ECA: High Dimensional Elliptical Component Analysis in non-Gaussian Distributions
(Supplementary Appendix)

Fang Han* and Han Liu†

The supplementary materials provide more simulation results, as well as all technical proofs.

A  More Simulation Results

This section provides more simulation results.

A.1  More Results in Section 6.1.1

Following the results in Section 6.1.1, we evaluate ECA’s dependence on sample size and dimension with the sparsity values \( s = 5 \) and 20. Here all the nonzero entries in \( v_1 \) and \( v_2 \) are set to be equal to \( 1/\sqrt{s} \). All the other parameters remain same as in Section 6.1.1. The corresponding results are put in Figures 1, and the conclusion drawn in Section 6.1.1 still holds here.

A.2  More Results in Section 6.1.2

Following the results in Section 6.1.2, we further evaluate the performance of ECA and its competitors when the data are Cauchy distributed. In particular, we consider the following setting:

\[
(\text{Cauchy}) \quad X \sim EC_d(0, \Sigma, \xi_2 \sqrt{d}) \quad \text{with} \quad \xi_2 \overset{d}{=} \xi_1^*/\xi_2^*. \quad \text{Here} \quad \xi_1^d = \chi_d \quad \text{and} \quad \xi_2^d = \chi_1.
\]

All the other parameters remain same as in Schemes 1, 2, 3 in Section 6.1.2. Figures 2 and 3 illustrate the estimation and model selection efficiency of the competing methods. It could be

*Department of Statistics, University of Washington, Seattle, WA 98195, USA; e-mail: fanghan@uw.edu
†Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA; e-mail: hanliu@princeton.edu
Figure 1: Simulation for multivariate-$t$ with varying numbers of dimension $d$ and sample size $n$. Plots of averaged distances between the estimators and the true parameters are conducted over 1,000 replications. (A) Multivariate-$t$ distribution with $s = 5$; (B) Multivariate-$t$ distribution with $s = 20$.

observed that ECA’s performance remains best. Actually, its advantage over TP and TCA is more significant than that under the multivariate-$t$ with the degree of freedom 3. This is as expected since the model is even more heavy-tailed now. The results also, similar to the case EC1, empirically verify that ECA works well even if the covariance matrix does not exist.

B Proofs

In this section we provide the proofs of results shown in Sections 2, 3, 4, and 5.

B.1 Proofs of Results in Section 2

This section proves Proposition 2.1. The proof summarizes the results in Marden (1999) and Croux et al. (2002), and is provided only for completeness. In particular, we do not claim any original contribution.
Figure 2: Curves of averaged distances between the estimates and true parameters for different schemes and Cauchy distribution using the FTPM algorithm. Here we are interested in estimating the leading eigenvector. The horizontal-axis represents the cardinalities of the estimates’ support sets and the vertical-axis represents the averaged distances.

Figure 3: ROC curves for different methods in schemes 1 to 3 and Cauchy distributions using the FTPM algorithm. Here we are interested in estimating the sparsity pattern of the leading eigenvector.
**Lemma B.1.** Let $X \sim EC_d(\mu, \Sigma, \xi)$ be a continuous random vector. We have

$$K = \mathbb{E}\left(\frac{(X - \bar{X})(X - \bar{X})^T}{\|X - \bar{X}\|_2^2}\right) = \mathbb{E}\left(\frac{(X - \mu)(X - \mu)^T}{\|X - \mu\|_2^2}\right).$$

(B.1)

*Proof.* By the equivalent definition of the elliptical distribution, there exists a characteristic function $\psi$ uniquely determined by $\xi$ such that $X \sim EC_d(\mu, \Sigma, \psi)$ and $\tilde{X} \sim EC_d(\mu, \Sigma, \psi)$. Let $i := \sqrt{-1}$. Since $X$ and $\tilde{X}$ are independent, we have $\mathbb{E}\exp(it^T(X - \bar{X})) = \mathbb{E}\exp(it^T X)\mathbb{E}\exp(-it^T \tilde{X}) = \psi^2(t^T \Sigma t)$, implying that $X - \tilde{X} \sim EC_d(0, \Sigma, \psi^2)$. Again, by the equivalent definition of the elliptical distribution, there exists a nonnegative random variable $\xi'$ uniquely determined by $\psi^2$, such that $X - \tilde{X} \sim EC_d(0, \Sigma, \xi')$. Because $X$ is continuous, we have $\mathbb{P}(\xi' = 0) = 0$. Therefore,

$$K = \mathbb{E}\left(\frac{(X - \bar{X})(X - \bar{X})^T}{\|X - \bar{X}\|_2^2}\right) = \mathbb{E}\left(\frac{(\xi' \Sigma \xi')(\xi' \Sigma \xi')^T}{\|\xi' \Sigma \xi'\|_2^2}\right) = \mathbb{E}\left(\frac{\xi \Sigma \xi}{\|\xi\Sigma\|_2^2}\right) = \mathbb{E}\left(\frac{(X - \mu)(X - \mu)^T}{\|X - \mu\|_2^2}\right).$$

This completes the proof. $\square$

*Proof of Proposition 2.1.* Using Lemma B.1, it is equivalent to consider $K = \mathbb{E}\left(\frac{(X - \mu)(X - \mu)^T}{\|X - \mu\|_2^2}\right)$. Letting $\Omega := [u_1(\Sigma), \ldots, u_d(\Sigma)]$, $u_{d+1}(\Sigma)$ until $u_d(\Sigma)$ chosen to be orthogonal to $u_1(\Sigma), \ldots, u_d(\Sigma)$ (which are also specified to be orthogonal to each other), we have

$$\|X - \mu\|_2 = \|\Omega^T (X - \mu)\|_2.$$  

This implies that

$$\Omega^T \frac{X - \mu}{\|X - \mu\|_2} = \Omega^T \frac{X - \mu}{\|\Omega^T (X - \mu)\|_2} = \frac{Z}{\|Z\|_2},$$

where using the stochastic representation of $X$ in Equation (2.1), we have $Z = \Omega^T A U = D U$ with $D = (\text{diag}(\sqrt{\lambda_1(\Sigma)}, \ldots, \sqrt{\lambda_q(\Sigma)}), 0)^T \in \mathbb{R}^{d \times q}$. Therefore,

$$K = \mathbb{E}\left(\frac{(X - \mu)(X - \mu)^T}{\|X - \mu\|_2^2}\right) = \Omega \cdot \left[\mathbb{E}\left(\frac{Z Z^T}{\|Z\|_2^2}\right)\right] \cdot \Omega^T.$$

Secondly, we prove that $\mathbb{E}\left(\frac{Z Z^T}{\|Z\|_2^2}\right)$ is a diagonal matrix. This is because, for any matrix $P = \text{diag}(v)$, where $v = (v_1, \ldots, v_d)^T$ satisfies that $v_j = 1$ or $-1$ for $j = 1, \ldots, d$, we have

$$P \frac{Z}{\|Z\|_2} = \frac{PZ}{\|PZ\|_2} \overset{d}{=} \frac{Z}{\|Z\|_2} \Rightarrow \mathbb{E}\left(\frac{Z Z^T}{\|Z\|_2^2}\right) = P \left[\mathbb{E}\left(\frac{Z Z^T}{\|Z\|_2^2}\right)\right] P.$$

It holds if and only if $\mathbb{E}\left(\frac{Z Z^T}{\|Z\|_2^2}\right)$ is a diagonal matrix.
To finish the proof, we need to show that the diagonals of \( \mathbb{E}\left( \frac{ZZ^T}{\|Z\|_2^2} \right) \) are decreasing. Reminding that \( Z = DU \), we have that
\[
\mathbb{E}\left( \frac{ZZ^T}{\|Z\|_2^2} \right) = \mathbb{E}\left( \frac{D U U^T D}{U^T D U} \right).
\]
Letting \( U := (U_1, \ldots, U_q)^T \), by algebra, for \( j = 1, \ldots, q \),
\[
\left[ \mathbb{E}\left( \frac{ZZ^T}{\|Z\|_2^2} \right) \right]_{jj} = \mathbb{E}\left( \frac{\lambda_j(\Sigma)U_j^2}{\lambda_1(\Sigma)U_1^2 + \ldots + \lambda_q(\Sigma)U_q^2} \right).
\]
Actually, we have for any \( k < j \),
\[
\frac{\lambda_k(K)}{\lambda_j(K)} = \frac{\mathbb{E}\frac{\lambda_k(\Sigma)U_k^2}{\lambda_j(\Sigma)U_j^2 + \lambda_k(\Sigma)U_k^2 + E}}{\mathbb{E}\frac{\lambda_j(\Sigma)U_j^2}{\lambda_j(\Sigma)U_j^2 + \lambda_k(\Sigma)U_k^2 + E}} < \frac{\mathbb{E}\frac{\lambda_k(\Sigma)U_k^2}{\lambda_k(\Sigma)U_k^2 + E}}{\mathbb{E}\frac{\lambda_j(\Sigma)U_j^2}{\lambda_j(\Sigma)U_j^2 + E}} = \frac{U_k^2}{U_j^2 + E/\lambda_k(\Sigma)} < 1,
\]
where we let \( E := \sum_{i \in \{j, k\}} \lambda_i(\Sigma)U_i^2 \). This completes the proof. \( \square \)

B.2 Proofs of Results in Section 3

In this section we provide the proofs of Theorems 3.1 and 3.5. To prove Theorem 3.1, we exploit the U-statistics version of the matrix Bernstein’s inequality (Tropp, 2012), which is given in the following theorem.

**Theorem B.2** (Matrix Bernstein’s inequality for U-statistics). Let \( k(\cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^{d \times d} \) be a matrix value function. Let \( X_1, \ldots, X_n \) be \( n \) independent observations of an random variable \( X \in \mathcal{X} \). Suppose that, for any \( i \neq i' \in \{1, \ldots, n\} \), \( \mathbb{E}k(X_i, X_{i'}) \) exists and there exist two constants \( R_1, R_2 > 0 \) such that
\[
\|k(X_i, X_{i'}) - \mathbb{E}k(X_i, X_{i'})\|_2 \leq R_1 \text{ and } \|\mathbb{E}\{k(X_i, X_{i'}) - \mathbb{E}k(X_i, X_{i'})\}^2\|_2 \leq R_2. \tag{B.2}
\]
We then have
\[
\mathbb{P}\left( \left\| \frac{1}{n} \sum_{i < i'} k(X_i, X_{i'}) - \mathbb{E}k(X_1, X_2) \right\|_2 \geq t \right) \leq d \exp\left( -\frac{(n/4)t^2}{R_2 + R_1 t/3} \right) \leq \begin{cases} d \cdot \exp\left( -\frac{3nt^2}{16R_2} \right), & \text{for } t \leq R_2/R_1; \\ d \cdot \exp\left( -\frac{3nt}{16R_1} \right), & \text{for } t > R_2/R_1. \end{cases}
\]

**Proof.** The proof is the combination of the Hoeffding’s decoupling trick and the proof of the independent matrix Bernstein’s inequality shown in Tropp (2012). A detailed analysis is given in the proofs of Theorem 2.1 in Wegkamp and Zhao (2016) and Theorem 3.1 in Han and Liu (2016). We refer to theirs for details. \( \square \)
With the matrix Bernstein inequality of U-statistics, we proceed to prove Theorem 3.1. This is equivalent to calculating $R_1$ and $R_2$ in (B.2) for the particular U-statistics $k_{MK}$ defined in (2.5).

**Proof of Theorem 3.1.** Let’s first calculate the terms $R_1$ and $R_2$ in Theorem B.2 for the particular kernel function

$$k_{MK}(X_i, X_{i'}) := \frac{(X_i - X_{i'})(X_i - X_{i'})^T}{\|X_i - X_{i'}\|_2^2}.$$ 

First, we have

$$\left\| \frac{(X_i - X_{i'})(X_i - X_{i'})^T}{\|X_i - X_{i'}\|_2^2} - K \right\|_2 \leq \left\| \frac{(X_i - X_{i'})(X_i - X_{i'})^T}{\|X_i - X_{i'}\|_2^2} \right\|_2 + \|K\|_2 = 1 + \|K\|_2,$$

where in the last equality we use the fact that

$$\left\| \frac{(X_i - X_{i'})(X_i - X_{i'})^T}{\|X_i - X_{i'}\|_2^2} \right\|_2 = \text{Tr}\left( \frac{(X_i - X_{i'})(X_i - X_{i'})^T}{\|X_i - X_{i'}\|_2^2} \right) = 1.$$

Secondly, by simple algebra, we have

$$\|E\{k_{MK}(X_i, X_{i'}) - E_{MK}(X_i, X_{i'})\}^2 \|_2 \leq \|K\|_2 + \|K\|_2^2.$$ 

Accordingly, applying $R_1 = 1 + \|K\|_2$ and $R_2 = \|K\|_2 + \|K\|_2^2$ to Theorem B.2, we have, for any small enough $t$ such that $t \leq (\|K\|_2 + \|K\|_2^2)/1 + \|K\|_2 = \|K\|_2$,

$$P\left( \left\| \frac{1}{n^2} \sum_{i<i'} (X_i - X_{i'})(X_i - X_{i'})^T}{\|X_i - X_{i'}\|_2^2} - K \right\|_2 \geq t \right) \leq d \exp\left( -\frac{3nt^2}{16(\|K\|_2 + \|K\|_2^2)} \right).$$

Setting

$$t = \sqrt{\frac{16}{3} \cdot \frac{(\|K\|_2 + \|K\|_2^2)(\log d + \log(1/\alpha))}{n}} = \|K\|_2 \sqrt{\frac{16}{3} \cdot \frac{(1 + r^*(K))(\log d + \log(1/\alpha))}{n}},$$

we get the desired concentration result.

We then proceed to the proofs of Theorem 3.5 and Corollary 3.2, which exploit the results in Proposition 2.1 and the concentration inequality of the quadratic terms of the Gaussian distribution.

**Proof of Theorem 3.5.** Using Proposition 2.1, the population multivariate Kendall’s tau statistic $K$ has, for $j = 1, \ldots, d$,

$$\lambda_j(K) = E\left( \frac{\lambda_j(\Sigma)Y_j^2}{\lambda_1(\Sigma)Y_1^2 + \cdots + \lambda_d(\Sigma)Y_d^2} \right) = E\left( \frac{Z_j^2}{\sum_{i=1}^d Z_i^2} \right),$$
where \((Z_1, \ldots, Z_d)^T \sim N_d(0, \Lambda)\). Here \(\Lambda\) is a diagonal matrix with \(\Lambda_{jj} = \lambda_j(\Sigma)\). Using Lemma B.8 and the fact that \(0 \leq Z_j^2/(Z_1^2 + \cdots + Z_d^2) \leq 1\), by setting \(t = 4 \log d\), \(A = Z_j^2/\sum_{i=1}^d Z_i^2\), and recalling \(\mathbb{I}(\cdot)\) to be the indicator function, we have

\[
\mathbb{E}A = \mathbb{E}\left( A \mathbb{I}\left( \sum_{i=1}^d Z_i^2 \geq \text{Tr}(\Sigma) - 4\|\Sigma\|_F \sqrt{\log d} \right) \right) + \mathbb{E}\left( A \mathbb{I}\left( \sum_{i=1}^d Z_i^2 < \text{Tr}(\Sigma) - 4\|\Sigma\|_F \sqrt{\log d} \right) \right)
\]

\[
\leq \frac{\lambda_j(\Sigma)}{\text{Tr}(\Sigma) - 4\|\Sigma\|_F \sqrt{\log d}} + \mathbb{P}\left( \sum_{i=1}^d Z_i^2 < \text{Tr}(\Sigma) - 4\|\Sigma\|_F \sqrt{\log d} \right)
\]

\[
\leq \frac{\lambda_j(\Sigma)}{\text{Tr}(\Sigma) - 4\|\Sigma\|_F \sqrt{\log d}} + \frac{1}{d^4}.
\]

Similarly, we have

\[
\mathbb{E}A = \mathbb{E}\left( A \mathbb{I}\left( \sum_{i=1}^d Z_i^2 \leq \text{Tr}(\Sigma) + 4\|\Sigma\|_F \sqrt{\log d} + 8\|\Sigma\|_2 \log d \right) \right)
\]

\[
+ \mathbb{E}\left( A \mathbb{I}\left( \sum_{i=1}^d Z_i^2 > \text{Tr}(\Sigma) + 4\|\Sigma\|_F \sqrt{\log d} + 8\|\Sigma\|_2 \log d \right) \right)
\]

\[
\geq \frac{\lambda_j(\Sigma)}{\text{Tr}(\Sigma) + 4\|\Sigma\|_F \sqrt{\log d} + 8\|\Sigma\|_2 \log d} - \frac{\mathbb{E}Z_j^2 \mathbb{I}\left( \sum_{i=1}^d Z_i^2 > \text{Tr}(\Sigma) + 4\|\Sigma\|_F \sqrt{\log d} + 8\|\Sigma\|_2 \log d \right)}{\text{Tr}(\Sigma) + 4\|\Sigma\|_F \sqrt{\log d} + 8\|\Sigma\|_2 \log d}.
\]

For the above second term, by Cauchy-Swartz inequality, we have

\[
\mathbb{E}Z_j^2 \mathbb{I}\left( \sum_{i=1}^d Z_i^2 > \text{Tr}(\Sigma) + 4\|\Sigma\|_F \sqrt{\log d} + 8\|\Sigma\|_2 \log d \right)
\]

\[
\leq (\mathbb{E}Z_j^2)^{1/2} \cdot \left( \mathbb{P}\left( \sum_{i=1}^d Z_i^2 > \text{Tr}(\Sigma) + 4\|\Sigma\|_F \sqrt{\log d} + 8\|\Sigma\|_2 \log d \right) \right)^{1/2}
\]

\[
\leq \sqrt{3} \lambda_j(\Sigma) \cdot d^{-2}.
\]

This completes the proof. \(\square\)

**Proof of Corollary 3.2.** Noticing that

\[
\|K\|_{\text{max}} = \mathbb{E} \frac{(X_j - \mu_j)^2}{\sum_{i=1}^d (X_i - \mu_i)^2}
\]

where \(\mathbb{E}(X_j - \mu_j)^2 = \|\Sigma\|_{\text{max}}\) and \(\sum_{i=1}^d (X_i - \mu_i)^2 = \sum_{i=1}^d Z_i^2\), the proof is a line-by-line follow of that of Theorem 3.5. \(\square\)

### B.3 Proofs of Results in Section 4

In this section we provide the proof of Theorems 4.1 and 4.2.
B.3.1 Proof of Theorem 4.1

The proof of Theorem 4.1 is via combining the following two lemmas.

**Lemma B.3.** Remind that \( S(X) \) is the self-normalized version of \( X \) defined in (4.4) and \( K := ES(X)S(X)^T \) is the multivariate Kendall’s tau statistic. Suppose that \( X \sim EC_d(\mu, \Sigma, \xi) \) is elliptically distributed. For any \( v \in S^{d-1} \), suppose that

\[
\mathbb{E} \exp \left( t \left[ (v^T S(X))^2 - v^T K v \right] \right) \leq \exp(\eta t^2), \quad \text{for } t \leq c_0 / \sqrt{\eta},
\]

where \( \eta > 0 \) only depends on the eigenvalues of \( \Sigma \) and \( c_0 \) is an absolute constant. We then have, with probability no smaller than \( 1 - 2\alpha \), for large enough \( n \),

\[
\sup_{v \in S^{d-1} \cap B_0(s)} \left| v^T (\hat{K} - K) v \right| \leq 2 \left( 8 \eta \right)^{1/2} \sqrt{s(3 + \log(d/s)) + \log(1/\alpha)} / n.
\]

**Proof.** This is a standard argument for sparse PCA, combined with the Hoeffding’s decoupling trick. We defer the proof to the last section.

The next lemma calculates the exact value of \( \eta \) in Equation (B.3).

**Lemma B.4.** For any \( v = (v_1, \ldots, v_d)^T \in S^{d-1} \), Equation (B.3) holds with

\[
\eta = \sup_{v \in S^{d-1}} 2\|v^T S(X)\|_{\psi_2}^2 + \|K\|_2
\]

and

\[
\sup_{v \in S^{d-1}} \|v^T S(X)\|_{\psi_2} = \sup_{v \in S^{d-1}} \left\| \sum_{i=1}^d v_i \lambda_i^{1/2}(\Sigma) Y_i \right\|_{\psi_2},
\]

where \((Y_1, \ldots, Y_d)^T \sim N_d(0, I_d)\) is standard Gaussian.

**Proof.** The first assertion is a simple consequence of the relationship between the subgaussian and sub-exponential distributions and the property of the sub-exponential distribution (check, for example, Section 5.2 in Vershynin (2010)).

We then focus on the second assertion. Remind that \( S(X) \) is defined as:

\[
S(X) = \frac{X - \bar{X}}{\|X - \bar{X}\|_2}, \quad \frac{X^*}{\|X^*\|_2} = \frac{Z^0}{\|Z^0\|_2},
\]

where \( X^* \sim EC_d(0, \Sigma, \xi^*) \) for some random variable \( \xi^* \geq 0 \) with \( \mathbb{P}(\xi^* = 0) = 0 \) and \( Z^0 \sim N_d(0, \Sigma) \). Here the second equality is due to the fact that the summation of two independently and identically distributed elliptical random vectors are elliptical distributed (see, for example, Lemma
where \( w \) is full rank, there is a one to one map between \( w \) of Theorem 4.1, it is sufficient to bound its all higher moments: distributed with subgaussian norm uniformly bounded by 

\[
\text{Proof of Theorem 4.2.}
\]

\[
\text{B.3.2 Proof of Theorem 4.2}
\]

In this section we focus on the proof of Theorem 4.2. We aim at providing sharp subgaussian constant of \( S(X) \).

\[
\text{Proof of Theorem 4.2. For any } v \in \mathbb{S}^{d-1}, \text{it is enough to show that } v^T S(X) \in \mathbb{R} \text{ is subgaussian distributed with subgaussian norm uniformly bounded by } \sqrt{2} \lambda_1(\Sigma)/(q \lambda_q(\Sigma)). \text{ For notational simplicity, with an abuse of notation, we let } \lambda_i(\Sigma) \text{ be abbreviated as } \lambda_i \text{ for } i = 1, \ldots, q.
\]

For any \( p = 1, 2, \ldots \) and \( v \in \mathbb{S}^{d-1}, \) to show \( v^T S(X) \) is subgaussian, following (B.5) in the proof of Theorem 4.1, it is sufficient to bound its all higher moments:

\[
\mathbb{E}\left( \left\| v^T Z^0 \right\|_2^p \right) = \mathbb{E}\left( \left\| \frac{\sum_{i=1}^q w_i \lambda_i^{1/2} Y_i}{\sqrt{\sum_{i=1}^q \lambda_i Y_i^2}} \right\|_2^p \right) \leq \mathbb{E}\left( \left\| \frac{\sum_{i=1}^q w_i \lambda_i^{1/2} Y_i}{\sqrt{\sum_{i=1}^q \lambda_i Y_i^2}} \right\|_2^p \right),
\]

where \( w = (w_1, \ldots, w_d)^T = U^T v, Y = (Y_1, \ldots, Y_d)^T \sim N_d(0, I_d), \) and we have \( w^T w = 1. \) Because \( U \) is full rank, there is a one to one map between \( v \) and \( w, \) and hence taking supremum over \( v \) is equivalent to taking supremum over \( w. \) This completes the proof. \( \Box \)

B.3.2 Proof of Theorem 4.2

In this section we focus on the proof of Theorem 4.2. We aim at providing sharp subgaussian constant of \( S(X). \)

\[
\text{Proof of Theorem 4.2. For any } v \in \mathbb{S}^{d-1}, \text{it is enough to show that } v^T S(X) \in \mathbb{R} \text{ is subgaussian distributed with subgaussian norm uniformly bounded by } \sqrt{2} \lambda_1(\Sigma)/(q \lambda_q(\Sigma)). \text{ For notational simplicity, with an abuse of notation, we let } \lambda_i(\Sigma) \text{ be abbreviated as } \lambda_i \text{ for } i = 1, \ldots, q.
\]

For any \( p = 1, 2, \ldots \) and \( v \in \mathbb{S}^{d-1}, \) to show \( v^T S(X) \) is subgaussian, following (B.5) in the proof of Theorem 4.1, it is sufficient to bound its all higher moments:

\[
\mathbb{E}\left( \left\| v^T Z^0 \right\|_2^p \right) = \mathbb{E}\left( \left\| \frac{\sum_{i=1}^q w_i \lambda_i^{1/2} Y_i}{\sqrt{\sum_{i=1}^q \lambda_i Y_i^2}} \right\|_2^p \right) \leq \mathbb{E}\left( \left\| \frac{\sum_{i=1}^q w_i \lambda_i^{1/2} Y_i}{\sqrt{\sum_{i=1}^q \lambda_i Y_i^2}} \right\|_2^p \right),
\]

where \( w = (w_1, \ldots, w_d)^T = U^T v, Y = (Y_1, \ldots, Y_d)^T \sim N_d(0, I_d), \) and we have \( w^T w = 1. \) Because \( U \) is full rank, there is a one to one map between \( v \) and \( w, \) and hence taking supremum over \( v \) is equivalent to taking supremum over \( w. \) This completes the proof. \( \Box \)

Next we prove that the rightest term at (B.6) reaches its maximum at \( w^* = (1, 0, \ldots, 0)^T. \) To this end, we adopt a standard technique in calculating the distribution of the quadratic ratio (see, for example, Provost and Cheong (2000)). Let \( \zeta(w) = (\zeta_1, \ldots, \zeta_q) \) with \( \zeta_i = w_i(\lambda_i/\lambda_q)^{1/2}. \) For any constant \( c \geq 0, \) we have

\[
\mathbb{P}\left( \left\| \zeta(w)^T Y \right\|_2 > c \right) = \mathbb{P}\left( \frac{Y^T \zeta(w) \zeta(w)^T Y}{Y^T Y} > c^2 \right) = \mathbb{P}(Y^T (\zeta(w) \zeta(w)^T - c^2 I_q) Y > 0).
\]

Then it is immediate to have

\[
\mathbb{P}(Y^T (\zeta(w) \zeta(w)^T - c^2 I_q) Y > 0) = \mathbb{P}\left( \sum_{j=1}^q l_j Y_j^2 > 0 \right),
\]

1 in Lindskog et al. (2003) for a proof). The third equality holds because \( X^* = d \xi^T Z^0 \) for some random variable \( \xi^T \geq 0 \) with \( \mathbb{P}(\xi^T = 0) = 0. \) Accordingly, we have \( Z^0/\|Z^0\|_2 = S(X) \).
where \( l_1 = \zeta(w)^T \zeta(w) - c^2 \) and \( l_j = -c^2 \) for \( j = 2, \ldots, q \). This implies that

\[
\arg \max_{w \in \mathbb{S}^{d-1}} P(Y^T (\zeta(w)\zeta(w)^T - c^2 I_d) Y > 0) = \arg \max_{w \in \mathbb{S}^{d-1}} \zeta(w)^T \zeta(w) = (1, 0, \ldots, 0)^T = w^*.
\]

In other words, for any \( c > 0 \) and \( w \in \mathbb{S}^{d-1} \), we have

\[
P(|\zeta(w)^T Y| |Y|_2 > c) \geq P(|\zeta(w^*)^T Y| |Y|_2 > c).
\]

Then (B.6) further implies that

\[
\sup_{v \in \mathbb{S}^{d-1}} \mathbb{E} \left( \left| \frac{v^T Z_0}{|Z_0|_2} \right| \right)^p \leq \mathbb{E} \left( \left| \frac{(\lambda_1/\lambda_q)^{1/2} Y_1}{\sqrt{\sum_{i=1}^q Y_i^2}} \right| \right)^p = (\lambda_1/\lambda_q)^{p/2} \cdot \mathbb{E} \left( \frac{Y_1}{\sqrt{\sum_{i=1}^q Y_i^2}} \right)^p.
\]

In the end, combining (B.8) and Lemma B.9, we have

\[
\sup_{v \in \mathbb{S}^{d-1}} \|v^T S(X)\|_{\psi_2} = \sup_{v \in \mathbb{S}^{d-1}} \left\| \frac{v^T Z_0}{|Z_0|_2} \right\|_{\psi_2} \leq \sqrt{\frac{\lambda_1}{\lambda_q}} \frac{2}{q}.
\]

This completes the proof. \( \square \)

### B.4 Proofs of Results in Section 5

In this section we provide the proofs of Theorems 5.2, 5.3, and 5.4.

#### B.4.1 Proof of Theorem 5.2

Theorem 5.2 is the direct consequence of the following lemma by setting \( A = X_m \), \( C = \hat{X}_m \), and \( P = \Pi_m \).

**Lemma B.5.** For any positive semidefinite symmetric matrix \( A \in \mathbb{R}^{d \times d} \) (not necessarily satisfying \( A \preceq I_d \)) and rank \( m \) projection matrix \( P \), letting \( C = \sum_{j=1}^m u_j(A)u_j(A)^T \), we have

\[
\|C - P\|_F \leq 4\|A - P\|_F.
\]

**Proof.** We let \( \epsilon := \|A - P\|_F \). We first define \( B := \sum_{j=1}^m \lambda_j(A)u_j(A)u_j(A)^T \) to be the best rank \( m \) approximation to \( A \). By simple algebra, we have

\[
\|A - B\|_F^2 = \sum_{j>m} (\lambda_j(A))^2,
\]

and accordingly, using triangular inequality,

\[
\|B - P\|_F \leq \|A - B\|_F + \|A - P\|_F = \left( \sum_{j>m} (\lambda_j(A))^2 \right)^{1/2} + \epsilon.
\]

This completes the proof. \( \square \)
Using Lemma B.10 and the fact that $\lambda_j(P) = 0$ for all $j > m$, we further have
\[
\sum_{j > m} (\lambda_j(A))^2 \leq \sum_{j=1}^{d} (\lambda_j(A) - \lambda_j(P))^2 \leq \|A - P\|_F^2 = \epsilon^2,
\]
so that $\|B - P\|_F \leq 2\epsilon$. With a little abuse of notation, for $j = 1, \ldots, d$, we write $\lambda_j = \lambda_j(A)$ and $u_j = u_j(A)$ for simplicity. Therefore, we have
\[
\|B - P\|_F^2 = \left( \sum_{j=1}^{m} \lambda_j u_j u_j^T - P \right)^2 = \sum_{j=1}^{m} \lambda_j^2 + 2 \sum_{j=1}^{m} \lambda_j u_j^T P u_j \leq 4\epsilon^2.
\]
This further implies that
\[
\sum_{j=1}^{m} (\lambda_j^2 + 1) - 4\epsilon^2 \leq 2 \sum_{j=1}^{m} \lambda_j u_j^T P u_j \leq 2 \sum_{j=1}^{m} \lambda_j \Rightarrow \sum_{j=1}^{m} (1 - \lambda_j)^2 \leq 4\epsilon^2.
\]
Noticing that, by the definition of $B$,
\[
\|C - B\|_F^2 = \sum_{j=1}^{m} (1 - \lambda_j)^2 \leq 4\epsilon^2,
\]
we finally have
\[
\|C - P\|_F \leq \|B - P\|_F + \|C - B\|_F \leq 4\epsilon.
\]
This completes the proof. \hfill \Box

B.4.2 Proof of Theorem 5.3

To prove Theorem 5.3, we first provide a general theorem, which quantifies the convergence rate of a U-statistic estimate of the covariance matrix.

**Theorem B.6** (Concentration inequality for U-statistics estimators of covariance matrix). Let $k_1(\cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ and $k_2(\cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be two real functions and $X_1, \ldots, X_n$ be $n$ observations of the random variable $X \in \mathcal{X}$ satisfying $\|k_i(X_1, X_2)\|_{\psi_2} \leq K_0$ for $i = 1, 2$, and $\tau^2 := \mathbb{E}k_1(X_1, X_2)k_2(X_1, X_2)$. We then have,
\[
P\left( \frac{1}{n} \sum_{i < i'} k_1(X_i, X_{i'})k_2(X_i, X_{i'}) - \mathbb{E}k_1(X_i, X_{i'})k_2(X_i, X_{i'}) \geq t \right) \leq \exp\left( -\frac{nt^2}{8C(4K_0^2 + \tau^2)^2} \right),
\]
for all $t < 2Cc(4K_0^2 + \tau^2)$ and some generic constants $C$ and $c$. When the two kernel functions $k_1(\cdot)$ and $k_2(\cdot)$ are equal, the term $4K_0^2$ above can be further relaxed to be $2K_0^2$. 
Proof. First, let’s calculate $\mathbb{E} \exp(\frac{t}{(2^{1/2})} \sum_{i < i'} k_1(X_i, X_{i'})k_2(X_i, X_{i'}) - \tau^2)$. Using a similar decoupling technique as in Theorem B.2, we have

$$
\mathbb{E} \exp\left(\frac{t}{(2^{1/2})} \sum_{i < i'} k_1(X_i, X_{i'})k_2(X_i, X_{i'}) - \mathbb{E}k_1(X_i, X_{i'})k_2(X_i, X_{i'})\right)
= \left(\mathbb{E}e^{\frac{t}{m}(k_1(X_1, X_2)k_2(X_1, X_2) - \mathbb{E}k_1(X_1, X_2)k_2(X_1, X_2))}\right)^m,
$$

where $m := n/2$. We then have

$$
\|k_1(X_1, X_2)k_2(X_1, X_2)\|_{\psi_1} = \left\| \frac{1}{4} \{(k_1(X_1, X_2) + k_2(X_1, X_2))^2 - (k_1(X_1, X_2) - k_2(X_1, X_2))^2\}\right\|_{\psi_1}
\leq \frac{1}{4} \|(k_1(X_1, X_2) + k_2(X_1, X_2))^2\|_{\psi_1} + \frac{1}{4} \|(k_1(X_1, X_2) - k_2(X_1, X_2))^2\|_{\psi_1}.
$$

(B.9)

Using Minkowski’s inequality, for any two random variables $X, Y \in \mathbb{R}$, we have $(\mathbb{E}|X + Y|^p)^{1/p} \leq (\mathbb{E}|X|^p)^{1/p} + (\mathbb{E}|Y|^p)^{1/p}$. Accordingly, we have $\|X + Y\|_{\psi_2} \leq \|X\|_{\psi_2} + \|Y\|_{\psi_2}$. This implies that

$$
\|k_1(X_1, X_2) + k_2(X_1, X_2)\|_{\psi_2} \leq \|k_1(X_1, X_2)\|_{\psi_2} + \|k_2(X_1, X_2)\|_{\psi_2} \leq 2K_0.
$$

Therefore, using the relationship between $\| \cdot \|_{\psi_1}$ and $\| \cdot \|_{\psi_2}$, we have

$$
\|(k_1(X_1, X_2) + k_2(X_1, X_2))^2\|_{\psi_1} \leq 2\|(k_1(X_1, X_2) + k_2(X_1, X_2))\|_{\psi_2}^2 \leq 8K_0^2.
$$

This, combined with (B.9), implies that

$$
\|k_1(X_1, X_2)k_2(X_1, X_2)\|_{\psi_1} \leq 4K_0^2.
$$

(B.10)

Accordingly, $k_1(X_1, X_2)k_2(X_1, X_2) - \mathbb{E}k_1(X_1, X_2)k_2(X_1, X_2)$ is sub-exponential and has sub-exponential norm

$$
\|k_1(X_1, X_2)k_2(X_1, X_2) - \mathbb{E}k_1(X_1, X_2)k_2(X_1, X_2)\|_{\psi_1} \leq 4K_0^2 + \tau^2.
$$

We can then apply Lemma 5.15 in Vershynin (2010) to deduce that

$$
\mathbb{E} \exp\left(\frac{t}{(2^{1/2})} \sum_{i < i'} k_1(X_i, X_{i'})k_2(X_i, X_{i'}) - \tau^2\right) \leq \exp\left(\frac{C(4K_0^2 + \tau^2)^2 t^2}{m}\right), \text{ for } \left|\frac{t}{m}\right| \leq c/(4K_0^2 + \tau^2),
$$

where $C$ and $c$ are two absolute constants. We then use the Markov’s inequality to have the final concentration inequality. This completes the proof of the first part.

Furthermore, when $k_1(\cdot) = k_2(\cdot)$, we can improve the upper bound in (B.10) to be $2K_0^2$. And the whole proof still proceeds. This completes the proof of the second part.

Using Theorem B.6, we are now ready to prove Theorem 5.3.
Proof of Theorem 5.3. Using the result in Theorem 4.2, we have for any \( v \in S^{d-1} \),
\[
\|v^T S(X)\|_{\psi_2} \leq \sqrt{\frac{\lambda_1(\Sigma)}{\lambda_q(\Sigma)} \cdot \frac{2}{q}}.
\]
In particular, for any \( j, k \in \{1, \ldots, d\} \), setting
\[
k_1(X_1, X_2) = \frac{e_j^T (X_1 - X_2)}{\|X_1 - X_2\|_2} \quad \text{and} \quad k_2(X_1, X_2) = \frac{e_k^T (X_1 - X_2)}{\|X_1 - X_2\|_2},
\]
we have
\[
\|k_i(X_1, X_2)\|_{\psi_2} \leq \sqrt{\frac{\lambda_1(\Sigma)}{\lambda_q(\Sigma)} \cdot \frac{2}{q}}, \quad \text{for } i = 1, 2.
\]
Accordingly, using Theorem B.6, we have
\[
\mathbb{P}(|\hat{K}_{jk} - K_{jk}| \geq t) \leq \exp\left(-\frac{nt^2}{8C(8\lambda_1(\Sigma)/q\lambda_q(\Sigma) + K_{jk})^2}\right), \quad \text{for } t < 2Cc(8\lambda_1(\Sigma)/q\lambda_q(\Sigma) + K_{jk}).
\]
We then use the union bound to deduce that
\[
\mathbb{P}(\|\hat{K} - K\|_{\max} \geq t) \leq d^2 \exp\left(-\frac{nt^2}{8C(8\lambda_1(\Sigma)/q\lambda_q(\Sigma) + \|K\|_{\max})^2}\right),
\]
which implies that, for large enough \( n \), with probability larger than \( 1 - \alpha^2 \),
\[
\|\hat{K} - K\|_{\max} \leq 4\sqrt{C} \left(\frac{8\lambda_1(\Sigma)}{q\lambda_q(\Sigma)} + \|K\|_{\max}\right) \sqrt{\frac{\log d + \log(1/\alpha)}{n}}.
\]
This completes the proof. \(\square\)

B.4.3 Proof of Theorem 5.4

Proof of Theorem 5.4. Without loss of generality, we assume that \((u_1(X_1))^T u_1(K) \geq 0\). Using Corollary 5.1, we have
\[
\|u_1(X_1) - u_1(K)\|_\infty \leq \|u_1(X_1) - u_1(K)\|_2 = O_P(s\sqrt{\log d/n}).
\]
It is then immediate that, with high probability, the support set of \(u_1(X_1)\), denoted as \(\hat{J}\), must include \(J_0\) and belong to \(J_1 := \{j : |(u_1(K))_j| = \Omega(s\log d/\sqrt{n})\}\), i.e.,
\[
\mathbb{P}(J_0 \subset \hat{J} \subset J_1) \to 1, \quad \text{when } n \to \infty.
\]
Therefore, with high probability, \(\|v^{(0)}\|_0 \leq \text{card}(J_1) \leq s\). Moreover, we have, for any \( j \in J_0 \), because \(s\sqrt{\log d/n} = o(s\log d/\sqrt{n})\), we have
\[
(u_1(X_1))_j = (u_1(K))_j(1 + o_P(1))
\]
for \( j \in J_0 \). Using the above result, we have

\[
v^{(0)} = \text{TRC}(u_1(K), \hat{J}) / \| \text{TRC}(u_1(K), \hat{J}) \|_2 \cdot (1 + o_P(1)).
\]

Accordingly, under the condition of Theorem 5.4, we have,

\[
(v^{(0)})^T u_1(K) = \| \text{TRC}(u_1(K), \hat{J}) \|_2 (1 + o_P(1)) \geq C_3 (1 + o_P(1)),
\]

and accordingly is asymptotically lower bounded by absolute constant. The rest can be proved by using Theorem 4 in Yuan and Zhang (2013).

\[\square\]

### B.5 Auxiliary Lemmas

In this section we provide the auxiliary lemmas. The first lemma shows that any elliptical distribution is a random scaled version of the Gaussian.

**Lemma B.7.** Let \( X \sim EC_d(\mu, \Sigma, \xi) \) be an elliptical distribution with \( \Sigma = AA^T \). It takes another stochastic representation:

\[
X \overset{\text{d}}{=} \mu + \xi Z / \| A^\dagger Z \|_2,
\]

where \( Z \sim N_d(0, \Sigma) \), \( \xi \geq 0 \) is independent of \( Z / \| A^\dagger Z \|_2 \), and \( A^\dagger \) is the Moore-Penrose pseudoinverse of \( A \).

**Proof.** Let \( X = \mu + \xi AU \) and \( q := \text{rank}(\Sigma) = \text{rank}(A) \) as in (2.1). Let \( U = \epsilon / \| \epsilon \|_2 \) with a standard normal vector \( \epsilon \) in \( \mathbb{R}^q \). Note that if \( A = V_1DV_2^T \) is the singular value decomposition of \( A \in \mathbb{R}^{d \times q} \) with \( V_1 \in \mathbb{R}^{d \times q} \) and \( D, V \in \mathbb{R}^{q \times q} \), then \( A^\dagger = V_2D^{-1}V_1^T \). Since \( \text{rank}(A) = q \), we have \( A^\dagger A = I_q \). Accordingly, let \( Z = Ae \sim N(0, \Sigma) \). It follows that

\[
X - \mu = \xi AU = \xi Z / \| \epsilon \|_2 = \xi Z / \| A^\dagger Z \|_2.
\]

The proof is complete. \[\square\]

The next lemma gives two Hanson-Wright type inequalities for the quadratic term of the Gaussian distributed random vectors.

**Lemma B.8.** Let \( Z \sim N_d(0, \Sigma) \) be a \( d \)-dimensional Gaussian distributed random vector. Then for every \( t \geq 0 \), we have

\[
\mathbb{P}(Z^T Z - \text{Tr}(\Sigma) \leq -2\| \Sigma \|_F \sqrt{t}) \leq \exp(-t),
\]

and

\[
\mathbb{P}(Z^T Z - \text{Tr}(\Sigma) \geq 2\| \Sigma \|_F \sqrt{t} + 2\| \Sigma \|_2 t) \leq \exp(-t).
\]
Proof. Let $U \Lambda U^T$ be the SVD decomposition of $\Sigma$. Then letting $Y = (Y_1, \ldots, Y_d)^T \sim N_d(0, I_d)$, using the fact that $U^TY \overset{d}{=} Y$, we have

$$Z^TZ \overset{d}{=} Y^T \Sigma Y = Y^T U \Lambda U^TY \overset{d}{=} Y^T \Lambda Y = \sum \lambda_j(\Sigma)Y_j^2.$$  

The rest follows from Lemma 1 in Laurent and Massart (2000). □

The next lemma shows that a simple version of the quadratic ratio under the Gaussian assumption is subgaussian.

Lemma B.9. For $Y = (Y_1, \ldots, Y_q)^T \sim N_q(0, I_q)$, we have

$$\left\| \frac{Y_j}{\sqrt{\sum_{i=1}^q Y_i^2}} \right\|_{\psi_2} \leq \sqrt{\frac{2}{q}}, \quad \text{for } j = 1, \ldots, d,$$

where we remind that the subgaussian norm $\| \cdot \|_{\psi_2}$ is defined in (4.3).

Proof. It is known (see, for example, Chapter 3 in Bilodeau and Brenner (1999)) that $Y_1^2/\sum_{i=1}^q Y_i^2 \sim \text{Beta}(\frac{1}{2}, \frac{q-1}{2})$.

Accordingly, using the Jensen’s inequality and the property of the beta distribution, we have

$$\mathbb{E}\left( \left( \frac{Y_1^2}{\sum_{i=1}^q Y_i^2} \right)^{p/2} \right) \leq \left( \mathbb{E}\left( \frac{Y_1^2}{\sum_{i=1}^q Y_i^2} \right)^p \right)^{1/2} = \left( \prod_{r=0}^{p-1} \frac{2r+1}{2r+q} \right)^{1/2},$$

where the last equality is using the moment formula of the beta distribution (check, for example, Page 36 in Gupta and Nadarajah (2004)). When $q$ is even, using the Sterling’s inequality, we can continue to write

$$\left( \prod_{r=0}^{p-1} \frac{2r+1}{2r+q} \right)^{1/2} = \left( \frac{(2p-1)!!}{(2p+q-2)!!} \right)^{1/2} \leq \left( \frac{p!}{(p+(q-2)/2)!} \right)^{1/2} \leq \left( \frac{p^{p+1/2}e^{(q-2)/2}}{(p+(q-2)/2)^{p+(q-1)/2}} \right)^{1/2} \leq p^{p/2}/(p+(q-2)/2)^{p/2}.$$  

Accordingly, we have

$$\left( \mathbb{E}\left( \frac{Y_1^2}{\sum_{i=1}^q Y_i^2} \right)^{p/2} \right)^{1/p} \leq \sqrt{p} \cdot \sqrt{\frac{T}{q}},$$

Similarly, when $q$ is odd, we have $\left( \mathbb{E}(Y_1^2/\sum_{i=1}^q Y_i^2)^{p/2} \right)^{1/p} \leq \sqrt{p} \cdot \kappa_U \cdot 1/(q-1/2)^{1/2} \leq \sqrt{p} \cdot \sqrt{2}/q$. This completes the proof. □
The final lemma states a Wyel type inequality and is well known in the matrix perturbation literature (check, for example, Equation (3.3.32) in Horn and Johnson (1991)).

**Lemma B.10.** For any positive semidefinite symmetric matrices $A, B \in \mathbb{R}^{d \times d}$ (so that the eigenvalues and singular values are equal), we have

$$
\sum_{i=1}^{d} (\lambda_i(A) - \lambda_i(B))^2 \leq \|A - B\|_F^2.
$$

### C. The Proof of Lemma B.3

The proof of Lemma B.3 is shown in this section. The idea is to combine the proof of sparse PCA (see, for example, Lounici (2013)) with the Hoeffding’s decoupling trick. We present the proof here mainly for completedness.

**Proof of Lemma B.3.** Let $a \in \mathbb{Z}^+$ be an integer no smaller than 1 and $J_a$ be any subset of $\{1, \ldots, d\}$ with cardinality $a$. For any $s$-dimensional sphere $\mathbb{S}^{s-1}$ equipped with Euclidean distance, we let $\mathcal{N}_\varepsilon$ be a subset of $\mathbb{S}^{s-1}$ such that for any $v \in \mathbb{S}^{s-1}$, there exists $u \in \mathcal{N}_\varepsilon$ subject to $\|u - v\|_2 \leq \varepsilon$. It is known that the cardinal number of $\mathcal{N}_\varepsilon$ has an upper bound: $\text{card}(\mathcal{N}_\varepsilon) < (1 + \frac{2}{s})^s$. Let $\mathcal{N}_{1/4}$ be a $(1/4)$-net of $\mathbb{S}^{s-1}$. We then have $\text{card}(\mathcal{N}_{1/4})$ is upper bounded by $9^s$. Moreover, for any symmetric matrix $M \in \mathbb{R}^{s \times s}$, we have

$$
\sup_{v \in \mathbb{S}^{s-1}} |v^T M v| \leq \frac{1}{1 - 2\varepsilon} \sup_{v \in \mathcal{N}_\varepsilon} |v^T M v|, \text{ implying } \sup_{v \in \mathbb{S}^{s-1}} |v^T M v| \leq 2 \sup_{v \in \mathcal{N}_{1/4}} |v^T M v|.
$$

Let $\beta > 0$ be a quantity defined as $\beta := (8\eta)^{1/2} \sqrt{\frac{s(3 + \log(d/s)) + \log(1/\alpha)}{n}}$. By the union bound, we have

$$
P\left( \sup_{b \in \mathbb{S}^{s-1}} \sup_{J_a \subset \{1, \ldots, d\}} \left| b^T \begin{bmatrix} \hat{K} - K \end{bmatrix}_{J_a, J_a} b \right| > 2\beta \right) \leq P\left( \sup_{b \in \mathcal{N}_{1/4}} \sup_{J_a \subset \{1, \ldots, d\}} \left| b^T \begin{bmatrix} \hat{K} - K \end{bmatrix}_{J_a, J_a} b \right| > \beta \right)
$$

$$
\leq 9^s \left( \frac{d}{s} \right) P\left( \left| b^T \begin{bmatrix} \hat{K} - K \end{bmatrix}_{J_s, J_s} b \right| > (8\eta)^{1/2} \sqrt{\frac{s(3 + \log(d/s)) + \log(1/\alpha)}{n}} \right), \text{ for any } b \text{ and } J_s.
$$

Thus, if we can show that for any $b \in \mathbb{S}^{s-1}$ and $J_s$, we have

$$
P\left( \left| b^T \begin{bmatrix} \hat{K} - K \end{bmatrix}_{J_s, J_s} b \right| > t \right) \leq 2e^{-nt^2/(8\eta)}, \tag{C.1}
$$

for $\eta$ defined in Equation (B.3). Then, using the bound $\left( \frac{d}{s} \right) < (ed/s)^s$, we have

$$
9^s \left( \frac{d}{s} \right) P\left( \left| b^T \begin{bmatrix} \hat{K} - K \end{bmatrix}_{J_s, J_s} b \right| > (8\eta)^{1/2} \sqrt{\frac{s(3 + \log(d/s)) + \log(1/\alpha)}{n}} \right), \text{ for any } b \text{ and } J
$$

$$
\leq 2 \exp\{s(1 + \log 9 - \log(s)) + s \log d - s(3 + \log d - \log s) - \log(1/\alpha)\} \leq 2\alpha.
$$
It shows that, with probability greater than $1 - 2\alpha$, the bound in Equation (B.4) holds.

We now show that Equation (C.1) holds. For any $t$, we have

$$
\mathbb{E} \exp \left( t \cdot b^T \left[ \hat{K} - K \right]_{J_s, J_s} b \right) = \mathbb{E} \exp \left( t \cdot \frac{1}{2} \sum_{i < i'} b^T \left( \frac{(X_i - X_{i'})_{J_s} (X_i - X_{i'})_{J_s}^T}{\|X_i - X_{i'}\|^2} - K_{J_s, J_s} \right) b \right).
$$

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 $O(\sigma) := \{(i_1, i_2), (i_3, i_4), \ldots, (i_{n-1}, i_n)\}$. In particular, we denote $O(\sigma_0) := \{(1, 2), (3, 4), \ldots, (n-1, n)\}$. By simple calculation,

$$
\mathbb{E} \exp \left( t \cdot \frac{1}{\text{card}(S_n)} \sum_{\sigma \in S_n} \frac{2}{n} \sum_{(i, i') \in O(\sigma)} b^T \left( \frac{(X_i - X_{i'})_{J_s} (X_i - X_{i'})_{J_s}^T}{\|X_i - X_{i'}\|^2} - K_{J_s, J_s} \right) b \right)
\leq \mathbb{E} \exp \left( t \cdot \frac{2}{n} \sum_{(i, i') \in O(\sigma_0)} b^T \left( \frac{(X_i - X_{i'})_{J_s} (X_i - X_{i'})_{J_s}^T}{\|X_i - X_{i'}\|^2} - K_{J_s, J_s} \right) b \right),
$$

where the inequality is due to the Jensen’s inequality.

Let $m := n/2$ and recall that $X = (X_1, \ldots, X_d)^T \in \mathcal{M}_d(\Sigma, \xi, q; \kappa_L, \kappa_U)$. Letting $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_d)^T$ be an independent copy of $X$, by Equation (B.3), we have that, for any $t \in \mathbb{R}$ and $v \in \mathbb{S}^{d-1}$,

$$
\mathbb{E} \exp \left( t \cdot v^T \left( \frac{(X - \tilde{X})(X - \tilde{X})}{\|X - \tilde{X}\|^2} - K \right) v \right) \leq e^{\eta t^2}.
$$

In particular, letting $v_{J_s} = b$ and $v_{J_s^c} = 0$, we have

$$
\mathbb{E} \exp \left( t \cdot b^T \left( \frac{(X - \tilde{X})_{J_s} (X - \tilde{X})_{J_s}^T}{\|X - \tilde{X}\|^2} - K_{J_s, J_s} \right) b \right) \leq e^{\eta t^2}.
$$

Then we are able to continue Equation (C.2) as

$$
\mathbb{E} \exp \left( t \cdot \frac{2}{n} \sum_{(i, i') \in O(\sigma_0)} b^T \left( \frac{(X_i - X_{i'})_{J_s} (X_i - X_{i'})_{J_s}^T}{\|X_i - X_{i'}\|^2} - K_{J_s, J_s} \right) b \right)
\leq \mathbb{E} \exp \left( t \cdot \frac{2m}{m} b^T \left( \frac{(X_{2i} - X_{2i-1})_{J_s} (X_{2i} - X_{2i-1})_{J_s}^T}{\|X_{2i} - X_{2i-1}\|^2} - K_{J_s, J_s} \right) b \right)
\leq \left( \mathbb{E} e^{\frac{1}{m}(b^T S(X)^2 - b^T K_{J_s, J_s} b)} \right)^m \leq e^{\eta t^2/m},
$$

(4)
where by Equation (B.3), the last inequality holds for any $|t/m| \leq c/\sqrt{\eta}$. Accordingly, choosing $t = \beta m/(2\eta)$, using the Markov inequality, we have

$$P\left( b^T \left[ \hat{K} - K \right]_{J_s,J_s} b > \beta \right) \leq e^{-n\beta^2/(8\eta)}, \quad \text{for } \beta \leq 2c_0\sqrt{\eta}. \tag{C.5}$$

By symmetry, we have the same bound for $P\left( b^T \left[ \hat{K} - K \right]_{J_s,J_s} b < -\beta \right)$ as in Equation (C.5). Together, they give us Equation (C.1). This completes the proof.

References

Bilodeau, M. and Brenner, D. (1999). *Theory of Multivariate Statistics*. Springer.

Croux, C., Ollila, E., and Oja, H. (2002). Sign and rank covariance matrices: statistical properties and application to principal components analysis. In *Statistics for Industry and Technology*, pages 257–69. Birkhauser.

Gupta, A. K. and Nadarajah, S. (2004). *Handbook of Beta Distribution and Its Applications*. CRC Press.

Han, F. and Liu, H. (2016). Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution. *Bernoulli (in press).*

Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis*. Cambridge University Press.

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338.

Lindskog, F., McNeil, A., and Schmock, U. (2003). Kendall’s tau for elliptical distributions. *Credit Risk–Measurement, Evaluation and Management*, pages 149—156.

Lounici, K. (2013). Sparse principal component analysis with missing observations. In *High Dimensional Probability VI*, pages 327–356. Springer.

Marden, J. (1999). Some robust estimates of principal components. *Statistics and Probability Letters*, 43(4):349–359.

Provost, S. B. and Cheong, Y.-H. (2000). On the distribution of linear combinations of the components of a Dirichlet random vector. *Canadian Journal of Statistics*, 28(2):417–425.

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing*, pages 210–268. Cambridge University Press.
Wegkamp, M. and Zhao, Y. (2016). Adaptive estimation of the copula correlation matrix for semiparametric elliptical copulas. *Bernoulli*, 22(2):1184–1226.

Yuan, X. and Zhang, T. (2013). Truncated power method for sparse eigenvalue problems. *Journal of Machine Learning Research*, 14:899–925.