risk of the considered estimators they are not required. Assumptions \((A_1^*)\) and \((A_6)\) guarantee that the entries \(\hat{\tau}_{ij}\) based on components with weak correlation satisfy a Gaussian approximation even sufficiently far in the tails. This is useful to provide the critical threshold level.

We study three highly related threshold estimators based on Kendall’s tau sample correlation matrix \(\hat{\tau}\). For the estimation of Kendall’s tau correlation matrix we investigate \(\hat{\tau}^* := \hat{\tau}^*_{\alpha} := T_\alpha(\hat{\tau})\). Based on Kendall’s tau correlation matrix two natural estimators appear for correlation matrix estimation. We consider both, \(\hat{\rho}^* := \hat{\rho}^*_{\alpha} := \sin(\pi \hat{\tau}^*)\) and \(\hat{\rho} := \hat{\rho}_{\alpha} := T_\alpha(\sin(\pi \hat{\tau}))\). The difference between them is in the order of thresholding and transformation by the sine function. Technically it is favorable to deduce the properties of \(\hat{\rho}\) from \(\hat{\rho}^*\). In the next section we write \(\hat{\rho}\) and \(\hat{\tau}\) also for an arbitrary estimator. The meaning of the notation is obtained from the context.
3 Minimax rates of estimation for sparse correlation matrices

As already mentioned before we want to study the minimax rates under the Frobenius norm such that \( \rho \) and \( \tau \) lie a fixed sparsity class \( G_{w,q}(c_{n,p}) \), i.e. we bound

\[
\inf_{\hat{\rho}} \sup_{\rho \in G_{w,q}(c_{n,p})} \frac{1}{p} E \| \hat{\rho} - \rho \|_F^2 \text{ and } \inf_{\hat{\tau}} \sup_{\tau \in G_{w,q}(c_{n,p})} \frac{1}{p} E \| \hat{\tau} - \tau \|_F^2 ,
\]

where the infimum is taken over all estimators for \( \rho \) resp. \( \tau \). In the supremum we have a slight abuse of notation, since the maximal risk

\[
\sup_{\rho \in G_{w,q}(c_{n,p})} E \| \hat{\rho} - \rho \|_F^2 \text{ resp. } \sup_{\tau \in G_{w,q}(c_{n,p})} \frac{1}{p} E \| \hat{\tau} - \tau \|_F^2 ,
\]

of a fixed estimator \( \hat{\rho} \) resp. \( \hat{\tau} \) is to read as the supremum over all permissible elliptical distributions \( EC_p(\mu, \Sigma, \xi) \) for the underlying sample \( X_1, ..., X_n \) of i.i.d. observations such that \( \rho \) resp. \( \tau \) lies in \( G_{w,q}(c_{n,p}) \). Notice that \( \rho \) and \( \tau \) do only depend on \( \Sigma \).

We first present the sharp minimax lower bounds of estimation for the correlation matrix and Kendall’s tau correlation matrix. The lower bound for estimating \( \rho \) over some class \( G_{w,q}(c_{n,p}) \) is an immediate consequence of the proof of Theorem 4 in the article Cai and Zhou (2012)(2), where the authors use a novel generalization of Le Cam’s method and Assouad’s lemma to treat the “two-directional” problem of estimating sparse covariance matrices. In the proof the authors consider the minimax lower bound over a finite subset of sparse covariance matrices, where the diagonal entries are equal to one, for a normally distributed sample. Hence, their proof implies the first part of Theorem 1.

**Theorem 1 (Minimax lower bound of estimation for sparse correlation matrices)**

Under the assumptions (A1) – (A5) the following minimax lower bounds hold

\[
\inf_{\hat{\rho}} \sup_{\rho \in G_{w,q}(c_{n,p})} \frac{1}{p} E \| \hat{\rho} - \rho \|_F^2 \geq C_{M,m,q,c_{n,p}} \left( \frac{\log p}{n} \right)^{1 - \frac{2}{q}} ,
\]

and

\[
\inf_{\hat{\tau}} \sup_{\tau \in G_{w,q}(c_{n,p})} \frac{1}{p} E \| \hat{\tau} - \tau \|_F^2 \geq \tilde{C}_{M,m,q,c_{n,p}} \left( \frac{\log p}{n} \right)^{1 - \frac{2}{q}} ,
\]

for some constants \( C_{M,m,q,c_{n,p}}, \tilde{C}_{M,m,q,c_{n,p}} > 0 \).

**Proof.** We only need to prove the second part of the Theorem and it is sufficient to restrict to estimators such that \( |\hat{\tau}_{ij}| \leq 1 \) for all \( i, j = 1, ..., p \). Therefore let \( \hat{\tau} \) be an estimator of Kendall’s tau sample correlation matrix, then \( \sin[\hat{\tau}] \) is an estimator of the correlation matrix \( \rho \). Moreover, by Lemma 2 \( \sin[\hat{\tau}] \in G_{w,q}(c_{n,p}) \) implies \( \tau \in G_{w,q}(c_{n,p}) \). We conclude by Lemma
Critical threshold levels

1 and by the first part of Theorem 1

\[
\inf_{\hat{\tau}} \sup_{\tau \in \mathcal{G}_{w,q}(c_{n,p})} \mathbb{E}[||\hat{\tau} - \tau||^2_F] \geq \frac{4}{\pi^2} \inf_{\hat{\tau}} \sup_{\tau \in \mathcal{G}_{w,q}(c_{n,p})} \mathbb{E}[||\sin[\frac{\pi}{2}\hat{\tau}] - \sin[\frac{\pi}{2}\tau]||^2_F].
\]

\[
\geq \frac{4}{\pi^2} \inf_{\hat{\rho}} \sup_{\rho \in \mathcal{G}_{w,q}(c_{n,p})} \mathbb{E}[||\sin[\frac{\pi}{2}\hat{\rho}] - \rho||^2_F].
\]

\[
\geq \frac{4}{\pi^2} \inf_{\hat{\rho}} \sup_{\rho \in \mathcal{G}_{w,q}(c_{n,p})} \mathbb{E}[||\hat{\rho} - \rho||^2_F].
\]

\[
\geq C_{c_{n,p}} \left( \frac{\log p}{n} \right)^{1 - \frac{2q}{2}}.
\]

Although Cai and Zhou (2012) do not mention the condition \(2m^q \left( \frac{\log p}{n} \right)^{q/2} < c_{n,p} \) explicitly, they require this assumption. For \(2m^q \left( \frac{\log p}{n} \right)^{q/2} > c_{n,p} \) their lower bound on the minimax risk in the proof is zero. Indeed one needs to ensure that the permissible correlation matrices are not too sparse in order that Theorem 1 provides the correct minimax rates. The following elementary proposition shows that estimating the correlation matrix by the identity matrix achieves a better rate than stated in Theorem 1 if \(c_{n,p} = o\left( \left( \frac{\log p}{n} \right)^{q/2} \right) \):

**Proposition 2**

Let the assumptions \((A_1), (A_2)\) and \((A_3)\) hold, then

\[
\sup_{\rho \in \mathcal{G}_{w,q}(c_{n,p})} \frac{1}{p} \mathbb{E}[||\hat{\rho} - \rho||^2_F] = o \left( c_{n,p} \left( \frac{\log p}{n} \right)^{1 - \frac{q}{2}} \right).
\]

(3.3)

and

\[
\sup_{\rho \in \mathcal{G}_{w,q}(c_{n,p})} \frac{1}{p} \mathbb{E}[||\hat{\tau} - \tau||^2_F] = o \left( c_{n,p} \left( \frac{\log p}{n} \right)^{1 - \frac{q}{2}} \right).
\]

(3.4)

**Proof.** Obviously both statements are equivalent. Therefore we only prove the first one. In case of \(q = 0\) the set \(\mathcal{G}_{w,q}(c_{n,p}) = \{\text{Id}\}\) for sufficiently large sample size \(n\) and hence there is nothing to prove. So assume that \(0 < q < 1\). Denote for \(j = 1, ..., p - 1\)

\[
(j)_i := \begin{cases} [j]_{i}, & \text{for } [j]_{i} < i, \\ [j]_{i} + 1 & \text{else.} \end{cases}
\]

Since the considered estimator is non-random, we have

\[
p^{-1} \mathbb{E}[||\hat{\tau} - \tau||^2_F] = p^{-1} ||\hat{\tau} - \tau||^2_F = p^{-1} \sum_{i=1}^{p} \sum_{j=1}^{p-1} |(j)_i|^2 \leq c_{n,p}^2 \left( 1 + \sum_{j=2}^{p} j^{-2/q} \right)
\]

\[
\leq c_{n,p}^2 \left( 1 + \int_{1}^{\infty} x^{-2/q} dx \right) = \left( 1 + \frac{q}{2 - q} \right) c_{n,p}^2 \left( \frac{n}{p} \right)^{2/q - 1}.
\]
The claim is now obtained from the fact that \( c_{n,p}^{2/q-1} = o\left(\left(\frac{\log p}{n}\right)^{1-\eta/2}\right) \). \( \square \)

Henceforth we only consider classes \( \mathcal{G}_{w,q}(c_{n,p}) \) of sparse correlation matrices, such that \( c_{n,p} > 2m^q \left(\frac{\log p}{n}\right)^{q/2} \). In the simplest case \( q = 0 \) this implies that there exist matrices \( A \in \mathcal{G}_{w,0}(c_{n,p}) \) such that some row has at least two nonzero entries off the diagonal. In this case the estimation of the correlation matrix and Kendall’s tau correlation matrix can be accomplished by the threshold estimators \( \hat{\rho}^\ast \) and \( \hat{\tau}^\ast \) with suitable threshold levels \( \alpha \sqrt{\frac{\log p}{n}} \). The following theorem proves that the choice \( \alpha > 2 \) is sufficient to guarantee that both estimators achieve the minimax rates. Hence, the proposed estimators provide even for small sample sizes a non-trivial estimate of the correlation matrix and Kendall’s tau correlation matrix, where by “non-trivial estimate” we understand that the estimators are with positive probability not the identity matrix.

**Theorem 2** (Minimax upper bound for Kendall’s tau correlation matrix estimation)

Under the assumptions \((A_1) - (A_5)\) the threshold estimators \( \hat{\tau}^\ast = T_\alpha(\hat{\tau}) \) and \( \hat{\rho}^\ast = \sin[\frac{\pi}{2}\hat{\tau}^\ast] \) based on Kendall’s tau sample correlation matrix \( \hat{\tau} \) attain the minimax rate over the set \( \mathcal{G}_{w,q}(c_{n,p}) \) for any threshold constant \( \alpha > 2 \), particularly

\[
\sup_{\tau \in \mathcal{G}_{w,q}(c_{n,p})} \frac{1}{n} \mathbb{E}[||\hat{\tau}^\ast - \tau||^2_F] \leq C_\alpha c_{n,p} \left(\frac{\log p}{n}\right)^{1-\eta/2},
\]

and

\[
\sup_{\rho \in \mathcal{G}_{w,q}(c_{n,p})} \frac{1}{n} \mathbb{E}[||\hat{\rho}^\ast - \rho||^2_F] \leq \tilde{C}_\alpha c_{n,p} \left(\frac{\log p}{n}\right)^{1-\eta/2},
\]

where the constants \( \tilde{C}_\alpha, C_\alpha > 0 \) depend on the threshold constant \( \alpha \).

As we will see in the next section Theorem 2 cannot be extended to any threshold constant \( \alpha < 2 \) without using an improved large deviation inequality in comparison to Hoeffding’s inequality for U-statistics (Hoeffding [1963]). Before we start to discuss the issue of critical threshold constants, we close the section with a final result. We show that it is irrelevant whether we first apply a threshold operator on Kendall’s tau sample correlation matrix and afterwards transform the obtained matrix to an appropriate estimator for the correlation matrix or vice versa. Heuristically, this seems evident, as asymptotically the ratio between the implied threshold level of \( \sin[\frac{\pi}{2} T_\alpha(\hat{\tau})] \) with respect to entries of \( \sin[\frac{\pi}{2}\hat{\tau}] \) and the threshold level of \( T_\frac{\pi}{2}(\sin[\frac{\pi}{2}\hat{\tau}]) \) is one.

**Theorem 3**

Under the assumptions \((A_1) - (A_5)\) the threshold estimator \( \hat{\rho} := T_\alpha(\sin[\frac{\pi}{2}\hat{\tau}]) \) based on Kendall’s tau sample correlation matrix \( \hat{\tau} \) attains the minimax rate over the set \( \mathcal{G}_{w,q}(c_{n,p}) \) for any \( \alpha > \pi \), particularly

\[
\sup_{\rho \in \mathcal{G}_{w,q}(c_{n,p})} \frac{1}{n} \mathbb{E}[||\hat{\rho} - \rho||^2_F] \leq C_\alpha c_{n,p} \left(\frac{\log p}{n}\right)^{1-\eta/2},
\]

where the factor \( C_\alpha > 0 \) depends on the threshold constant \( \alpha \).
4 Critical threshold levels for minimax estimation

In this section we discuss the critical threshold levels for which the threshold estimators just attain the minimax rate. In opposite to the previous section we need throughout that the observations are Gaussian random vectors and the dimension $p$ does not grow to fast in $n$. Precisely, $p$ is only allowed to grow polynomially in $n$. Otherwise we cannot apply the large deviation inequality \[ \text{(5.6)} \] from section 5. As mentioned earlier the arguments in the proof of Theorem 2 are optimized up to the concentration inequality for the entries of Kendall’s tau sample correlation matrix. This is supposed to mean that if inequality \[ \text{(5.6)} \] would hold for any entry $\hat{\tau}_{ij}$ then replacing Hoeffding’s inequality by \[ \text{(5.6)} \] at some places of the proof already provides the critical threshold level. Indeed we do not need inequality \[ \text{(5.6)} \] for all entries $\hat{\tau}_{ij}$. It is sufficient to stay with Hoeffding’s inequality for strongly correlated components. We say that two real-valued random variables $(Y, Z)$ are strongly correlated (with respect to $p$ and $n$) if $\rho(Y, Z) \geq \frac{\log p}{\sqrt{pn}}$. Otherwise $(Y, Z)$ are called weakly correlated. Certainly the constant $\frac{\log p}{\sqrt{pn}}$ is chosen arbitrarily and could be replaced by any sufficiently large value. For the proof that threshold estimators with threshold constant $\alpha < \frac{2\sqrt{p}}{3}$ do not achieve the minimax rate over some classes $G_{\alpha, \eta}$ we just have to study $E\rho^{-1}||\hat{\rho} - \tilde{\rho}||_F^2$ by the lower bound in inequality \[ \text{(4.6)} \], where the identity matrix is the underlying correlation matrix. Note that obviously $\tilde{\rho} \in G_{\alpha, \eta}$ for any sparsity class $G_{\alpha, \eta}$.

Similar to the previous section we first present the results for the estimators $\hat{\tau}^*$ and $\hat{\rho}^*$. Afterwards analogous statements for $\tilde{\rho}$ are formulated.

**Theorem 4**

Under the assumptions $(A_1^*)$ and $(A_2) - (A_6)$ let $\alpha < \frac{2\sqrt{p}}{3}$, then $\hat{\tau}^*$ and $\hat{\rho}^*$ are minimax rate optimal over all sets $G_{w,q}(c_{n,p})$ if $\alpha > \frac{2\sqrt{p}}{3}$. Hence, for $\alpha > \frac{2\sqrt{p}}{3}$ and an arbitrary set $G_{w,q}(c_{n,p})$ we have

\[
\sup_{\tau \in G_{w,q}(c_{n,p})} E[||\hat{\tau}^* - \tau||_F^2] \leq C_{\alpha, \eta_\alpha} c_{n,p} \left( \frac{\log p}{n} \right)^{1 - \frac{2}{p}} \tag{4.1}
\]

and

\[
\sup_{\rho \in G_{w,q}(c_{n,p})} E[||\hat{\rho}^* - \rho||_F^2] \leq C_{\alpha, \eta_\alpha} c_{n,p} \left( \frac{\log p}{n} \right)^{1 - \frac{2}{p}} \tag{4.2}
\]

where the constants $C_{\alpha, \eta_\alpha}, \tilde{C}_{\alpha, \eta_\alpha} > 0$ depend on $\alpha$ and $\eta$.

Moreover, for $\alpha < \frac{4}{3}$ there is no permissible set $G_{w,q}(c_{n,p})$ such that $\hat{\tau}^*$ or $\hat{\rho}^*$ attains the minimax rate over $G_{w,q}(c_{n,p})$.

**Theorem 5**

Under the assumptions $(A_1^*)$ and $(A_2) - (A_6)$ let $\alpha < \frac{2\sqrt{p}}{3}$, then $\tilde{\rho}$ is minimax rate optimal over all sets $G_{w,q}(c_{n,p})$ if $\alpha > \frac{2\sqrt{p}}{3}$. Hence, for $\alpha > \frac{2\sqrt{p}}{3}$ and an arbitrary set $G_{w,q}(c_{n,p})$ we
have
\[
\sup_{\rho \in \mathcal{G}_{w,q}(c_{n,p})} \left\| \frac{1}{p} \mathbb{E}[\hat{\rho} - \rho]^2 \right\|_F \leq C_{\alpha,\eta} c_{n,p} \left( \frac{\log p}{n} \right)^{1 - \frac{\alpha}{2}} ,
\]
(4.3)
where the constant \( C_{\alpha,\eta} > 0 \) depends on \( \alpha \) and \( \eta \).
Moreover, for \( \alpha < \frac{\pi}{3} \) there is no permissible set \( \mathcal{G}_{w,q}(c_{n,p}) \) such that \( \hat{\rho} \) attains the minimax rate over \( \mathcal{G}_{w,q}(c_{n,p}) \).

5 On Kendall’s tau sample correlation for normal distributions with weak correlation

In this section we discuss the properties of the tails of Kendall’s tau sample correlation based on a sample \( (Y_1, Z_1), \ldots, (Y_n, Z_n) \sim \mathcal{N}_2(\mu, \Sigma) \) for \( \Sigma = \text{Id} \) and small perturbations of the identity. Specifically, we need preferably sharp upper and lower bounds on its tails. The essential argument for our investigation is the natural linkage between Kendall’s tau sample correlation and the number of inversions in a random permutation. So we can apply “an asymptotic expansion for the number of permutations with a certain number of inversions” developed by Clark [2000].

5.1 Kendall’s tau sample correlation for the standard normal distribution

Before studying the tails of Kendall’s tau sample correlation \( \hat{\tau}(Y_1, Z_1, \ldots, Y_n, Z_n) \) based on a sample \( (Y_1, Z_1), \ldots, (Y_n, Z_n) \sim \mathcal{N}_2(\mu, \Sigma) \) for \( \Sigma = \text{Id} \) and small perturbations of the identity, we first prove that \( \hat{\tau}(Y_1, Z_1, \ldots, Y_n, Z_n) \) has - after centering and rescaling - the same distribution as the number of inversions in a random permutation on \( \{1, \ldots, n\} \). This result is probably known for long time but to the best of the author’s knowledge in no work mentioned explicitly. Particularly, so far statisticians have not taken advantage of any developments on inversions in random permutations.

The number of inversions in a permutation is an old and well-studied object. We say that a permutation \( \pi \) on \( \{1, \ldots, n\} \) has an inversion at \( (i, j) \), \( 1 \leq i < j \leq n \), iff \( \pi(j) > \pi(i) \). This concept was originally introduced by Cramer [1750] in the context of the Leibniz formula for the determinant of quadratic matrices. Denote by \( I_n(k) \) the number of permutations on \( \{1, \ldots, n\} \) with exactly \( k \) inversions. The generating function \( G \) for the numbers \( I_n(k) \) is given by \( G(z) = \prod_{l=1}^{n-1} (1 + z + \ldots + z^l) \) as already known at least since Muir [1900]. Note that Kendall [1938] studied the generating function of \( \hat{\tau}(Y_1, Z_1, \ldots, Y_n, Z_n) \) independently on prior works on inversions in permutations and thereby derived a central limit theorem for \( \hat{\tau}(Y_1, Z_1, \ldots, Y_n, Z_n) \) when \( n \) tends to infinity. However such a result is not strong enough for our purposes. We actually need a Gaussian approximation for the tails of order \( (\log n)^{1/2} \). This will be concluded by the work of Clark [2000], who gives an asymptotic expansion for \( I_n(k) \), where \( k = \frac{n(n-1)}{2} \pm l \) and \( l \) is allowed to grow moderately with \( n \). Therefore, at this point we need that \( p \) is not increasing faster than polynomially in \( n \), in other words \( \frac{\log n}{n} \) is bounded above by some absolute constant.
Certainly, one could show the connection between Kendall’s tau correlation and the number of inversion in random permutations by their generating functions. We prefer a direct proof which is more intuitive.
Proposition 3
Let \( I_- \) be the number of inversions in a random permutation on \( \{1, \ldots, n\} \), \( n \geq 2 \), and \( \hat{\tau}(Y_{1:n}, Z_{1:n}) \) be Kendall’s sample correlation based on \((Y_1, Z_1), \ldots, (Y_n, Z_n) \overset{i.i.d.}{\sim} \mathcal{N}_2(\mu, \text{Id}) \).
Then, it holds
\[
\hat{\tau}(Y_{1:n}, Z_{1:n}) \overset{p}{=} 1 - \frac{4}{n(n - 1)} I_-.
\]

Proof. Recall that by definition
\[
\hat{\tau}(Y_{1:n}, Z_{1:n}) = \frac{1}{n(n - 1)} \sum_{i,j} \text{sign}(Y_i - Y_j) \text{sign}(Z_i - Z_j).
\]

Now let \( \pi: i \mapsto [i] \) be the permutation induced by the order statistics \( Z_{[1]} \geq \ldots \geq Z_{[n]} \). Therefore, rewrite
\[
\hat{\tau}(Y_{1:n}, Z_{1:n}) = \frac{2}{n(n - 1)} \sum_{[i] \neq [j]} \text{sign}(Y_{[i]} - Y_{[j]}) \text{sign}(Z_{[i]} - Z_{[j]})
\] (a.s.)
\[
= \frac{2}{n(n - 1)} \sum_{i > j} \text{sign}(Y_i - Y_j) (\text{by independence of } Y \text{ and } Z)
\]
Remap \( \pi: i \mapsto [i] \) by the permutation induced by the order statistics \( Y_{[1]} \geq \ldots \geq Y_{[n]} \). Obviously, \( \text{sign}(X_i - X_j) = -1 \) if \( \pi \) has an inversion at \((i, j)\). Otherwise \( \text{sign}(X_i - X_j) = 1 \). Denote by \( I_- \) the number of inversions in \( \pi \) and by \( I_+ \) the number of all the other pairs \((i, j)\). Clearly, \( I_- + I_+ = \binom{n}{2} \).
Finally, we conclude
\[
\hat{\tau}(Y_{1:n}, Z_{1:n}) \overset{p}{=} \frac{2}{n(n - 1)} \sum_{i > j} \text{sign}(Y_i - Y_j)
\] (a.s.)
\[
= \frac{2}{n(n - 1)} (I_+ - I_-)
\] (by independence of \( Y \) and \( Z \))
\[
= 1 - \frac{4}{n(n - 1)} I_-.
\]

Now we reformulate the result of Clark (2000) for the number of permutations whose number of inversions differ exactly by \( l \) from \( \frac{n(n - 1)}{4} \).

Theorem 6 (Clark (2000))
Fix \( \lambda > 0 \). Let \( m = \lfloor \lambda^2/2 \rfloor + 2 \) and \( l \in \text{Im}(I_- - EI_-) \), then we have
\[
P(|I_- - EI_-| = l) = 12(2\pi)^{-1/2} n^{-3/2} e^{-\lambda^2/4n} + r_{n,\lambda,1}(l) + r_{n,\lambda,2}(l),
\] (5.1)
where the error terms $r_{n,\gamma,1}$ and $r_{n,\gamma,2}$ satisfy for a certain constant $C_\gamma > 0$

\[
|r_{n,\lambda,1}(l)| \leq C_{\lambda} \left( \frac{n^{-5/2} e^{-1/2 n^{-2}}} n^{6m + 6 4m^{-4}} \right)
\]

and

\[
|r_{n,\lambda,2}(l)| \leq C_{\lambda} \left( \frac{\log^{2m^2 + 1} n}{n^{m + 3/2}} \right).
\]

Notice that for the second error term $r_{n,\gamma,2}$ we have a uniform upper bound for all $l \in \text{Im}(I_\tau - E_\tau)$. Obviously if

\[
36 n^{-3/4} \leq \gamma^2 \log n,
\]

the leading term on the right hand side of inequality (5.1) is the dominating one. Now Theorem 6 enables to calculate asymptotically sharp bounds on the tail probabilities of $\hat{\tau}(X_1:n, Y_1:n)$.

**Proposition 4**

Under the assumptions of Proposition 3 let $\gamma, \beta > 0$ and $p \in \mathbb{N}$, such that $p < n^\beta$. Then,

\[
P \left( |\hat{\tau}(Y_1:n, Z_1:n)| \geq \gamma \left( \frac{\log p}{n} \right) \right) = 2 \left( 1 - \Phi \left( \frac{3}{2} \gamma \sqrt{\log p} \right) \right) + R_{n,p,\beta,\gamma},
\]

where the error term $R_{n,p,\beta,\gamma}$ satisfies for $m = \left[ \frac{3}{2} + \beta \gamma^2 \right] + 1$ and some constant $C_{\gamma,\beta} > 0$

\[
|R_{n,p,\beta,\gamma}| \leq C_{\gamma,\beta} \log^{2m^2 + 1} n \frac{n^{-\beta}}{p \gamma^2}. \quad (5.3)
\]

**Proof.** Let $\hat{\tau} := \hat{\tau}(Y_1:n, Z_1:n)$, $I_0 := 6n^{-3/2} (I_\tau - E_\tau)$, $\gamma_{n,p} := \frac{3}{2} \gamma \sqrt{\log p} - \frac{3}{2} \gamma \sqrt{\log p}$ and $\lambda > 0$ be first arbitrary, then we have

\[
P \left( |\hat{\tau}| \geq \gamma \left( \frac{\log p}{n} \right) \right) = P \left( |I_\tau - E_\tau| \geq \frac{1}{2} \gamma n^{3/2} \sqrt{\log p} - \frac{1}{4} n^{1/2} \sqrt{\log p} \right)
\]

\[
= P \left( |I_0| \geq \frac{3}{2} \gamma \sqrt{\log p} - \frac{3}{2} \gamma \sqrt{\log p} \right)
\]

\[
= \sum_{x \in \text{Im}(I_0): x \geq \gamma_{n,p}} \sqrt{\frac{2 e^{-x^2}}{x^2 \gamma_{n,p}}} + \sum_{x \in \text{Im}(I_0): x \geq \gamma_{n,p}} r_{n,\lambda,1} \left( \frac{n^{3/2} \log p}{6} \right) + \sum_{x \in \text{Im}(I_0): x \geq \gamma_{n,p}} r_{n,\lambda,2} \left( \frac{n^{3/2} \log p}{6} \right)
\]

\[
= J_1 + J_2 + J_3.
\]

We evaluate $J_1$, $J_2$ and $J_3$ separately. We first give upper bounds on the expressions.
Upper bound on $J_1$:

\[
J_1 \leq 2 \int_{\gamma_n - 6n^{-3/2}}^{\infty} \phi(x)dx = 2 \left( 1 - \Phi \left( \frac{3}{2} \gamma \sqrt{\log p} \right) \right) + 2 \int_{\gamma_n - 6n^{-3/2}}^{\infty} \phi(x)dx \\
\leq 2 \left( 1 - \Phi \left( \frac{3}{2} \gamma \sqrt{\log p} \right) \right) + C_{\gamma, \beta} \frac{\log p}{n} \phi \left( \frac{3}{2} \gamma \sqrt{\log p} \right) - \frac{3}{2} \gamma \sqrt{\log p} - 6n^{-3/2} \\
\leq 2 \left( 1 - \Phi \left( \frac{3}{2} \gamma \sqrt{\log p} \right) \right) + C_{\gamma, \beta} \frac{\log p}{n} \phi \left( \frac{3}{2} \gamma \sqrt{\log p} \right) \\
= 2 \left( 1 - \Phi \left( \frac{3}{2} \gamma \sqrt{\log p} \right) \right) + C_{\gamma, \beta} \frac{\log p}{n} p^{-\frac{3}{2} \gamma^2}.
\]

Upper bound on $J_2$: By the error bound in Theorem 6 on $r_{n, \lambda, 1}$ and integration by parts we similarly conclude

\[
J_2 \leq C_{\gamma, \beta, m} \log \frac{4}{m^2} - \frac{3}{2} \gamma \sqrt{\log p} - 9 \gamma^2.
\]

Lower bound on $J_1$: Clearly,

\[
J_1 \geq C_m \log^{m-1} \frac{\log p}{n^{m-\frac{2}{3}}}. 
\]

For an integer $m > 3/2 + \frac{9}{16} \beta \gamma^2$ we finally have

\[
P \left( |\hat{\tau}| \geq \gamma \sqrt{\frac{\log p}{n}} \right) \geq 2 \left( 1 - \Phi \left( \frac{3}{2} \gamma \sqrt{\log p} \right) \right) - C_{\gamma, \beta} \log^{m-1} \frac{\log \frac{4}{m^2}}{n^{m-\frac{2}{3}}} p^{-\frac{3}{2} \gamma^2}. 
\]

Lower bound on $J_1$: Analogously to the upper bound on $J_1$ we obtain

\[
J_1 \geq 2 \left( 1 - \Phi \left( \frac{3}{2} \gamma \sqrt{\log p} \right) \right) \int_{\gamma_n - 6n^{-3/2}}^{\infty} \phi(x)dx.
\]

Lower bounds on $J_2$ and $J_3$: We have

\[
J_2 \geq -C_{\gamma, \beta, m} \log^{m-1} \frac{\log \frac{4}{m^2}}{n} p^{-\frac{3}{2} \gamma^2} 
\]

and

\[
J_3 \geq C_m \log^{m-1} \frac{\log \frac{4}{m^2}}{n^{m-\frac{2}{3}}}. 
\]

Hence, again pick $m > 3/2 + \frac{9}{16} \beta \gamma^2$ and derive

\[
P \left( |\hat{\tau}| \geq \gamma \sqrt{\frac{\log p}{n}} \right) \geq 2 \left( 1 - \Phi \left( \frac{3}{2} \gamma \sqrt{\log p} \right) \right) - C_{\gamma, \beta} \log^{m-1} \frac{\log \frac{4}{m^2}}{n} p^{-\frac{3}{2} \gamma^2}. 
\]

Combing both bounds provides the desired statement.
5.2 Tails of Kendall’s tau sample correlation for normal distributions with weak correlation

In this subsection we transfer Proposition 4 to normal distributed random variables \( (Y, Z) \) with weak correlation \( \sigma \) and give some conclusions from it. The crucial argument to evaluate the tail probabilities of \( \hat{\tau}(Y_{1:n}, Z_{1:n}) - \tau(Y, Z) \) for an i.i.d. sample \( (Y_1, Z_1), \ldots, (Y_n, Z_n) \) is to approximate \( \hat{\tau}(Y_{1:n}, Z_{1:n}) - \tau(Y, Z) \) by Kendall’s tau sample correlation for an appropriate sample \( (W_1, Z_1), \ldots, (W_n, Z_n) \) with uncorrelated components. For the ease of notations suppose that \( Y \) and \( Z \) are standardized. Then \( Y \) can be written as \( Y = \sqrt{1 - \sigma^2} W + \sigma Z \) for \( (W, Z) \sim \mathcal{N}_2(\mu, \text{Id}) \). This will be the natural candidate for the approximation argument.

**Lemma 3**

Let \( (W_1, Z_1), \ldots, (W_n, Z_n) \overset{i.i.d.}{\sim} \mathcal{N}_2(\mu, \text{Id}) \), \( Y_i = \sqrt{1 - \sigma^2} W_i + \sigma Z_i, \ i = 1, \ldots, n \), where \( \sigma^2 \leq \zeta \frac{\log n}{n} \lambda \), \( 1 < n < p \leq n^\beta \) for a constant \( \beta > 1 \) and \( \zeta > 0 \). Then for \( c_n = \lambda \frac{\log n}{n^{1/4}}, \lambda > 0 \), holds

\[
P \left( |\hat{\tau}(Y_{1:n}, Z_{1:n}) - \tau(Y, Z_1) - \hat{\tau}(W_{1:n}, Z_{1:n})| \geq c_n \sqrt{\frac{\log p}{n}} \right) \leq 2np^{-c_\beta, \zeta, \lambda, \sqrt{\log p}}.
\]

**Proof.** By 1-factorization of the complete graph on \( n + 2 \left( \frac{n}{2} - \left[ \frac{n}{2} \right] \right) \) vertices and by the union bound we conclude

\[
P \left( |\hat{\tau}(Y_{1:n}, Z_{1:n}) - \tau(Y, Z_1) - \hat{\tau}(W_{1:n}, Z_{1:n})| \geq c_n \sqrt{\frac{\log p}{n}} \right)
= \mathbb{P} \left( \left| \frac{2}{n - 1} \sum_{k, l = 1}^{n} \left( (\text{sign}(Y_k - Y_l) - \text{sign}(W_k - W_l)) \text{sign}(Z_k - Z_l) - \tau(Y_1, Z_1) \right) \right| \geq c_n \sqrt{n \log p} \right)
\leq n \mathbb{P} \left( \left| \frac{2}{n - 1} \sum_{i = 1}^{\left[ \frac{n}{2} \right]} \left( (\text{sign}(Y_{2k - 1} - Y_{2k}) - \text{sign}(W_{2k - 1} - W_{2k})) \times \text{sign}(Z_{2k - 1} - Z_{2k}) - \tau(Y_1, Z_1) \right) \right| \geq c_n \sqrt{\frac{\log p}{n}} \right).
\]

Now let

\[
\varepsilon_k := \frac{1}{3} \left( (\text{sign}(Y_{2k-1} - Y_{2k}) - \text{sign}(W_{2k-1} - W_{2k})) \text{sign}(Z_{2k-1} - Z_{2k}) - \tau(X, Y) \right).
\]

Obviously, the random variables \( \varepsilon_k \) are centered and bounded in absolute value by 1. We
evaluate the variance of $\varepsilon_k$ to apply Bernstein inequality. We have

$$\text{Var} (\varepsilon_k) = E \varepsilon_k^2 \leq \frac{1}{9} E (\text{sign} (Y_{2k-1} - Y_{2k}) - \text{sign} (W_{2k-1} - W_{2k}))^2$$

$$\leq P(\text{sign}(Y_{2k-1} - Y_{2k}) \neq \text{sign}(W_{2k-1} - W_{2k}))$$

$$= P\left( \text{sign} \left( \sqrt{1-\sigma^2} (W_{2k-1} - W_{2k}) + \sigma (Z_{2k-1} - Z_{2k}) \right) \neq \text{sign} (W_{2k-1} - W_{2k}) \right)$$

$$\leq P \left( \sqrt{1-\sigma^2} |W_{2k-1} - W_{2k}| < |\sigma| |Z_{2k-1} - Z_{2k}| \right)$$

$$\leq P \left( \frac{|W_{2k-1} - W_{2k}|}{|Z_{2k-1} - Z_{2k}|} < 2|\sigma| \right)$$

where the last line follows easily from the fact that $\frac{W_{2k-1} - W_{2k}}{Z_{2k-1} - Z_{2k}}$ is standard Cauchy distributed and therefore its density is bounded by $\pi^{-1}$. Finally, we conclude by Bernstein inequality

$$P \left( |\hat{\tau}(Y_{1,n}, Z_{1,n}) - \tau(Y_{1}, Z_{1}) - \hat{\tau}(W_{1,n}, Z_{1,n})| \geq c_n \sqrt{\frac{\log p}{n}} \right) \leq nP \left( \sum_{l=1}^{n} \varepsilon_l \geq c_n \frac{n-1}{6n} \sqrt{n \log p} \right)$$

$$\leq 2np^{-C_{\beta, \zeta, \lambda} \sqrt{\log p}}.$$

\[\square\]

**Proposition 5**

Let $(Y_1, Z_1), \ldots, (Y_n, Z_n) \overset{i.i.d.}{\sim} \mathcal{N}_2(\mu, \Sigma), \Sigma_{11}, \Sigma_{22} > 0$, where $\frac{\Sigma_{12}}{\sqrt{\Sigma_{11} \Sigma_{22}}} \leq \zeta \sqrt{\frac{\log p}{n}} \wedge \frac{3}{2}$ for an arbitrary constant $\zeta > 0$ and $n < p \leq n^\beta$, $\beta > 1$. Then for any real number $\gamma > 0$ holds

$$P \left( |\hat{\tau}(Y_{1,n}, Z_{1,n}) - \tau(Y_{1}, Z_{1})| \geq \gamma \sqrt{\frac{\log p}{n}} \right) = 2 \left( 1 - \Phi \left( \frac{3}{2} \gamma \sqrt{\frac{\log p}{n}} \right) \right) + R_{n,p,\beta,\zeta,\gamma}, \quad (5.4)$$

where the error $R_{n,p,\beta,\zeta,\gamma}$ satisfies for some constant $C_{\beta, \zeta, \gamma} > 0$

$$|R_{n,p,\beta,\zeta,\gamma}| \leq C_{\beta, \zeta, \gamma} \frac{\log p}{n^{1/4} p^{2/3}}. \quad (5.5)$$

**Proof.** Let $\sigma := \frac{\Sigma_{12}}{\sqrt{\Sigma_{11} \Sigma_{22}}}$. W.l.o.g. assume that $(W_1, Z_1), \ldots, (W_n, Z_n) \overset{i.i.d.}{\sim} \mathcal{N}_2(0, \text{Id})$ and $Y_i = \sqrt{1-\sigma^2} W_i + \sigma Z_i$, $i = 1, \ldots, n$. Pick $c_n := 12 \frac{\sqrt{\log p} \sqrt{\log p}}{(n-1)n^{1/4}}$. First we give an upper bound
on \( P \left( |\hat{\tau}(Y_{1:n}, Z_{1:n}) - \tau(Y_1, Z_1)| \geq \gamma \sqrt{\frac{\log p}{n}} \right) \). We have

\[
P \left( |\hat{\tau}(Y_{1:n}, Z_{1:n}) - \tau(Y_1, Z_1)| \geq \gamma \sqrt{\frac{\log p}{n}} \right)
\leq P \left( |\hat{\tau}(Y_{1:n}, Z_{1:n}) - \hat{\tau}(Y_{1:n}, Z_{1:n})| + |\hat{\tau}(Y_{1:n}, Z_{1:n}) - \tau(Y_1, Z_1)| \right) \geq \gamma \sqrt{\frac{\log p}{n}}
\leq P \left( |\hat{\tau}(W_{1:n}, Z_{1:n})| \geq (\gamma - c_n) \sqrt{\frac{\log p}{n}} \right)
+ P \left( |\hat{\tau}(Y_{1:n}, Z_{1:n}) - \tau(Y_1, Z_1) - \hat{\tau}(W_{1:n}, Z_{1:n})| \geq \tau(Y_1, Z_1) \right) \geq \gamma \sqrt{\frac{\log p}{n}}
\].

The second summand is easily handled by Lemma 3. For the first summand notice that \( \gamma - c_n \) is bounded above uniformly for all \( n \). Therefore we apply equation (5.2), where the constant in the error bound (5.3) can be chosen independently from \( n \), such that

\[
|\hat{R}_{m,p,\beta,\gamma-c_n}| \leq C_{\gamma,\beta,\zeta} \log^{2m^2+1} n \frac{p^{\frac{5}{4}(\gamma-c_n)^2}}{n}
\]

Hence, equation (5.2) yields

\[
P \left( |\hat{\tau}(W_{1:n}, Z_{1:n})| \geq (\gamma - c_n) \sqrt{\frac{\log p}{n}} \right) \leq 2 \left( 1 - \Phi \left( \frac{3}{2} (\gamma - c_n) \sqrt{\log p} \right) \right) + C_{\gamma,\beta,\zeta} \log^{2m^2+1} n \frac{p^{\frac{5}{4}(\gamma-c_n)^2}}{n}
\leq 2 \left( 1 - \Phi \left( \frac{3}{2} \sqrt{\log p} \right) \right) + C_{\gamma,\beta,\zeta} \frac{\log p}{n^{1/4}} p^{-\frac{5}{8} \gamma^2}.
\]

This provides the upper bound. The lower bound arises from the following computation, where we finally apply Proposition 4 and Lemma 3 again:

\[
P \left( |\hat{\tau}(Y_{1:n}, Z_{1:n}) - \tau(Y_1, Z_1)| \geq \gamma \sqrt{\frac{\log p}{n}} \right)
\geq P \left( |\hat{\tau}(W_{1:n}, Z_{1:n})| - |\hat{\tau}(W_{1:n}, Z_{1:n}) - \hat{\tau}(Y_{1:n}, Z_{1:n}) + \tau(Y_1, Z_1)| \geq \gamma \sqrt{\frac{\log p}{n}} \right)
\geq P \left( |\hat{\tau}(W_{1:n}, Z_{1:n})| - |\hat{\tau}(W_{1:n}, Z_{1:n}) - \tau(Y_{1:n}, Z_{1:n}) + \tau(Y_1, Z_1)| \geq \gamma \sqrt{\frac{\log p}{n}},
|\hat{\tau}(W_{1:n}, Z_{1:n})| - \tau(Y_{1:n}, Z_{1:n}) + \tau(Y_1, Z_1)| < c_n \sqrt{\frac{\log p}{n}} \right)
\geq P \left( |\hat{\tau}(W_{1:n}, Z_{1:n})| \geq (\gamma + c_n) \sqrt{\frac{\log p}{n}} \right)
\]
Critical threshold levels

- \[ P \left( |\hat{\tau}(W_{1:n}, Z_{1:n}) - \hat{\tau}(Y_{1:n}, Z_{1:n}) + \tau(Y_1, Z_1) | \geq c_n \sqrt{\frac{\log p}{n}} \right) . \]

We close this section with two straightforward consequences from the last proposition.

**Corollary 1**

Under the assumptions of Proposition \( \Box \) let \( \gamma > 0 \). Then,

\[
\tilde{C}_{\beta, \zeta, \gamma} \Phi \left( \frac{3}{2} \gamma \sqrt{\log p} \right) \leq P \left( |\hat{\tau}(Y_{1:n}, Z_{1:n}) - \tau(Y_1, Z_1)| \geq \gamma \sqrt{\frac{\log p}{n}} \right) \leq C_{\beta, \zeta, \gamma} \Phi \left( \frac{3}{2} \gamma \sqrt{\log p} \right),
\]

where the constants \( \tilde{C}_{\beta, \zeta, \gamma}, C_{\beta, \zeta, \gamma} > 0 \) depend on \( \beta, \zeta \) and \( \gamma \).

In the final corollary the sample \( (Y_1, Z_1), \ldots, (Y_n, Z_n) \) and the quantities \( \mu, \Sigma \) depend on \( n \) even if it is not apparent from the notation.

**Corollary 2**

Let \( (Y_1, Z_1), \ldots, (Y_n, Z_n) \overset{i.i.d.}{\sim} N_2(\mu, \Sigma), \Sigma_{11}, \Sigma_{22} > 0, \) where \( \left| \frac{\Sigma_{12}}{\sqrt{\Sigma_{11} \Sigma_{22}}} \right| \leq \zeta \sqrt{\frac{\log p}{n}} \wedge \frac{3}{4} \) for an arbitrary constant \( \zeta > 0 \) and \( n < p \leq n^\beta, \beta > 1 \). Then for any real number \( \gamma > 0 \) holds

\[
\lim_{n \to \infty} \frac{P \left( |\hat{\tau}(Y_{1:n}, Z_{1:n}) - \tau(Y_1, Z_1)| \geq \gamma \sqrt{\frac{\log p}{n}} \right)}{2 \left( 1 - \Phi \left( \frac{3}{2} \gamma \sqrt{\log p} \right) \right)} = 1.
\]

**6 Discussion**

In this article we studied the question how much information an entrywise hard threshold matrix operator \( T_{\alpha,n} \) is allowed to keep from the pilot estimator \( \sin(\hat{\pi} \hat{\tau}) \) to obtain a minimax optimal estimator for the correlation matrix. It is shown that \( \sqrt{\frac{\pi}{2} \frac{1}{\log n}} \) is a critical threshold level for Gaussian observations. This means that any threshold constant \( \alpha > \alpha^* := \sqrt{\frac{\pi}{2}} \) provides a minimax optimal estimator whereas for \( \alpha < \sqrt{\frac{\pi}{2}} \) the threshold estimator \( \hat{\rho} \) does not achieve the optimal rate over sparsity classes \( \mathcal{G}_{w,q}(c_n, p) \) without sufficiently dense correlation matrices. It is not clear how to prove analogous statements for broader classes of elliptical distributions since even the asymptotic variance of the entries \( \sin(\hat{\pi} \hat{\tau}_{ij}) \) does not only depend on \( \Sigma \) but on \( \xi \) as well - see Lehmann and Casella (1998).

Threshold estimators with small threshold constants are meaningful because of two essential reasons. On the one hand the smaller the threshold constant is the more dependency structure is captured by the estimator, on the other hand small threshold constants are of practical interest. To discuss the second reason let us suppose that we have a threshold level \( 10(\frac{\log p}{n})^{1/2} \) for instance. Then we need already a sample size larger than hundred just to compensate the
threshold constant 10, where we have not even considered the dimension $p$ which is typically at least in the thousands. Thus, in practice already moderate threshold constants may lead to a trivial estimator which provides the identity as an estimate for the correlation matrix, no matter what we actually observe.

It is an open question, which rate the threshold estimator attains for $\alpha = \alpha^*$, since Lemma 6 is not applicable for that threshold constant. This case is also important because the constant $C_{\alpha, \eta}$ in the upper bound on the maximal risk of $\hat{\rho}_\alpha$ tends to infinity as $\alpha \downarrow \alpha^*$ in the proof of Theorem 5. So if $\hat{\rho}_\alpha^*$ attains the minimax rate, the constant $C_{\alpha, \eta}$ in Theorem 5 should be substantially improvable. The results are so far restricted to minimax estimation under the Frobenius norm loss. The author supposes that $\frac{\sqrt{2 \pi}}{3}$ is also the critical threshold level under any operator norm induced by a $l_q$-norm for $1 \leq q \leq \infty$. By few adjustments of the proofs in this article we can show that under the spectral norm the critical threshold level lies within $[\sqrt{\frac{2\pi}{3}}, \frac{2\pi}{3}]$. For the exact critical threshold constant under the spectral and any other operator norm one needs an appropriate upper bound on the expectation of the squared $l_1$-norm of the adjacency matrix $\hat{M} = (\hat{M}_{ij})_{i,j=1,...,p}$ with

$$K_{ij} := \mathbf{1}\left( |\hat{\tau}_{ij}^* - \tau_{ij}| > \beta \min \left( \tau_{ij}, \alpha \sqrt{\log \frac{p}{n}} \right) \right)$$

for a sufficiently large value $\beta > 0$ and $\tau \in \mathcal{G}_{w,q}(c_{n,p})$. However, a solution to this task seems currently out of reach.

The results in this paper can easily be transferred to the estimation of sparse latent generalized correlation matrices in nonparanormal distributions (Liu, Lafferty, and Wasserman 2009, Liu et al. 2012, Xue and Zou 2012, Han, Liu, and Zhao 2013). Moreover the results of section 3 hold for the estimation of sparse latent generalized correlation matrices of meta-elliptical distributions (Fang, Fang, and Kotz 2002) and transelliptical distributions (Han and Liu 2014) as well as for the estimation of sparse copula correlation matrices for elliptical copulas (Wegkamp and Zhao 2013).

A further open problem is the identification of the threshold estimator $\hat{\rho}_\alpha$ with the asymptotically smallest maximal risk. At least under the Frobenius norm Corollary 2 should enable to compute the exact asymptotic constant of the maximal risk for any minimax optimal estimator $\hat{\rho} = \rho_\alpha$ under slight regularization, i.e. we have to evaluate the limit

$$\limsup_{n \to \infty} \sup_{\rho \in \mathcal{G}_{w,q}(c_{n,p})} c_{n,p}^{-1} \left( \frac{n}{\log p} \right)^{1-q/2} p^{-1} E ||\rho - \hat{\rho}||_F^2.$$  

Minimizing over all $\alpha$ would provide the threshold estimator with the asymptotically smallest maximal risk.

7 Proofs of Section 3 and 4

We start with the proof of Theorem 4. Therefore we need the following lemma, which is an advancement of Lemma 8 of Cai and Zhou (2012) for our purpose of minimizing the necessary threshold in Theorem 4.
Lemma 4
Let \((Y_1, Z_1), \ldots, (Y_n, Z_n) \in \mathbb{R}^2\) be i.i.d. random variables and for any \(\alpha > 2\) and \(\beta > \frac{2\alpha^2}{(\alpha + 1)^2}\) let the event
\[
B := \left\{ |\hat{\tau}^*(Y_{1:n}, Z_{1:n}) - \tau(Y_1, Z_1)| \leq \beta \min\left( |\tau(Y_1, Z_1)|, \alpha \sqrt{\frac{\log p}{n}} \right) \right\}.
\]
Then we have
\[
P(B) \geq 1 - 4p^{-(1 + \varepsilon)},
\]
where \(\varepsilon := \frac{\alpha^2 - 1}{4} - \frac{\alpha^2 (\beta - 1)^2 - (\beta + 1)^2}{4(\beta + 1)^2} > 0\).

Proof. Let \(\hat{\tau} := \hat{\tau}(Y_{1:n}, Z_{1:n}), \hat{\tau}^* := \hat{\tau}^*(Y_1, Z_1)\) and \(\tau := \tau(Y_1, Z_1)\). First note that by Hoeffding’s inequality for U-statistics (Hoeffding (1963))
\[
P(|\hat{\tau} - \tau| > t) \leq 2 \exp \left( -\frac{nt^2}{4} \right)
\]
for any \(t > 0\). (7.1)
We distinguish three cases:

(i) \(|\tau| \leq \frac{2\alpha}{\beta + 1} \sqrt{\frac{\log p}{n}}\): Since on the event \(A := \left\{ |\hat{\tau} - \alpha \sqrt{\frac{\log p}{n}}| \right\}\) holds \(|\hat{\tau}^* - \tau| = |\tau| \leq \beta \min\left( |\tau|, \alpha \sqrt{\frac{\log p}{n}} \right)\), we have \(A \subset B\). Now by the triangle inequality and inequality (7.1) follows
\[
P(B) \geq P(A) \geq P\left( |\hat{\tau} - \tau| < \frac{\beta - 1}{\beta + 1} \sqrt{\frac{\log p}{n}} \right)
\]
\[
\geq 1 - 2p^{\frac{(\beta - 1)^2\alpha^2}{(\beta + 1)^2}}.
\]

(ii) \(\frac{2\alpha}{\beta + 1} \sqrt{\frac{\log p}{n}} < |\tau| < \frac{2\alpha\beta}{\beta + 1} \sqrt{\frac{\log p}{n}}\): On \(A\) we have again \(|\hat{\tau}^* - \tau| = |\tau| \leq \beta \min\left( |\tau|, \alpha \sqrt{\frac{\log p}{n}} \right)\).
This implies \(A \subset B\). Furthermore consider the event \(C := A^c \cap \left\{ |\hat{\tau} - \tau| \leq \frac{2\alpha\beta}{\beta + 1} \sqrt{\frac{\log p}{n}} \right\}\).

On \(C\) holds \(|\hat{\tau}^* - \tau| = |\hat{\tau} - \tau| \leq \frac{2\alpha\beta}{\beta + 1} \sqrt{\frac{\log p}{n}} \leq \beta \min\left( |\tau|, \alpha \sqrt{\frac{\log p}{n}} \right)\). So \(C \subset B\). Finally inequality (7.1) yields
\[
P(B) \geq P\left( \left\{ |\hat{\tau} - \tau| \leq \frac{2\alpha\beta}{\beta + 1} \sqrt{\frac{\log p}{n}} \right\} \cap A^c \right) \cup A
\]
\[
\geq P\left( |\hat{\tau} - \tau| \leq \frac{2\alpha\beta}{\beta + 1} \sqrt{\frac{\log p}{n}} \right)
\]
\[
\geq 1 - 2p^{\frac{\alpha^2 \beta^2}{(\beta + 1)^2}}.
\]
(iii) $|\tau| > \frac{2\alpha\beta}{\beta + 1} \sqrt{\frac{\log p}{n}}$: The union bound, the triangle inequality and inequality (7.1) yield

$$P(B) = P\left( |\hat{\tau} - \tau| \leq \alpha\beta \sqrt{\frac{\log p}{n}} \right) \geq P\left( \left| \hat{\tau} - \tau \right| \leq \alpha\beta \sqrt{\frac{\log p}{n}} \right) \cap A'$$

$$\geq 1 - P\left( |\hat{\tau} - \tau| > \alpha\beta \sqrt{\frac{\log p}{n}} \right) - P(A)$$

$$\geq 1 - P\left( |\hat{\tau} - \tau| > \alpha\beta \sqrt{\frac{\log p}{n}} \right) - P\left( |\hat{\tau} - \tau| > \frac{\beta - 1}{\beta + 1} \alpha \sqrt{\frac{\log p}{n}} \right)$$

$$\geq 1 - 2p^{-\frac{\alpha^2\beta^2}{4}} - 2p^{-\frac{(\beta - 1)^2\alpha^2}{4(\beta + 1)^2}} \geq 1 - 4p^{-\frac{(\beta - 1)^2\alpha^2}{4(\beta + 1)^2}}$$

We have in each case

$$P(B) \geq 1 - 4p^{-\frac{(\beta - 1)^2\alpha^2}{4(\beta + 1)^2}}.$$  

Finally note that by the choice of $\alpha$ and $\beta$

$$(\beta - 1)^2\alpha^2 > 1.$$

**Proof of Theorem 2.** The essential part of Proof is to show inequality (3.5). Let $A_{ij}$ be the event that $\hat{\tau}_{ij}$ estimates $\tau_{ij}$ by 0, i.e.

$$A_{ij} := \left\{ |\hat{\tau}_{ij}| \leq \alpha \sqrt{\frac{\log p}{n}} \right\}.$$

Moreover define the event

$$B_{ij} := \left\{ |\hat{\tau}_{ij} - \tau_{ij}| \leq \beta \min\left( |\tau_{ij}|, \alpha \sqrt{\frac{\log p}{n}} \right) \right\}.$$

We only prove the case $0 < q < 1$. The case $q = 0$ is even easier and therefore omitted. Denote $[\cdot] := [\cdot]_q$. Analogously to Cai and Zhou (2012) we split $\frac{1}{p} E||\hat{\tau} - \tau||_F^2$ into two parts

$$\frac{1}{p} E||\hat{\tau} - \tau||_F^2 = \frac{1}{p} \sum_{i=1}^{p} \sum_{j=1, j \neq i}^{p} E|\tau_{ij} - \hat{\tau}_{ij}|^2$$

$$= \frac{1}{p} \sum_{i=1}^{p} \sum_{j=1, j \neq i}^{p} E|\tau_{ij}|^2 \mathbb{1}_{N_{ij}} + \frac{1}{p} \sum_{i=1}^{p} \sum_{j=1, j \neq i}^{p} E|\tau_{ij} - \hat{\tau}_{ij}|^2 \mathbb{1}_{N_{ij}}$$

$$= I_1 + I_2.$$

**Bounding $I_1$:** First consider the case that

$$\frac{(2 - q)^{y/2} e^{-n_p}}{q^{y/2} \alpha^y (\log p)^{y/2}} < 1.$$
Then

\[ I_1 = \frac{1}{p} \sum_{i=1}^{p} \sum_{j=1 \atop j \neq i}^{p} \mathbb{E}|\tau_{ij} - \tilde{\tau}_{ij}|^2 \mathbb{I}_{B_{ij}} \leq \beta^2 c_{n,p} \left( 1 + \sum_{i=2}^{p} i^{-2/q} \right) \]

\[ \leq \beta^2 c_{n,p} \left( 1 + \int_{1}^{\infty} x^{-2/q} dx \right) = \left( 1 + \frac{q}{2} - q \right) \beta^2 c_{n,p} \leq 2 \beta^2 \alpha^2 m \left( 2 - \frac{q}{q} \right) \sum_{i=1}^{p} \frac{\log p}{n} 1/q \geq 2. \]  

Otherwise fix \( m > 1 \). Then we have

\[ I_1 = \frac{1}{p} \sum_{i=1}^{p} \sum_{j=1 \atop j \neq i}^{p} \mathbb{E}|\tau_{ij} - \tilde{\tau}_{ij}|^2 \mathbb{I}_{B_{ij}} + \frac{1}{p} \sum_{i=1}^{p} \sum_{j=m+1}^{p} \mathbb{E}|\tau_{ij} - \tilde{\tau}_{ij}|^2 \mathbb{I}_{B_{ij}} \]

\[ \leq m \beta^2 \alpha^2 \log \frac{p}{n} + \beta^2 c_{n,p} \sum_{i=m+1}^{\infty} i^{-2/q} \]

\[ \leq m \beta^2 \alpha^2 \log \frac{p}{n} + \beta^2 c_{n,p} \int_{m}^{\infty} x^{-2/q} dx \]

\[ = m \beta^2 \alpha^2 \log \frac{p}{n} + \left( \frac{2}{q} - 1 \right) \beta^2 c_{n,p} m^{-2/q+1}. \]

For \( m = \left( \frac{2-q}{q} \right)^{1/2} c_{n,p} n^{-1/2} / \log (\log p)^{1/2} \) + 1 we get

\[ I_1 \leq \left( 2 \alpha^{-q} \left( \frac{2 - q}{q} \right) \frac{1}{q} \right) \beta^2 \alpha^2 c_{n,p} \left( \frac{\log p}{n} \right)^{1/q}. \]  

(7.3)

Bounding \( I_2 \): It remains to show that \( I_2 \) is of equal or smaller order than \( I_1 \). We have

\[ I_2 = p^{-1} \sum_{i,j=1 \atop i \neq j}^{p} \mathbb{E}(\tilde{\tau}_{ij} - \tau_{ij})^2 \mathbb{I}_{B_{ij}^c} + p^{-1} \sum_{i,j=1 \atop i \neq j}^{p} \mathbb{E} \left( \tau_{ij}^2 \mathbb{I}_{B_{ij}^c} \right). \]  

(7.4)

The first summand in \( I_2 \) can be assessed by Hölder’s inequality. For any \( N \geq 3 \) we have

\[ p^{-1} \sum_{i,j=1 \atop i \neq j}^{p} \mathbb{E}(\tilde{\tau}_{ij} - \tau_{ij})^2 \mathbb{I}_{B_{ij}^c} \leq p^{-1} \sum_{i,j=1 \atop i \neq j}^{p} \mathbb{E}^{1/N}(\tilde{\tau}_{ij} - \tau_{ij})^2 \mathbb{I}_{B_{ij}^c} \]

\[ \leq \frac{16 N p}{n} p^{-1-N/(1+\epsilon)} \]

where we used inequality \( I_1 \) and Stirling’s approximation to bound the expectation \( \mathbb{E}(\tilde{\tau}_{ij} - \tau_{ij})^2 N \mathbb{I}_{B_{ij}^c} \) by the formula

\[ \mathbb{E}(\tilde{\tau}_{ij} - \tau_{ij})^2 N = \int_{0}^{\infty} \mathbb{P}((\tilde{\tau}_{ij} - \tau_{ij})^{2N} \geq x) dx. \]
and Lemma 3 for $\mathbb{P}^{1-1/N}(B_{ij}^c)$. By taking $N = \frac{1}{\varepsilon}$ we conclude

$$p^{-1} \sum_{i,j=1 \atop i \neq j}^p E(\hat{\tau}_{ij} - \tau_{ij})^2 \mathbb{I}_{B_{ij}^c \cap A_{ij}} \leq 16 \frac{1 + \varepsilon}{n \varepsilon}$$

$$\leq 16 \frac{1 + \varepsilon}{\varepsilon} c_{n,p} \left( \frac{\log p}{n} \right)^{1-q/2}.$$  \hspace{1cm} (7.5)

Lastly consider the second summand in (7.4). We observe that the event $B_{ij}^c \cap A_{ij}$ can only occur if $|\tau_{ij}| \geq \alpha \sqrt{\log p n}$. Therefore we obtain

$$p^{-1} \sum_{i,j=1 \atop i \neq j}^p E\tau_{ij}^2 \mathbb{I}_{B_{ij}^c \cap A_{ij}} \leq p^{-1} \sum_{i,j=1 \atop i \neq j}^p \tau_{ij}^2 \mathbb{E}A_{ij} \mathbb{I}_{\{|\tau_{ij}| - |\hat{\tau}_{ij} - \tau_{ij}| < \alpha \sqrt{\log p n}\}} \mathbb{I}_{\{|\tau_{ij}| \geq \alpha \beta \sqrt{\log p n}\}}$$

$$\leq p^{-1} \sum_{i,j=1 \atop i \neq j}^p \tau_{ij}^2 \mathbb{E} \mathbb{I}_{\{|\tau_{ij} - (1 - \frac{\beta}{2}) |\tau_{ij}| > \alpha \beta \sqrt{\log p n}\}} \mathbb{I}_{\{|\tau_{ij}| \geq \alpha \beta \sqrt{\log p n}\}}$$

$$\leq p^{-1} \sum_{i,j=1 \atop i \neq j}^p \tau_{ij}^2 \mathbb{E} \mathbb{I}_{\{|\tau_{ij}| \geq \alpha \beta \sqrt{\log p n}\}} \mathbb{I}_{\{|\tau_{ij}| \leq \alpha \beta \sqrt{\log p n}\}}$$

$$\leq p^{-1} \frac{2}{n} \sum_{i,j=1}^p n \tau_{ij}^2 \exp \left( - \frac{(1 - \beta - \frac{1}{2})^2 n \tau_{ij}^2}{4} \right) \mathbb{I}_{\{|\tau_{ij}| \geq \alpha \beta \sqrt{\log p n}\}}$$

$$\leq \frac{2 \alpha^2 \beta^2}{3e} \sum_{i,j=1}^p 3n \tau_{ij}^2 \exp \left( - \frac{3n \tau_{ij}^2}{\alpha^2 \beta^2} \right)$$

$$\leq \frac{2 \alpha^2 \beta^2}{3e}.$$  \hspace{1cm} (7.6)

The last line follows by the the fact the function $f(x) = x \exp(-x)$ is bounded from above by $e^{-1}$. Therefore,

$$p^{-1} \sum_{i,j=1 \atop i \neq j}^p E\tau_{ij}^2 \mathbb{I}_{B_{ij}^c \cap A_{ij}} \leq \frac{2 \alpha^2 \beta^2}{3e} c_{n,p} \left( \frac{\log p}{n} \right)^{q/2}.$$  \hspace{1cm} (7.6)

Summarizing the bounds (7.2), (7.3), (7.5) and (7.6) yields inequality (3.5). Now inequality (3.6) is a conclusion of (3.5) since

$$\sup_{\tau \in \mathbb{R}^{w,q}(c_{n,p})} E\|\hat{\tau} - \tau\|_p^2 \geq \frac{4}{\pi^2} \sup_{\rho \in \mathbb{R}^{w,q}(c_{n,p})} E\|\hat{\rho} - \rho\|_p^2.$$  \hspace{1cm} \Box

**Proof of Theorem** 5 By definition an entry $\sin[\frac{\pi}{2} \hat{\tau}_{ij}]$, $i \neq j$ will be deleted iff $|\sin[\frac{\pi}{2} \tau_{ij}]| \leq $
Critical threshold levels

\[ \alpha \sqrt{\frac{\log p}{n}}. \] Rearranging yields

\[ |\sin \left( \frac{\pi}{2} \hat{\tau}_{ij} \right) | \leq \alpha \sqrt{\frac{\log p}{n}} \Leftrightarrow |\sin \left( \frac{\pi}{2} \hat{\tau}_{ij} \right) | \leq \alpha \sqrt{\frac{\log p}{n}} \wedge 1 \]

\[ \Leftrightarrow |\hat{\tau}_{ij} | \leq \frac{2}{\pi} \arcsin \left( \alpha \sqrt{\frac{\log p}{n}} \wedge 1 \right). \]

By the mean value theorem there exists \( \theta \in \left( 0, \alpha \sqrt{\frac{\log p}{n}} \wedge 1 \right) \) such that

\[ \arcsin \left( \alpha \sqrt{\frac{\log p}{n}} \wedge 1 \right) = \left( \alpha \sqrt{\frac{\log p}{n}} \wedge 1 \right) \frac{1}{\sqrt{1-\theta^2}}. \]

Moreover by convexity of the arcsine function on \([0,1]\) we have \( \lambda := \frac{1}{\sqrt{1-\theta^2}} \in [1, \frac{2}{\alpha}] \), where \( \lambda = \frac{2}{\alpha} \) if \( \alpha \sqrt{\frac{\log p}{n}} \geq 1 \). Therefore we have \( \hat{\rho} = T_\alpha (\sin \left( \frac{\pi}{2} \hat{\tau} \right)) = \sin \left( \frac{\pi}{2} T_{\frac{2}{\alpha} \lambda} (\hat{\tau}) \right) \). Finally we conclude by Theorem 2 that

\[ \sup_{\rho \in \mathcal{G}_{\omega,q}(c_{n,p})} \frac{1}{p} E \| \hat{\rho} - \rho \|_F^2 \leq \sup_{\lambda \in \left[ 1, \frac{2}{\alpha} \right]} \sup_{\rho \in \mathcal{G}_{\omega,q}(c_{n,p})} \frac{1}{p} E \| \sin \left( \frac{\pi}{2} T_{\frac{2}{\alpha} \lambda} (\hat{\tau}) \right) - \rho \|_F^2 \]

\[ = \sup_{\lambda \in \left[ 1, \frac{2}{\alpha} \right]} \sup_{\rho \in \mathcal{G}_{\omega,q}(c_{n,p})} \frac{1}{p} E \| \sin \left( \frac{\pi}{2} T_{\frac{2}{\alpha} \lambda} (\hat{\tau}) \right) - \rho \|_F^2 \]

\[ \leq \sup_{\alpha \in \left[ 1, \frac{2}{\alpha} \right]} C_{\alpha,c_{n,p}} \left( \frac{\log p}{n} \right)^{1-\frac{q}{2}} \]

\[ \leq C_{\alpha,c_{n,p}} \left( \frac{\log p}{n} \right)^{1-\frac{q}{2}}, \]

where by the proof of Theorem \( 2 \) the expression \( \sup_{\alpha \in \left[ 1, \frac{2}{\alpha} \right]} C_{\alpha} \) is bounded for any \( \alpha > \pi \). \( \square \)

Next we provide the proof of Theorem 3. It is sufficient to restrict to the case that \( \alpha \leq 2 \) in the upper bound (4.1). In view of the proof of Theorem 2 we need to improve Lemma 4. This can be done by distinguishing between entries \( \hat{\tau}_{ij}^* \) based on weakly correlated components and all the other entries \( \hat{\tau}_{ij} \). Clearly, if \( \tau_{ij} \) is sufficiently large compared to the threshold level \( \alpha \sqrt{\frac{\log p}{n}} \), then \( |\hat{\tau}_{ij}^* - \tau_{ij}| = |\hat{\tau}_{ij} - \tau_{ij}| = O \left( \frac{\log p}{n} \right) \) with an appropriately large probability.

The next lemma gives a more precise statement of the last thought.

**Lemma 5**

Let \((Y_1, Z_1), \ldots, (Y_n, Z_n) \sim N_2(\mu, \Sigma)\) with correlation \( |\rho| \geq \frac{2\pi}{\alpha} \sqrt{\frac{\log p}{n}} \) Then for any \( \frac{\alpha}{2} < \alpha \leq 2 \) we have

\[ \mathbb{P} \left( |\hat{\tau}^*(Y_1, n, Z_{1:n}) - \tau(Y_1, Z_1)| \leq 3 \sqrt{\frac{\log p}{n}} \right) \geq 1 - 4p^{-\frac{q}{2}}. \]
Proof. The proof is similar to the third part of Lemma \[4\] Let \( \tau := \tau(Y_1, Z_1) \), \( \hat{\tau} := \hat{\tau}(Y_{1:n}, Z_{1:n}) \) and \( \hat{\tau}^* := \hat{\tau}^*(Y_{1:n}, Z_{1:n}) \). First note that \( |\tau| \geq \frac{2}{\pi} |\rho| \geq 5 \sqrt{\log n} \). Remind that

\[ A = \left\{ |\hat{\tau}| < \alpha \sqrt{\frac{\log p}{n}} \right\}. \]

Then we have by Hoeffding’s inequality

\[
\begin{align*}
P\left(|\hat{\tau}^* - \tau| \geq 3 \sqrt{\frac{\log p}{n}}\right) &\geq P\left(|\hat{\tau} - \tau| \geq 3 \sqrt{\frac{\log p}{n}} \cap A^c\right) \\
&\geq 1 - 2p^{-\frac{3}{4}} - P\left(|\hat{\tau} - \tau| \geq 3 \sqrt{\frac{\log p}{n}}\right) \\
&\geq 1 - 2p^{-\frac{3}{4}} - P\left(|\hat{\tau} - \tau| \geq \beta \min\left(|\tau(Y_1, Z_1)|, \alpha \sqrt{\frac{\log p}{n}}\right)\right) \\
&\geq 1 - 2p^{-\frac{3}{4}},
\end{align*}
\]

where the second last line follows by triangle inequality since \( \alpha \leq 2 \), \( |\hat{\tau}| < \alpha \sqrt{\frac{\log p}{n}} \) and \( |\tau| \geq 5 \sqrt{\frac{\log p}{n}} \) implies that \( |\hat{\tau} - \tau| \geq 3 \sqrt{\frac{\log p}{n}} \). Hence the claim holds true.

Now we can give a more refined version of Lemma \[4\] for Gaussian random vectors by treating the entries \( \hat{\tau}^*_{ij} \) based on weakly correlated components more carefully.

**Lemma 6**

Let \((Y_1, Z_1), \ldots, (Y_n, Z_n) \sim_{i.i.d.} N_2(\mu, \Sigma) \) where \( p \leq n^{\eta_u} \). Then for any \( \frac{2}{\pi} < \alpha \leq 2 \) and \( \beta > \frac{3\alpha + 2}{3\alpha - 2} \) let the event

\[
B := \left\{ |\hat{\tau}^*(Y_{1:n}, Z_{1:n}) - \tau(Y_1, Z_1)| \leq \beta \min\left(|\tau(Y_1, Z_1)|, \alpha \sqrt{\frac{\log p}{n}}\right) \right\}.
\]

Then we have

\[
P(B) \geq 1 - C_{\alpha, \beta, \eta_u} p^{-\varepsilon},
\]

where \( \varepsilon > 0 \) depends on \( \alpha, \beta \) and \( \eta_u \).

**Proof.** For \( |\rho(Y_1, Z_1)| \geq \frac{5\pi}{7} \sqrt{\frac{\log p}{n}} \) we have

\[
\beta \min\left(|\tau(Y_1, Z_1)|, \alpha \sqrt{\frac{\log p}{n}}\right) \geq \beta \alpha \sqrt{\frac{\log p}{n}} > 3 \sqrt{\frac{\log p}{n}}
\]

and therefore by Lemma \[4\] the inequality holds here. Otherwise for \( |\rho(Y_1, Z_1)| < \frac{5\pi}{7} \sqrt{\frac{\log p}{n}} \) we just need to replace Hoeffding’s inequality in the proof of Lemma \[4\] by the upper bound on the probability that \( \hat{\tau}^*(Y_{1:n}, Z_{1:n}) \) is close its mean \( \tau(Y_1, Z_1) \) in Corollary \[4\].
Proof of Theorem 4. The proof of the inequalities (4.1) and (4.2) is essentially analogous to the one of Theorem 2. Using Lemma 6 instead of Lemma 4 to compute $P_{i,j}$ provides the desired upper bound. We choose an appropriate example to show that for $\alpha < \frac{2}{\sqrt{3}}$ the corresponding estimators $\hat{\tau}^*$ and $\hat{\rho}^*$ in general do not attain the minimax rate. Since $\text{Id} \in G_{w,q}(c_{n,p})$ for any $0 \leq q < 1$ and any $c_{n,p} > 0$, we assume that $\Sigma = \text{Id}$. Furthermore let $\varepsilon := 1 - \frac{9}{8} \alpha^2$ and $c_{n,p} = o\left(\left(\frac{\log p}{n}\right)^{q/2} \frac{p^c}{\sqrt{n}}\right)$. The lower bound in Corollary 1 provides

$$E_p^{-1}\|\hat{\tau}^* - \tau\|_F^2 \geq p^{-1} \sum_{i,j=1}^p (\hat{\tau}_{ij}^* - \tau_{ij})^2 \|A_{ij}^c\|$$

$$\geq \frac{a^2}{np} \frac{\log p}{n} \sum_{i,j=1}^p \mathbb{P}(A_{ij}^c)$$

$$\geq C_{\alpha, a, q} \frac{p^{c} \log p}{n \sqrt{\log p}}$$

$$= C_{\alpha, a, q, c_{n,p}} \left(\frac{\log p}{n}\right)^{1-\frac{q}{2}} \frac{p^{c} \log p}{c_{n,p}}\left(\frac{\log p}{n}\right)^{q/2}.$$ 

The last line proves that $\hat{\tau}^*$ does not attain the minimax rate. It remains to give a suitable lower bound on $\sup_{\rho \in G_{w,q}(c_{n,p})} E\|\sin\left(\frac{\pi}{2} \hat{\tau}^*\right) - \rho\|_F^2$. Again we first bound this expression by the special case $\Sigma = \text{Id}$. We have

$$\sup_{\rho \in G_{w,q}(c_{n,p})} p^{-1} E\|\sin\left(\frac{\pi}{2} \hat{\tau}^*\right) - \rho\|_F^2 \geq p^{-1} E\|\sin\left(\frac{\pi}{2} \hat{\tau}^*\right) - \text{Id}\|_F^2 \geq p^{-1} E\|\hat{\tau}^* - \text{Id}\|_F^2$$

$$\geq C_{\alpha, a, q, c_{n,p}} \left(\frac{\log p}{n}\right)^{1-\frac{q}{2}} \frac{p^{c} \log p}{c_{n,p}}\left(\frac{\log p}{n}\right)^{q/2}.$$ 

The last statement of the Theorem follows analogously, where the upper bound on $c_{n,p}$ in assumption (A3) is to be regarded.

Proof of Theorem 5. Inequality (4.3) follows similarly to inequality (4.7). For the proof of the lower bound we use the implication

$$|\sin\left(\frac{\pi}{2} x\right)| \leq \alpha \sqrt{\frac{\log p}{n}} \implies |x| \leq \frac{2 \alpha}{\pi} \sqrt{\frac{\log p}{n}}$$

on $[-1, 1]$. We conclude

$$\sup_{\rho \in G_{w,q}(c_{n,p})} p^{-1} E\|\hat{\rho} - \rho\|_F^2 \geq p^{-1} E\|T_{\alpha}\left(\sin\left(\frac{\pi}{2} \hat{\tau}^*\right)\right) - \text{Id}\|_F^2$$

$$\geq p^{-1} E\|\sin\left(\frac{\pi}{2} \hat{\tau}^*\right) - \text{Id}\|_F^2$$

$$\geq C_{\alpha, a, q, c_{n,p}} \left(\frac{\log p}{n}\right)^{1-\frac{q}{2}} \frac{p^{c} \log p}{c_{n,p}}\left(\frac{\log p}{n}\right)^{q/2}.$$ 

Analogously we obtain the last statement of the Theorem.
References

Abramovich, F. et al. (2006). “Adapting to unknown sparsity by controlling the false discovery rate”. In: Ann. Stat. 34, pp. 584–653.

Amini, A. and Wainwright, M. (2009). “High-dimensional analysis of semidefinite relaxations for sparse principal components”. In: Ann. Statist. 37, pp. 2877–2921.

Baik, Jinbo and Silverstein, Jack W. (2006). “Eigenvalues of large sample covariance matrices of spiked population models”. In: J. Multivariate Anal. 97.6.

Berthet, Q. and Rigollet, P. (2013). “Optimal detection of sparse principal components in high dimensions”. In: Ann. Statist. 41, pp. 1780–1815.

Bickel, P. and Levina, E. (2008a). “Covariance regularization by thresholding”. In: Ann. Stat. 36.6, pp. 2577–2604.

— (2008b). “Regularized estimation of large covariance matrices”. In: Ann. Stat. 36.1, pp. 199–227.

Cai, T., Ma, Z., and Wu, Y. (2013). “Sparse PCA: Optimal rates and adaptive estimation”. In: Ann. Statist. 41, pp. 3074–3110.

— (2014+). “Optimal estimation and rank detection for sparse spiked covariance matrices”. To appear in: Prob. Theory Rel. Fields.

Cai, T. and Zhou, H. (2012a). “Covariance matrix estimation under $l_1$ norm”. In: Statist. Sinica 22, pp. 1319–1378.

— (2012b). “Optimal rates of convergence for sparse covariance matrix estimation”. In: Ann. Stat. 40, pp. 2389–2420.

Candès, E. et al. (2010). “Compressed sensing with coherent and redundant dictionaries”. In: Appl. Comput. Harmon. Anal. 31, pp. 59–73.

Clark, L. (2000). “An Asymptotic Expansion for the Number of Permutations with a Certain Number of Inversions”. In: Electron. J. Comb. 7.

Cramer, G. (1750). Introduction à l’analyse des lignes courbes algébriques. Genève: Frères Cramer et C. Philibert.

D’Aspremont, A. et al. (2007). “A direct formulation for sparse PCA using semidefinite programming”. In: SIAM Rev. Vol. 49, pp. 434–448.

Donoho, D. (2006). “Compressed sensing”. In: IEEE Trans. Inform. Theory 52, pp. 1289–1306.

El Karoui, N. (2008). “Operator norm consistent estimation of large-dimensional sparse covariance matrices”. In: Ann. Stat. 36, pp. 2717–2756.

Embrechts, P., Lindskog, F., and McNeil, A. (2003). “Modelling dependence with copulas and applications to risk management”. In: Handbook of Heavy Tailed Distributions in Finance 8 (1), pp. 329–384.

Fan, J., Yuan, L., and Mincheva, M (2013). “Large covariance estimation by thresholding principal orthogonal complements”. In: J. Roy. Statist. Soc. Ser. B 75, pp. 603–680.

Fang, H., Fang, K., and Kotz, S. (2002). “The meta-elliptical distributions with given marginals”. In: J. Mult. Anal. 82.1-16.

Fang, K., Kotz, S., and Ng, K. (1990). Symmetric Multivariate and Related Distributions. Chapman & Hall.
Han, F. and Liu, H. (2013). “Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution”. Preprint.
— (2014). “Scale-invariant sparse PCA on high dimensional meta-elliptical data”. In: J. Mach. Learn. Res. 14, pp. 629–671.
Han, F., Liu, H., and Zhao, T. (2013). “CODA: High dimensional copula discriminant analysis”. In: J. Mach. Learn. Res. 14, pp. 629–671.
Hoeffding, W. (1963). “Probability inequalities for sums of bounded random variables”.
Huang, J. and Shen, H. (2008). “Sparse principal component analysis via regularized low rank matrix approximation”. In: J. Multivariate Anal. 99, pp. 1015–1034.
Hult, H. and Lindskog, F. (2002). “Multivariate extremes, aggregation and dependence in elliptical distributions”. In: Adv. in Appl. Probab. 34, pp. 587–608.
Johnstone, I. and Lu, Y. (2009). “On consistency and sparsity for principal components analysis in high dimensions”. In: J. Am. Stat. Assoc. 104.486, pp. 682–693.
Johnstone, Iain (2001). “On the distribution of the largest eigenvalue in principal component analysis.” In: Ann. Stat. 29.2, pp. 295–327.
Johnstone, Iain and Paul, Debashis (2007). “Sparse principal component analysis for high dimensional data”. Technical report.
Kendall, M. (1938). “A new measure of rank correlation”. In: Biometrika 30, pp. 81–93.
Lehmann, E. and Casella, G. (1998). Theory of Point Estimation. Springer, New York.
Levina, E. and Vershynin, R. (2012). “Partial estimation of covariance matrices”. In: Prob. Theory Rel. Fields 153, pp. 405–419.
Liu, H., Lafferty, J., and Wasserman, L. (2009). “The nonparanormal: Semiparametric estimation of high dimensional undirected graphs”. In: J. Mach. Learn. Res. 10, pp. 2295–2328.
Liu, H. et al. (2012). “High-dimensional semiparametric gaussian copula graphical models”. In: Ann. Statist. 40, pp. 2293–2326.
Muir, T. (1900). “Determination of the sign of a single term of a determinant”. In: Proc. Roy. Soc. Edinb. 22, pp. 441–477.
Saulis, L. and Statulevicius, V. (1991). Limit Theorems for Large Deviations. Springer Science & Business Media.
Vu, V. and Lei, J. (2013). “Minimax sparse principal subspace estimation in high dimensions”. In: Ann. Statist. 41, pp. 2905–2947.
Wegkamp, M. and Zhao, Y. (2013). “Adaptive Estimation of the Copula Correlation Matrix for Semiparametric Elliptical Copulas”. Preprint.
Xue, L. and Zou, H. (2012). “Regularized rank-based estimation of high-dimensional nonparanormal graphical models”. In: Ann. Statist. 40, pp. 2541–2571.