
**High-dimensional semiparametric estimate
of latent covariance matrix for matrix-variates: Supplementary Material**

Lu Niu¹ AND JunLong Zhao²

1. School of Mathematics and System Science, Beihang University, China

2. School of Statistics, Beijing Normal University, China

The Supplementary Material consists of two parts. The first part contains the proof of Proposition 2, Lemma 1, Theorem 1, Theorem 2 and Theorem 3 in the main text. The second part contains the simulation results for the case of $(s_1, s_2) = (\lceil p/4 \rceil, \lceil q/4 \rceil)$ for Example 1–4 in Section 4, and the Figure 1–3 for real data analysis in Section 5.

S1 Proofs

S1.1 Proof of Proposition 2

Recall that $L_{kk'} = Z_k - Z_{k'}$ and

$$V_{kk'} = \text{vec}(\text{sign}(L_{kk'})\text{sign}(L_{kk'}) - E[\text{sign}(L_{kk'})\text{sign}(L_{kk'})]).$$

For any \mathbf{a} with $\|\mathbf{a}\| = 1$, we see that

$$\mathbf{a}^\top E_n = 2(n(n-1))^{-1} \sum_{1 \leq k \neq k' \leq n} \mathbf{a}^\top V_{kk'}.$$

Note that $E(V_{kk'}) = 0$. For any $t > 0$ and any $\lambda > 0$, we have

$$\begin{aligned} P(\sqrt{n}\mathbf{a}^\top E_n > t) &= P(\mathbf{a}^\top E_n > \frac{t}{\sqrt{n}}) \\ &= P(e^{\lambda \mathbf{a}^\top E_n} > e^{\lambda t / \sqrt{n}}) \\ &\leq e^{-\lambda t} E(e^{\lambda \sqrt{n} \mathbf{a}^\top E_n}). \end{aligned}$$

By the assumption and the Lemma C of Serfling (2009) (see Page 200), as

$0 < \lambda < 2^{-1}t_0n^{1/2}$, it holds that

$$E(e^{\lambda \sqrt{n} \mathbf{a}^\top E_n}) \leq [E(e^{(2/n)\lambda \sqrt{n} \mathbf{a}^\top V_{kk'}})]^{n/2} \leq e^{2c\lambda^2}.$$

Then, it holds that $P(\sqrt{n}\mathbf{a}^\top E_n > t) \leq \exp(-\lambda t + 2c\lambda^2)$. Minimizing the right hand side over λ with the constraint $\lambda > 0$, we have

$$\min_{\lambda > 0} (-\lambda t + 2c\lambda^2) = -(8c)^{-1}t^2,$$

where the minimum value is arrived at $\lambda = (4c)^{-1}t$, which is smaller than $2^{-1}t_0n^{1/2}$ under the assumption $t = o(n^{1/2})$. That is, $P(\sqrt{n}\mathbf{a}^\top E_n > t) \leq$

$e^{-(8c)^{-1}t^2}$ for any $t > 0$. This completes the proof. ■

Next, we show the conditions in Proposition 2 hold if Z_k is sign sub-Gaussian.

Lemma S.1 *Suppose that $\text{sign}(Z_k - Z_{k'})$ is sub-Gaussian with $\|\text{sign}(Z_k - Z_{k'})\|_{\psi_2} < K < \infty$ for some constant $K > 0$. Then the condition of Proposition 2 holds, that is, $E(\exp(\mathbf{t}\mathbf{a}^\top V_{kk'})) \leq e^{ct^2}$ for any $0 < t < t_0$, where $t_0 > 0$ and $c > 0$ are constants.*

Proof. Let $Y = \text{sign}(Z_k - Z_{k'})$. Recall that $V_{kk'} = \text{vec}(YY^\top - E(YY^\top))$. Denote $M_{\mathbf{a}} \in R^{pq \times pq}$ such that $\text{vec}(M_{\mathbf{a}}) = \mathbf{a}$. Let $\tilde{M}_{\mathbf{a}} = (M_{\mathbf{a}} + M_{\mathbf{a}}^\top)/2$ which is a symmetric matrix but may not be positive. Due to the fact that $\|M_{\mathbf{a}}\|_F = \|\mathbf{a}\| = 1$, it follows that $\|\tilde{M}_{\mathbf{a}}\|_F \leq 1$. Without loss of generality, by assuming that the first r eigenvalues of $\tilde{M}_{\mathbf{a}}$ are positive and rest eigenvalues are negative, we denote its eigenvalue decomposition as

$$\tilde{M}_{\mathbf{a}} = \sum_{i=1}^r \lambda_i u_i u_i^\top - \sum_{i=r+1}^{pq} \lambda_i u_i u_i^\top := U_1 \Lambda_1 U_1^\top - U_2 \Lambda_2 U_2^\top,$$

where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$ is a diagonal matrix are the positive eigenvalues, and $-\Lambda_2 = \text{diag}(-\lambda_{r+1}, \dots, -\lambda_{pq})$ are the negative ones. In addition, $U_1 = [u_1, \dots, u_r]$ and $U_2 = [u_{r+1}, \dots, u_{pq}]$ are the eigenvectors associated with positive and negative eigenvalues, respectively. Let $A_1 = \Lambda_1^{1/2} U_1^\top$ and $A_2 =$

$\Lambda_2^{1/2} U_2^T$. Then $\tilde{M}_{\mathbf{a}} = A_1^T A_1 - A_2^T A_2$.

Due to the symmetric of $YY^T - E(YY^T)$, we can rewrite $\mathbf{a}^T V_{kk'}$ as the difference of two quadratic forms

$$\begin{aligned}
\mathbf{a}^T V_{kk'} &= \langle \tilde{M}_{\mathbf{a}}, YY^T - E(YY^T) \rangle \\
&= Y^T \tilde{M}_{\mathbf{a}} Y - E(Y^T \tilde{M}_{\mathbf{a}} Y) \\
&= [Y^T A_1^T A_1 Y - E(Y^T A_1^T A_1 Y)] - [Y^T A_2^T A_2 Y - E(Y^T A_2^T A_2 Y)] \\
&= (\|A_1 Y\|^2 - E(\|A_1 Y\|^2)) - (\|A_2 Y\|^2 - E(\|A_2 Y\|^2)) \\
&:= \mathbf{B}_1 - \mathbf{B}_2. \tag{S1.1}
\end{aligned}$$

Consider \mathbf{B}_1 first. Recall that Y is a sub-Gaussian variable. Then $E(\exp(\alpha^T Y)) \leq \exp(\|Y\|_{\psi_2}^2 \|\alpha\|^2)$ for any vector $\alpha \in R^{pq}$. Moreover, for any $\eta \in R^{pa}$, we have $\|\eta^T A_1\|^2 \leq \|\eta\|^2$ by the fact that $\sum_i \lambda_i^2 = 1$. Then, it follows that

$$E(\exp(\eta^T A_1 Y)) \leq \exp(\|Y\|_{\psi_2}^2 \|\eta^T A_1\|^2) \leq \exp(\|Y\|_{\psi_2}^2 \|\eta\|^2),$$

which implies that $A_1 Y$ is sub-Gaussian and $\|A_1 Y\|_{\psi_2} \leq \|Y\|_{\psi_2}$. By Lemma 5.14 and Remark 5.18 of Vershynin (2015), we see that $\|A_1 Y\|^2$ and, consequently, $\|A_1 Y\|^2 - E(\|A_1 Y\|^2)$ are sub-Exponential variables with

$$\| \|A_1 Y\|^2 - E(\|A_1 Y\|^2) \|_{\psi_1} \leq 2(\|A_1 Y\|^2)_{\psi_1} \leq 4\|A_1 Y\|_{\psi_2}^2 \leq 4\|Y\|_{\psi_2}^2 = 4K^2.$$

Therefore, by Lemma 5.15 of Vershynin (2015), for $0 < t < c_1/(2\|A_1Y\|^2 - E(\|A_1Y\|^2))\|_{\psi_1} := t_0$ with $c_1 > 0$ being a constant, it holds that

$$\begin{aligned} E(\exp(2t\mathbf{B}_1)) &= E(\exp\{2t(\|A_1Y\|^2 - E(\|A_1Y\|^2))\}) \\ &\leq \exp\left(4t^2\|\|A_1Y\|^2 - E(\|A_1Y\|^2)\|_{\psi_1}^2\right) \\ &\leq \exp(t^2 \cdot 64K^4). \end{aligned} \tag{S1.2}$$

Similarly, we have $E\{\exp[2t(-\mathbf{B}_2)]\} \leq \exp(t^2 \cdot 64K^4)$. Then by Jensen's inequality, we have

$$E(\exp(\mathbf{a}^T V_{kk'})) \leq \frac{1}{2} [E(\exp(2t\mathbf{B}_1)) + E\{\exp[2t(-\mathbf{B}_2)]\}] \leq \exp(t^2 \cdot 64K^4).$$

Taking $c = 64K^4$ in the inequality of Proposition 2, we complete the proof.

■

S1.2 Proof of Lemma 1

Recall the definition that $X_{ij,k}^{(0)} = c_{ij}|X_{ij,k} - \xi_{ij,0.5}|$, and the fact that $\xi_{ij,0.5}$ and σ_{ij} are the 0.5 quantile of the distribution of $X_{ij,k}$ and $X_{ij,k}^{(0)}$, respectively.

To simplify the argument, we introduce some notations first. For any $\epsilon > 0$,

define $\phi_{ij,1} = P(X_{ij,k} > \xi_{ij,0.5} + \epsilon)$, $\phi_{ij,2} = P(X_{ij,k} < \xi_{ij,0.5} - \epsilon)$, and

$\phi_\epsilon = \min_{s=1,2, i \in [p], j \in [q]} |\phi_{ij,s} - 1/2|$. Moreover, define $\theta_{ij,1} = P(X_{ij,k}^{(0)} > \sigma_{ij} + \epsilon)$,

$\theta_{ij,2} = P(X_{ij,k}^{(0)} < \sigma_{ij} - \epsilon)$, and $\theta_\epsilon = \min_{s=1,2,i \in [p], j \in [q]} |\theta_{ij,s} - 1/2|$. We first point out the following simple fact. Recalling that $\xi_{ij,0.5}$ and σ_{ij} are the 0.5 quantile of the distribution of $X_{ij,k}$ and $X_{ij,k}^{(0)}$, respectively, as $\epsilon \rightarrow 0$, we have $|\phi_{ij,s} - 1/2|/\epsilon \rightarrow f_{ij}(\xi_{ij,0.5})$ and $|\theta_{ij,s} - 1/2|/\epsilon \rightarrow g_{ij}(\sigma_{ij})$. By the assumption (A2), we see that as $\epsilon \rightarrow 0$, we have $\min\{\phi_\epsilon, \theta_\epsilon\} > c_0\epsilon$.

Step 1. We show that $\sup_{i,j} |X_{ij}^{med} - \xi_{ij,0.5}| = O_p(\sqrt{(\log p + \log q)/n})$. It holds that

$$P(X_{ij,k}^{med} - \xi_{ij,0.5} > \epsilon) = P(\text{at least } \frac{n+1}{2} \text{ of } X_{ij,k} \text{'s exceeds } \xi_{ij,0.5} + \epsilon)$$

Then, we have, for $\eta_{ij,k} = I(X_{ij,k} > \xi_{ij,0.5} + \epsilon)$,

$$\begin{aligned} P(X_{ij,k}^{med} - \xi_{ij,0.5} > \epsilon) &= P\left(\sum_{k=1}^n \eta_{ij,k} > (n+1)/2\right) \\ &= P\left(n^{-1} \sum_{k=1}^n (\eta_{ij,k} - \phi_{ij,1}) > \frac{1}{2} - \phi_{ij,1} + \frac{1}{2n}\right) \\ &\leq P\left(n^{-1} \sum_{k=1}^n (\eta_{ij,k} - \phi_{ij,1}) > \frac{1}{2} - \phi_{ij,1}\right) \\ &\stackrel{(i)}{\leq} \exp(-2n(1/2 - \phi_{ij,1})^2) \\ &\leq \exp(-2n\phi_\epsilon^2) \end{aligned} \tag{S1.3}$$

where the inequality (i) is derived from the Hoeffding inequality. Similarly, we have $P(X_{ij,k}^{med} < \xi_{ij,0.5} - \epsilon) \leq \exp(-2n(\frac{1}{2} - \phi_{ij,2})^2) \leq \exp(-2n(\frac{1}{2} -$

$\phi_{ij,2}^2) \leq \exp(-2n\phi_\epsilon^2)$. Combining them together, we have $P(|X_{ij}^{med} - \xi_{ij,0.5}| > \epsilon) \leq 2 \exp(-2n\phi_\epsilon^2)$. For any $\delta > 0$, we take $\epsilon_n = \sqrt{C(\log p + \log q)/n}$ for some $C = C(\delta) > 0$ being large. Note that $\epsilon_n \rightarrow 0$ under our assumption, as $n \rightarrow \infty$, and recall that $\phi_{\epsilon_n} > c_0\epsilon_n$. We have

$$P(|X_{ij}^{med} - \xi_{ij,0.5}| > \epsilon_n) \leq 2 \exp(-2c_0^2 C(\log(p) + \log q)).$$

And consequently, it holds that

$$\begin{aligned} P\left(\max_{i \in [p], j \in [q]} |X_{ij}^{med} - \xi_{ij,0.5}| > \epsilon_n\right) &\leq 2pq \exp(-2c_0^2 C(\log(p) + \log q)) \\ &\leq 2 \exp(-(2c_0^2 C - 1)[\log(p) + \log q]) \\ &\leq \delta. \end{aligned} \tag{S1.4}$$

Step 2. Let $\tilde{\sigma}_{ij} = c_{ij} \cdot \text{median}\{|X_{ij,k} - \xi_{ij,0.5}|, k \in [n]\} = \text{median}\{X_{ij,k}^{(0)}, k \in [n]\}$, for $i \in [p], j \in [q]$. We show that $\sup_{ij} |\tilde{\sigma}_{ij} - \sigma_{ij}| = O_p(\sqrt{(\log p + \log q)/n})$.

The proof is similar to step 1. Let $\gamma_{k,ij} = I(c_{ij}|X_{k,ij} - \xi_{0.5,ij}| > \sigma_{ij} + \epsilon)$. Then by the definition of $\tilde{\sigma}_{ij}$, we have $P(\tilde{\sigma}_{ij} - \sigma_{ij} > \epsilon) = P\left(n^{-1} \sum_{k=1}^n \gamma_{ij,k} > (n+1)/2\right)$.

Similar to the argument of Step 1, we have $P(\tilde{\sigma}_{ij} - \sigma_{ij} > \epsilon) \leq 2 \exp(-2n\theta_\epsilon^2)$.

Similarly, we have $P(\tilde{\sigma}_{ij} < \sigma_{ij} - \epsilon) \leq 2 \exp(-2n\theta_\epsilon^2)$. Combining them together, we prove the conclusion in Step 2, taking the same ϵ_n as in Step 1 and using the same argument, as $n \rightarrow \infty$.

Step 3. It is easy to check that for constants $a_1 \leq a_2 \leq \cdots \leq a_n$ and $a_i \leq b_i$, $1 \leq i \leq n$, it holds that $\text{median}\{a_1, \cdots, a_n\} \leq \text{median}\{b_1, \cdots, b_n\}$.

Then, for any $(i, j) \in [p] \times [q]$, we have

$$\begin{aligned} \hat{\sigma}_{ij} &= \text{median}\{c_{ij}|X_{ij,k} - X_{ij}^{med}|, k \in [n]\} \\ &\leq \text{median}\{X_{ij,k}^{(0)}, k \in [n]\} + |X_{ij}^{med} - \xi_{ij,0.5}| \\ &= \tilde{\sigma}_{ij} + |X_{ij}^{med} - \xi_{ij,0.5}| \end{aligned}$$

Similarly, we have $\tilde{\sigma}_{ij} \leq \hat{\sigma}_{ij} + |X_{ij}^{med} - \xi_{ij,0.5}|$. Combined together, it follows that

$$\max_{i,j} |\hat{\sigma}_{ij} - \tilde{\sigma}_{ij}| \leq \max_{i,j} |X_{ij}^{med} - \xi_{ij,0.5}|.$$

Consequently, we have

$$\max_{i,j} |\hat{\sigma}_{ij} - \sigma_{ij}| \leq \max_{i,j} |\hat{\sigma}_{ij} - \tilde{\sigma}_{ij}| + \max_{i,j} |\tilde{\sigma}_{ij} - \sigma_{ij}| \leq \max_{i,j} |X_{ij}^{med} - \xi_{ij,0.5}| + \max_{i,j} |\tilde{\sigma}_{ij} - \sigma_{ij}|.$$

The final conclusion is derived by combining the conclusion in Step 1 and 2.

This completes the proof. ■

S1.3 Proof of Theorem 1

Step 1 We first show that $\|\mathcal{T}(\hat{\mathbf{T}} - \mathbf{T})\|_{op} = O_p(\sqrt{(p^2 + q^2 + \log n)/n})$.

Let $\Delta = \{(u, v) : u \in \mathbb{S}^{q^2-1}, v \in \mathbb{S}^{p^2-1}\}$. For any $(u, v) \in \Delta$, define

$M_u \in \mathbb{R}^{q \times q}$ and $M_v \in \mathbb{R}^{p \times p}$ such that $\text{vec}(M_u^\top) = u$ and $\text{vec}(M_v) = v$. For any two matrices Q_1 and Q_2 with the same dimension, we define $\langle Q_1, Q_2 \rangle = \text{trace}(Q_1^\top Q_2)$. Note that $\|\mathcal{T}(\hat{\mathbf{T}} - \mathbf{T})\|_{op} = \sup_{(u,v) \in \Delta} u^\top \mathcal{T}(\hat{\mathbf{T}} - \mathbf{T})v$. In addition, $u^\top \mathcal{T}(\hat{\mathbf{T}} - \mathbf{T})v$ can be written as $u^\top \mathcal{T}(\hat{\mathbf{T}} - \mathbf{T})v = \langle \hat{\mathbf{T}} - \mathbf{T}, M_u \otimes M_v \rangle = \langle E_n, \text{vec}(M_u \otimes M_v) \rangle$, where $E_n = \text{vec}(\hat{\mathbf{T}} - \mathbf{T})$ mentioned in Section 3.1.

Therefore, we have

$$\|\mathcal{T}(\hat{\mathbf{T}} - \mathbf{T})\|_{op} = \sup_{u,v \in \Delta} \langle E_n, \text{vec}(M_u \otimes M_v) \rangle.$$

We first consider the probability of $P(\langle E_n, \text{vec}(M_u \otimes M_v) \rangle > t)$ for any fixed $(u, v) \in \Delta$. By the assumption (A1), we have for any fixed $u \in \mathbb{S}^{q^2-1}, v \in \mathbb{S}^{p^2-1}$

$$P\left(\langle E_n, \text{vec}(M_u \otimes M_v) \rangle > \frac{t}{\sqrt{n}}\right) \leq C \exp\left(-\frac{t^2}{K}\right).$$

Let $N(\mathbb{S}^{d-1}, \epsilon)$ be the ϵ -net of the sphere \mathbb{S}^{d-1} . For any $u \in \mathbb{S}^{q^2-1}$ and $v \in \mathbb{S}^{p^2-1}$, we can find $u_1 \in N(\mathbb{S}^{q^2-1}, \epsilon)$ and $v_1 \in N(\mathbb{S}^{p^2-1}, \epsilon)$, such that $\|u - u_1\| < \epsilon$, $\|v - v_1\| < \epsilon$. By Lemma 5.2 in Vershynin (2015), we have $|N(\mathbb{S}^{d-1}, \epsilon)| \leq (1 + \frac{2}{\epsilon})^d$. Combing with the definition of M_u and M_{v_1} , it is easy to see that $\|M_u - M_{v_1}\|_F = \|v - v_1\| \leq \epsilon$. And it follows similarly that

$\|M_u - M_{u_1}\|_F = \|u - u_1\| \leq \epsilon$. Then it is easy to see that

$$\begin{aligned}
& \|\text{vec}(M_u \otimes M_v) - \text{vec}(M_{u_1} \otimes M_{v_1})\| \\
&= \|M_u \otimes M_v - M_{u_1} \otimes M_{v_1}\|_F \\
&\leq \|M_u \otimes (M_v - M_{v_1})\|_F + \|(M_u - M_{u_1}) \otimes M_{v_1}\|_F \leq 2\epsilon,
\end{aligned}$$

where we use the fact that $\|Q_1 \otimes Q_2\|_F = \|Q_1\|_F \|Q_2\|_F$ for any two matrices Q_1 and Q_2 , and that $\|M_u\|_F = \|u\| = 1$ and $\|M_{v_1}\|_F = \|v_1\| = 1$, for any $u \in \mathbb{S}^{q^2-1}$ and $v_1 \in \mathbb{S}^{p^2-1}$. Therefore, $\{\text{vec}(M_{u_1} \otimes M_{v_1}), u_1 \in N(\mathbb{S}^{q^2-1}, \epsilon), v \in N(\mathbb{S}^{p^2-1}, \epsilon)\}$ is the 2ϵ -cover of $\{\text{vec}(M_u \otimes M_v) : (u, v) \in \Delta\}$. And consequently, similar to the argument of Tsiligkaridis and Hero (2013), by assumption (A1), we have

$$\begin{aligned}
& P\left(\sup_{u,v \in \Delta} \langle E_n, \text{vec}(M_u \otimes M_v) \rangle > t\right) \\
&\leq P\left((1-2\epsilon)^{-1} \max_{\substack{u \in N(\mathbb{S}^{q^2-1}, \epsilon) \\ v \in N(\mathbb{S}^{p^2-1}, \epsilon)}} \langle E_n, \text{vec}(M_u \otimes M_v) \rangle > t\right) \\
&\leq \left(1 + \frac{2}{\epsilon}\right)^{p^2+q^2} \sup_{u,v \in \Delta} P\left(\langle E_n, \text{vec}(M_u \otimes M_v) \rangle > \frac{t(1-2\epsilon)}{\sqrt{n}}\right) \\
&\leq \left(1 + \frac{2}{\epsilon}\right)^{p^2+q^2} \exp\left(-\frac{[t(1-2\epsilon)]^2}{K}\right). \tag{S1.5}
\end{aligned}$$

Set $t = C\sqrt{p^2 + q^2 + \log n}$ with $C^2 > K(\log(1 + 2\epsilon^{-1}) + 1)/(1 - 2\epsilon)$. Con-

sequently, it follows that

$$P\left(\sup_{u,v \in \Delta} \langle E_n, \text{vec}(M_u \otimes M_v) \rangle > C \sqrt{(p^2 + q^2 + \log n)/n}\right) \leq n^{-1} \exp(-(p^2 + q^2)) \rightarrow 0.$$

Step 2 We consider the $\|\mathcal{T}(\hat{\mathbf{R}}^\tau) - \mathcal{T}(\mathbf{R})\|_{op}$. This proof is similar to the Theorem 3.2 of Han and Liu (2017). We omit the similar part and only state the parts that are different. Similar to Han and Liu (2017), we have $\mathcal{T}(\hat{\mathbf{R}}^\tau) - \mathcal{T}(\mathbf{R}) = \mathcal{T}(\tilde{E}_1) + \mathcal{T}(\tilde{E}_2)$, where $\tilde{E}_1 = (\tilde{E}_{1,jk})$, $\tilde{E}_2 = (\tilde{E}_{2,jk}) \in \mathbb{R}^{pq \times pq}$ are defined as, for $j \neq k$,

$$\tilde{E}_{1,jk} = \cos\left(\frac{\pi}{2} \mathbf{T}_{jk}\right) \frac{\pi}{2} (\hat{\mathbf{T}}_{jk} - \mathbf{T}_{jk}), \quad \tilde{E}_{2,jk} = -\frac{1}{2} \sin^2\left(\frac{\pi}{2} \mathbf{T}_{jk}\right) \left(\frac{\pi}{2}\right)^2 (\hat{\mathbf{T}}_{jk} - \mathbf{T}_{jk})^2,$$

and the diagonal elements of \tilde{E}_1 and \tilde{E}_2 are zero. For given $\alpha > 0$ being small, define $\Omega_\alpha = \{\exists 1 \leq k \neq j \leq pq, |E_{2,jk}| > \pi^2 \log(pq/\alpha)/n\}$. Han and Liu (2017) showed that $P(\Omega_\alpha) \leq \alpha^2$. Conditioning on Ω_α^c , for any $u \in \mathbb{S}^{q^2-1}$ and any $v \in \mathbb{S}^{p^2-1}$, it follows that

$$|u^\top \mathcal{T}(\tilde{E}_2)v| \leq \sqrt{\sum_{j,k} \tilde{E}_{2,jk}^2} \|u\| \|v\| \leq \sqrt{(pq)^2 (\pi^2 \log(pq/\alpha)/n)^2}.$$

Therefore, with probability $1 - \alpha^2$, $\|\mathcal{T}(\tilde{E}_2)\|_{op} \leq pq\pi^2 \log(pq/\alpha)/n$.

Moreover, let $F = (F_{jk}) \in \mathbb{R}^{pq \times pq}$ with $F_{jk} = \frac{\pi}{2} \cos(\frac{\pi}{2} \mathbf{T}_{jk})$, and note

that $\hat{\mathbf{T}} = (\hat{\mathbf{T}}_{jk})$ is the Kendall's τ correlation matrix. Then

$$\tilde{E}_1 = F \circ (\hat{\mathbf{T}} - \mathbf{T}),$$

and consequently $\mathcal{T}(\tilde{E}_1) = \mathcal{T}(F) \circ \mathcal{T}(\hat{\mathbf{T}} - \mathbf{T})$, where "o" denotes the Hadamard product. In addition, similar to the argument of Han and Liu (2017), for any matrix $Q \in \mathbb{R}^{q^2 \times p^2}$ and any $u \in \mathbb{S}^{q^2-1}, v \in \mathbb{S}^{p^2-1}$, we can show that

$$|u^\top (Q \circ \mathcal{T}(F))v| \leq \frac{\pi^2}{2} \|Q\|_{op}. \quad (\text{S1.6})$$

Consequently,

$$\|\mathcal{T}(\tilde{E}_1)\|_{op} = \max_{u \in \mathbb{S}^{q^2-1}, v \in \mathbb{S}^{p^2-1}} |u^\top (\mathcal{T}(F) \circ \mathcal{T}(\hat{\mathbf{T}} - \mathbf{T}))v| \leq \frac{\pi^2}{2} \|\mathcal{T}(\hat{\mathbf{T}} - \mathbf{T})\|_{op}$$

Combining together, we have with probability $1 - \alpha^2 - n^{-1} \exp(-(p^2 + q^2))$,

$$\begin{aligned} \|\mathcal{T}(\mathbf{R}^\tau) - \mathcal{T}(\mathbf{R})\|_{op} &\leq \|\mathcal{T}(\tilde{E}_1)\|_{op} + \|\mathcal{T}(\tilde{E}_2)\|_{op} \\ &\leq C_1 \sqrt{\frac{p^2 + q^2 + \log n}{n}} + \frac{pq\pi^2 \log(pq/\alpha)}{n}, \end{aligned}$$

for some constant $C_1 > 0$. This completes the proof. ■

S1.4 Proof of Theorem 2

Step 1. We first establish the bound $\|\mathcal{T}(\hat{\Sigma}^\tau) - \mathcal{T}(\Sigma)\|_{op} = O_p(\omega_n^{(0)})$.

Recalling that $\hat{\Sigma}^\tau = \hat{D}\hat{\mathbf{R}}^\tau\hat{D}$, we have

$$\begin{aligned} \|\mathcal{T}(\hat{\Sigma}^\tau) - \mathcal{T}(\Sigma)\|_{op} &= \|\mathcal{T}(\hat{D}\hat{\mathbf{R}}^\tau\hat{D}) - \mathcal{T}(D\mathbf{R}D)\|_{op} \\ &\leq \|\mathcal{T}(\hat{D}\hat{\mathbf{R}}^\tau\hat{D}) - \mathcal{T}(\hat{D}\mathbf{R}\hat{D})\|_{op} + \|\mathcal{T}(\hat{D}\mathbf{R}\hat{D}) - \mathcal{T}(D\mathbf{R}D)\|_{op} \\ &:= \|J_1\|_{op} + \|J_2\|_{op}, \end{aligned} \tag{S1.7}$$

where J_1 and J_2 are defined accordingly.

Consider $\|J_1\|_{op}$ first. Letting $\Delta = \{(u, v) : u \in \mathbb{S}^{q^2-1}, v \in \mathbb{S}^{p^2-1}\}$ and $\mathbf{F} = \hat{\mathbf{R}}^\tau - \mathbf{R}$, we have

$$\|J_1\|_{op} = \sup_{(u,v) \in \Delta} u^\top J_1 v = \sup_{(u,v) \in \Delta} u^\top \mathcal{T}(\hat{D}\mathbf{F}\hat{D})v.$$

Recall that $D^{(d)}$ is the vector of diagonal elements of D . Then $\hat{D}\mathbf{F}\hat{D} = \mathbf{F} \circ \hat{D}^{(d)}\hat{D}^{(d)\top}$ and $\mathcal{T}(\hat{D}\mathbf{F}\hat{D}) = \mathcal{T}(\mathbf{F}) \circ \mathcal{T}(\hat{D}^{(d)}\hat{D}^{(d)\top})$. Therefore,

$$u^\top \mathcal{T}(\hat{D}\mathbf{F}\hat{D})v = \langle \mathcal{T}(\mathbf{F}), \mathcal{T}(\hat{D}^{(d)}\hat{D}^{(d)\top}) \circ (u \otimes v^\top) \rangle \leq \|\hat{D}\|_{\max}^2 \langle \mathcal{T}(\mathbf{F}), u \otimes v^\top \rangle = \|\hat{D}\|_{\max}^2 u^\top \mathcal{T}(\mathbf{F})v.$$

In addition, by Lemma 1, we have $\|\hat{D}\|_{\max}^2 = \|D\|_{\max}^2 + O_p(\omega_n^{(2)})$. Therefore,

combining with Theorem 1, we have

$$\|J_1\|_{op} \leq \|\hat{D}\|_{\max}^2 \|\mathcal{T}(\mathbf{F})\|_{op} = \|\mathcal{T}(\hat{\mathbf{R}}^\tau) - \mathcal{T}(\mathbf{R})\|_{op} (\|D\|_{\max}^2 + O_p(\omega_n^{(2)})) = O_p(\omega_n^{(1)}) \|D\|_{\max}^2.$$

Now we consider $\|J_2\|_{op}$. It is easy to see that

$$\begin{aligned} J_2 &= \mathcal{T}(\hat{D}\mathbf{R}\hat{D}) - \mathcal{T}(D\mathbf{R}D) \\ &= \mathcal{T}((\hat{D} - D)\mathbf{R}D) + \mathcal{T}(D\mathbf{R}(\hat{D} - D)) + \mathcal{T}((\hat{D} - D)\mathbf{R}(\hat{D} - D)) \\ &:= J_{21} + J_{22} + J_{23}. \end{aligned}$$

Applying the same argument as that of J_1 on each term in J_2 , we get

$$\|J_2\|_{op} \leq 2\|\mathcal{T}(\mathbf{R})\|_{op} \|D\|_{\max} \|\hat{D} - D\|_{\max} + \|\mathcal{T}(\mathbf{R})\|_{op} \|\hat{D} - D\|_{\max}^2.$$

Combining together and noting the definition of $\omega_n^{(0)}$, we have

$$\|\mathcal{T}(\hat{\Sigma}^\tau) - \mathcal{T}(\Sigma)\|_{op} = O_p(\omega_n^{(1)}) \|D\|_{\max}^2 + \omega_n^{(2)} \|\mathcal{T}(\mathbf{R})\|_{op} \|D\|_{\max} = O_p(\omega_n^{(0)}).$$

Step 2. We prove the final conclusion. By the proof similar to that of the Theorem 2 of Tsiligkaridis and Hero (2013), as $\lambda > 2\|\mathcal{T}(\hat{\Sigma}^\tau) - \mathcal{T}(\Sigma)\|_{op}$,

we have

$$\|\hat{\Sigma}_{\mathcal{T}}^{\tau} - \mathcal{T}(\Sigma)\|_F^2 \leq \inf_{G \in \mathbb{R}^{q^2 \times p^2}} \left\{ \|G - \mathcal{T}(\Sigma)\|_F^2 + \frac{(1 + \sqrt{2})^2}{4} \lambda^2 \text{rank}(G) \right\}$$

Combining with results in Step 1 on the operator norm of $\|\mathcal{T}(\hat{\Sigma}^{\tau}) - \mathcal{T}(\Sigma)\|_{op}$, it follows that, as $\lambda > C\omega_n^{(0)}$ for some constant $C > 0$, with probability tending to 1,

$$\|\hat{\Sigma}_{\mathcal{T}}^{\tau} - \mathcal{T}(\Sigma)\|_F^2 \leq \inf_{\substack{G \in \mathbb{R}^{q^2 \times p^2} \\ \text{rank}(G) \leq r}} \|G - \mathcal{T}(\Sigma)\|_F^2 + r(\omega_n^{(0)})^2.$$

The finally conclusion is derived by noting that $\|\hat{\Sigma}_{LR}^{\tau} - \Sigma\|_F^2 = \|\hat{\Sigma}_{\mathcal{T}}^{\tau} - \mathcal{T}(\Sigma)\|_F^2$. ■

S1.5 Proof of Theorem 3

Denote $\hat{H} = \mathcal{T}(\hat{\Sigma})$, $H = \mathcal{T}(\Sigma)$. Note that the \hat{V}_A is the eigenvector associated with the largest eigenvalue of $\hat{H}\hat{H}^{\top}$. Recall the bound on $\|\mathcal{T}(\hat{\Sigma}^{\tau}) - \mathcal{T}(\Sigma)\|_{op}$ in Step 1 of Theorem 2. Note that $\|H\|_{op} = \|\Sigma\|_F$ when $r = 1$, and that $\|\hat{H}\|_{op} \leq \|\hat{H} - H\|_{op} + \|H\|_{op}$. Then we have

$$\begin{aligned} \|\hat{H}\hat{H}^{\top} - HH^{\top}\|_{op} &\leq \|\hat{H}\hat{H}^{\top} - \hat{H}H^{\top}\|_{op} + \|\hat{H}H^{\top} - HH^{\top}\|_{op} \\ &\leq (\|H\|_{op} + \|\hat{H} - H\|_{op})\|\hat{H} - H\|_{op} + \|\hat{H} - H\|_{op}\|H\|_{op}. \end{aligned}$$

Consequently, we have

$$\|\hat{H}\hat{H}^\top - HH^\top\|_{op} = O_p(\omega_n^{(0)}\|H\|_{op}) = O_p(\omega_n^{(0)}\|\Sigma\|_F).$$

Note that $\gamma^2 = \|\Sigma\|_F^2$ is the unique nonzero eigenvalue of HH^\top . By the Theorem 2 of Yu et al., (2014), we have

$$\|\hat{V}_A - cV_A\| \leq 2^{3/2}\gamma^{-2}\|\hat{H}\hat{H}^\top - HH^\top\|_{op} = O_p(\omega_n^{(0)}/\|\Sigma\|_F).$$

The conclusion on \hat{V}_B can be derived similarly. The conclusions on \hat{A}, \hat{B} can be derived by noting that $\|\hat{V}_A - cV_A\| = \|\hat{A} - cA\|_F$ and $\|\hat{V}_B - c'V_B\| = \|\hat{B} - c'B\|_F$.

We consider $\|\hat{\Sigma}_{(rk=1)}^\tau - \Sigma\|_F^2$. Note that $\|V_A\| = \|\hat{V}_A\| = \|V_B\| = \|\hat{V}_B\| = 1$. It is easy to see that $\|\hat{V}_A\hat{V}_B^\top - V_AV_B^\top\|_F = O_p(\omega_n^{(0)}/\|\Sigma\|_F)$. Noting that $\hat{\gamma}$ and γ are the largest eigenvalue of $\mathcal{T}(\hat{\Sigma}^\tau)$ and $\mathcal{T}(\Sigma)$, respectively, then we have

$$|\gamma - \hat{\gamma}| \leq \|\mathcal{T}(\hat{\Sigma}^\tau) - \mathcal{T}(\Sigma)\|_{op} = O_p(\omega_n^{(0)}).$$

Combining together, we have

$$\|\hat{\gamma}\hat{V}_A\hat{V}_B^\top - \gamma V_AV_B^\top\|_F \leq \gamma\|\hat{V}_A\hat{V}_B^\top - V_AV_B^\top\|_F + |\gamma - \hat{\gamma}| \cdot \|\hat{V}_A\hat{V}_B^\top\|_F = O_p(\omega_n^{(0)}),$$

where we use the fact that $\gamma = \|\Sigma\|_F$ and $\|\hat{V}_A \hat{V}_B^\top\|_F = 1$. The final conclusion is derived by noting that $\|\hat{\Sigma}_{(rk=1)}^\tau - \Sigma\|_F^2 = \|\hat{\gamma} \hat{V}_A \hat{V}_B^\top - \mathcal{T}(\Sigma)\|_F^2 = \|\hat{\gamma} \hat{V}_A \hat{V}_B^\top - \gamma V_A V_B^\top\|_F^2$. This completes the proof. ■

S2 Some simulation results for Section 4 and figures in Section 5

We present the simulation results on $s_1 = \lceil p/4 \rceil, s_2 = \lceil q/4 \rceil$ for Example 1–4, and the Figure 1–3 in the real data analysis.

Table S1: Simulation results for Example 1 and 2 with $s_1 = \lceil p/4 \rceil, s_2 = \lceil q/4 \rceil$

n, p, q		Example 1				Example 2
		$\delta_0 = 0$	$\delta_0 = 10$	$\delta_0 = 20$	$\delta_0 = 50$	
(100,15,15)	$Err_F^{(rob)}$	0.0241	0.0261	0.0314	0.0441	0.0025
	$Err_F^{(sam)}$	0.0101	0.0168	0.0331	0.1135	0.0032
	$Err_2^{(rob)}$	0.0201	0.0212	0.0258	0.0351	0.0021
	$Err_2^{(sam)}$	0.0077	0.0137	0.0268	0.0876	0.0026
	$Err_\infty^{(rob)}$	0.0045	0.0048	0.0057	0.0077	0.0005
	$Err_\infty^{(sam)}$	0.0017	0.0031	0.0058	0.0186	0.0009
(100,25,25)	$Err_F^{(rob)}$	0.0150	0.0166	0.0177	0.0227	0.0008
	$Err_F^{(sam)}$	0.0070	0.0097	0.0145	0.0424	0.0014
	$Err_2^{(rob)}$	0.0106	0.0119	0.0126	0.0158	0.0006
	$Err_2^{(sam)}$	0.0046	0.0068	0.0103	0.0282	0.0012
	$Err_\infty^{(rob)}$	0.0018	0.0019	0.0021	0.0025	0.0001
	$Err_\infty^{(sam)}$	0.0007	0.0010	0.0017	0.0045	0.0005

Table S2: Simulation results for Example 3 and 4 with $s_1 = \lceil p/4 \rceil, s_2 = \lceil q/4 \rceil$

n, p, q		Example 3				Example 4
		$\delta_0 = 0$	$\delta_0 = 10$	$\delta_0 = 20$	$\delta_0 = 50$	
(100,15,15)	$Err_F^{(rob)}$	0.2393	0.2493	0.2803	0.4370	0.0895
	$Err_F^{(sam)}$	0.0881	0.1456	0.2677	1.2288	0.4531
	$Err_2^{(rob)}$	0.1801	0.1925	0.2142	0.3589	0.0620
	$Err_F^{(sam)}$	0.0714	0.1118	0.2287	1.1345	0.3941
	$Err_\infty^{(rob)}$	0.0496	0.0521	0.0579	0.0883	0.0245
	$Err_\infty^{(sam)}$	0.0188	0.0275	0.0542	0.2254	0.1209
(100,25,25)	$Err_F^{(rob)}$	0.1457	0.1523	0.1642	0.2389	0.0563
	$Err_F^{(sam)}$	0.0537	0.0723	0.1688	0.7289	0.2908
	$Err_2^{(rob)}$	0.0956	0.1007	0.1076	0.1780	0.0338
	$Err_2^{(sam)}$	0.0365	0.0500	0.1510	0.6947	0.2278
	$Err_\infty^{(rob)}$	0.0205	0.0216	0.0238	0.0335	0.0099
	$Err_\infty^{(sam)}$	0.0081	0.0099	0.0190	0.0764	0.0675

Figure 1: Robust covariance matrix with estimated \hat{r} obtained by our proposal. There are seven tissues: Cerebrum, Hippocampus, Kidney, Lung, Muscle, Thymus and Spinal cord, which associate with the diagonal blocks from the lower left corner to the upper right corner, respectively. There are weak correlation between tissues Hippocampus and Thymus in male, and clear correlation between tissues Cerebrum and Thymus in female.

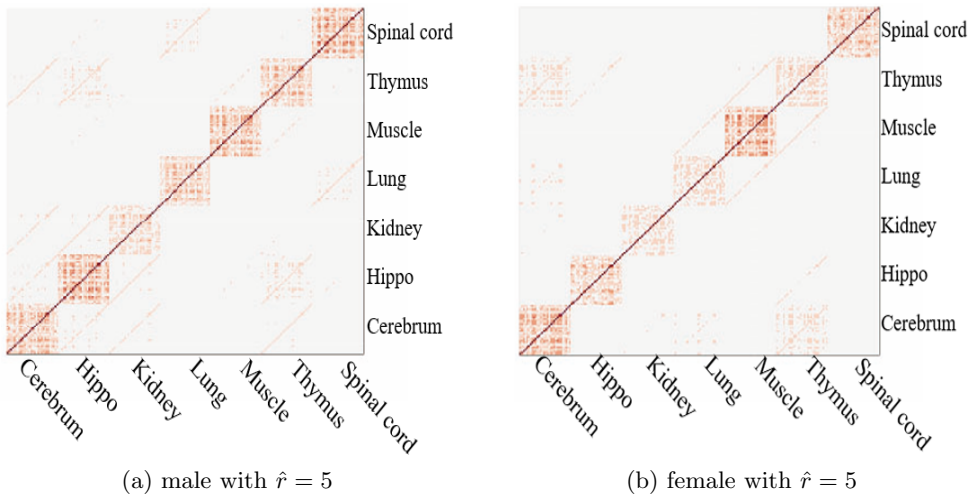


Figure 2: Non-robust covariance matrix estimation with estimated \hat{r} . This is the estimator of Tsiligkaridis and Hero (2013). There are clear dependency among Lung, Muscle and Thymus in male data and almost no correlation among different tissues for female data.

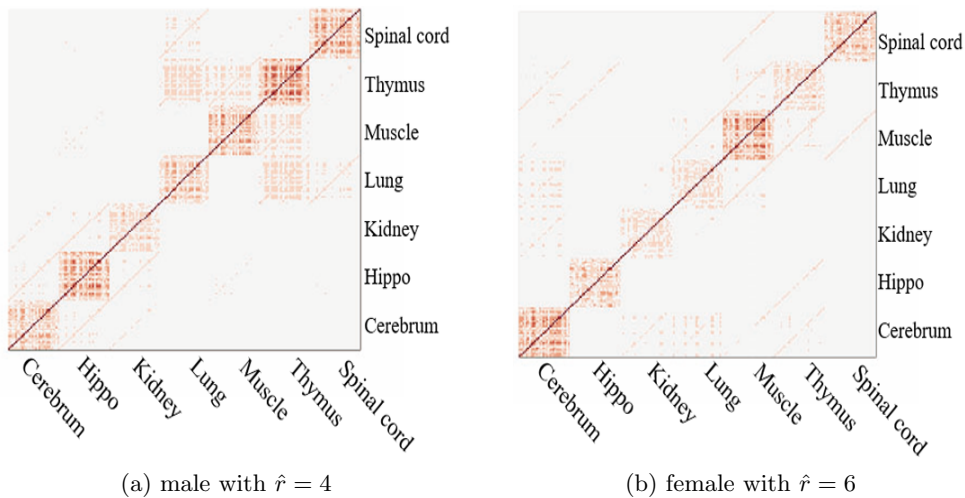
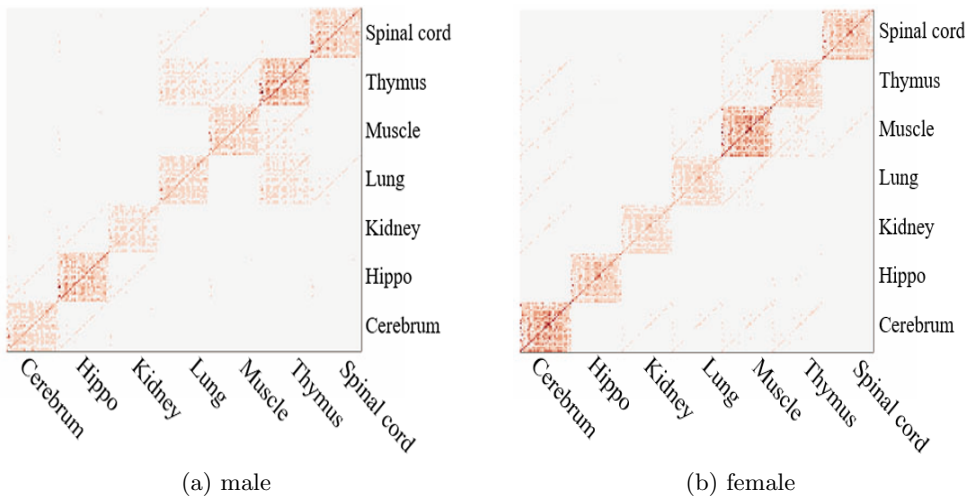


Figure 3: Non-robust covariance matrix estimation with fixed $r = 1$. This is the estimator of Leng and Pan (2017). There are clear dependency between Thymus and Lung in male data and almost no correlation among different tissues for female data.



Bibliography

- Han, F., Liu, H., 2017. Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution. *Bernoulli* 23 (1), 23–57.
- Leng, C., Pan, G., 2017. Covariance estimation via sparse kronecker structure. *Bernoulli* In press.
- Serfling, R. J., 2009. Approximation theorems of mathematical statistics. Vol. 162. John Wiley & Sons.
- Tsiligkaridis, T., Hero, A. O., 2013. Covariance estimation in high dimensions via kronecker product expansions. *Signal Processing IEEE Transactions on* 61 (21), 5347–5360.
- Vershynin, R., 2015. Introduction to the non-asymptotic analysis of random matrices. Eprint Arxiv.