SUPPLEMENT TO "ON CUMULATIVE SLICING ESTIMATION FOR HIGH DIMENSIONAL DATA"

Cheng Wang, Zhou Yu and Liping Zhu

Shanghai Jiao Tong University, East China Normal University,

Renmin University of China

This Supplement Material provides additional simulations, and proofs of theoretical results in the main context.

S1. Simulations

S2. Proof of the Inconsistency Issue in Example 1

In Example 1, $\mathbf{x}$ and Y are jointly normal. The normality yields that

$$\begin{aligned}\mathbf{m}(y) &= \text{cov}\{\mathbf{x}, I(Y \le y)\} = \left[-\frac{1}{\sqrt{2\pi}\sqrt{1+\sigma^2}}\exp\left\{-\frac{y^2}{2(1+\sigma^2)}\right\}, 0, \cdots, 0\right]^{\mathrm{T}}, \\ \mathbf{\Lambda}_{1,1} &= \frac{1}{2\sqrt{3}\pi(1+\sigma^2)},\ \mathbf{\Lambda}_{k,l} = 0 \text{ for } k^2 + l^2 > 2.\end{aligned}$$

We sort the response in an ascending order to obtain $Y_{(1)} \le Y_{(2)} \le \cdots \le Y_{(n)}$. Denote by $\mathbf{x}_{(i)}$ the corresponding quantity associated with $Y_{(i)}$. For

Table 1: The averages (standard deviations) of trace correlation($\times$ 100) for Scenario 1 where SIR_k denotes SIR with k slices and $\mathbf{\Sigma} = \mathbf{I}_{p\times p}$.

	p	CUME	SIR_2	SIR_5	SIR_{10}	SIR_{20}	NEW
(1.2)	10	99.2(0.4)	96.9(1.4)	99.1(0.4)	99.4(0.3)	99.6(0.2)	97.5(2.2)
	50	94.5(1.5)	81.8(4.0)	94.0(1.5)	96.2(1.0)	96.9(0.8)	97.5(3.5)
	100	84.2(3.7)	60.0(6.1)	83.8(3.3)	89.3(2.4)	90.8(2.0)	97.6(3.1)
	200	-	-	-	-	-	97.6(3.1)
	300	-	-	-	-	-	97.2(3.6)
(1.3)	10	97.1(1.5)	94.8(2.6)	96.8(1.7)	96.9(1.6)	96.7(1.8)	97.3(3.8)
	50	83.0(4.9)	72.8(6.4)	81.7(5.2)	81.7(5.4)	79.8(6.3)	98.3(5.5)
	100	61.6(7.9)	47.5(8.0)	58.8(8.3)	57.7(8.9)	51.6(11.2)	99.0(4.4)
	150	33.5(8.6)	23.3(7.1)	30.4(8.9)	25.8(10.2)	13.9(9.9)	99.2(3.8)
	200	-	-	-	-	-	99.3(3.7)
	300	-	-	-	-	-	99.4(3.6)
(1.4)	10	98.8(0.6)	95.4(2.2)	98.9(0.5)	99.3(0.4)	99.4(0.3)	97.6(2.3)
	50	91.6(2.4)	74.6(5.2)	92.9(1.8)	95.4(1.1)	96.2(1.0)	98.0(3.1)
	100	76.5(5.1)	49.6(6.7)	80.8(3.7)	86.9(2.8)	88.5(2.5)	98.2(2.8)
	200	-	-	-	-	-	98.1(3.0)
	300	-	-	-	-	-	98.1(3.1)
(1.5)	10	88.9(5.5)	94.3(3.4)	84.3(8.5)	86.9(8.0)	86.5(8.7)	94.9(7.1)
	50	59.0(6.0)	75.2(5.2)	53.6(6.7)	55.2(7.8)	52.4(7.8)	94.6(8.9)
	100	39.6(4.5)	60.3(7.5)	35.9(4.5)	36.0(4.7)	33.2(4.7)	94.4(8.6)
	200	-	-	-	-	-	93.8(9.4)
	300	-	-	-	-	-	94.1(9.2)
(1.6)	10	89.4(5.5)	58.6(3.6)	88.0(6.3)	89.3(6.1)	87.7(7.4)	90.8(9.4)
	50	54.5(7.3)	45.5(5.2)	55.6(7.6)	56.5(8.4)	51.8(8.8)	84.8(14.6)
	100	34.4(5.5)	31.7(7.5)	35.4(6.0)	35.0(6.2)	31.2(5.8)	82.3(15.5)
	200	-	-	-	-	-	79.9(16.2)
	300	-	-	-	-	-	77.6(15.9)
(1.7)	10	97.7(0.9)	48.9(0.7)	96.9(1.2)	97.5(1.0)	97.5(1.0)	95.6(3.1)
	50	84.5(2.8)	41.5(1.9)	80.2(3.3)	83.2(2.9)	82.9(3.2)	93.4(6.5)
	100	63.6(4.7)	30.9(3.1)	57.8(4.7)	60.9(5.1)	57.4(6.2)	88.6(13.2)
	200	-	-	-	-	-	76.6(22.2)
	300	-	-	-	-	-	71.2(24.0)

Table 2: The averages (standard deviations) of trace correlation($\times$ 100) for Scenario 3 where SIR_k denotes SIR with k slices and $\mathbf{\Sigma} = (0.5^{|k-l|})_{p\times p}$.

p_0	CUME	SIR_2	SIR_5	SIR_{10}	SIR_{20}	NEW
	(1.2) with $p = 1000$					
38	92.7(2.3)	74.4(5.8)	92.1(2.3)	95.7(1.3)	97.0(1.0)	97.2(3.2)
76	80.3(5.1)	52.0(6.4)	81.3(4.0)	89.2(2.6)	92.0(2.0)	96.8(3.7)
114	56.6(10.6)	33.1(6.1)	65.9(6.0)	78.2(4.6)	82.7(3.9)	96.6(4.0)
152	15.7(11.3)	17.0(4.6)	43.5(7.4)	57.2(7.7)	59.8(8.2)	96.5(4.4)
190	1.1(1.4)	3.5(2.3)	7.9(5.1)	4.2(4.8)	0.5(1.7)	96.3(4.7)
	(1.5) with $p = 1000$					
38	66.1(5.8)	80.6(7.6)	62.6(6.5)	67.0(6.8)	64.0(7.8)	93.6(9.8)
76	48.4(4.6)	65.5(9.2)	45.6(4.8)	47.4(5.9)	43.0(6.1)	92.8(11.3)
114	34.8(4.6)	48.6(12.5)	32.3(4.7)	32.0(5.4)	26.4(6.0)	92.4(11.6)
152	20.9(4.6)	29.4(13.7)	18.8(4.6)	16.6(5.3)	10.3(5.3)	91.5(12.2)
190	5.1(2.8)	10.0(9.5)	4.5(2.9)	2.7(2.0)	1.6(1.3)	90.6(12.7)
	(1.2) with $p = 5000$					
38	92.7(2.4)	75.1(5.8)	92.3(2.2)	95.8(1.3)	97(0.9)	97.4(3.1)
76	81.1(5.0)	53.8(6.8)	82.0(3.8)	89.5(2.6)	92.3(2.0)	97.2(3.5)
114	59.9(9.6)	35.0(6.5)	67.4(5.9)	79.0(4.7)	83.5(3.9)	97.0(3.6)
152	18.1(12.9)	18.4(5.0)	45.4(7.7)	58.7(7.7)	61.2(8.3)	96.8(4.0)
190	1.4(1.7)	3.9(2.5)	8.6(5.6)	4.7(5.2)	0.5(1.7)	96.7(4.2)
	(1.5) with $p = 5000$					
38	66.9(6.3)	79.3(9.7)	63.3(6.9)	67.8(7.0)	64.8(7.9)	93.4(10.4)
76	49.6(4.7)	64.6(11.2)	46.6(5.1)	48.5(6.0)	44.1(6.1)	93.4(10.7)
114	36.2(4.4)	47.9(13.6)	33.5(4.6)	33.2(5.3)	27.6(6.1)	93.0(11.2)
152	22.5(4.7)	30.1(13.5)	19.9(4.9)	17.7(5.5)	11.1(5.4)	92.3(12.0)
190	5.5(3.1)	12.4(12.4)	4.8(2.9)	3(2.3)	1.6(1.4)	91.7(12.4)

Example 1, we have $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$ almost surely. Thus

$$\begin{aligned}\widehat{\mathbf{\Lambda}} &= n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathbf{x}_{(j)}\mathbf{x}_{(k)}^{\mathrm{T}} I(Y_{(j)} \le Y_{(i)}) I(Y_{(k)} \le Y_{(i)}) \\ &= n^{-3}\sum_{j=1}^{n}\sum_{k=1}^{n}\{n+1-\max(j,k)\}\mathbf{x}_{(j)}\mathbf{x}_{(k)}^{\mathrm{T}} \\ &= n^{-1}\big(\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \cdots, \mathbf{x}_{(n)}\big)\mathbf{T}\big(\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \cdots, \mathbf{x}_{(n)}\big)^{\mathrm{T}},\end{aligned}$$

where $\mathbf{T}$ is an $n \times n$ matrix with its (j,k)-th element being $\mathbf{T}_{j,k} = n^{-2}\{n + 1 - \max(j,k)\}$. It can be verified that $\mathbf{T}$ is positive definite. By Gershgorin's circle theorem,

$$\|\mathbf{T}\| \le n^{-2}\sum_{j=1}^{n}(n+1-j) < 1.$$

We partition the matrix $\widehat{\mathbf{\Lambda}}$ into block matrices as

$$\widehat{\mathbf{\Lambda}} = \begin{pmatrix} w & \mathbf{w}^{\mathrm{T}} \\ \mathbf{w} & \mathbf{W}\end{pmatrix} = \begin{pmatrix} w & \mathbf{w}^{\mathrm{T}} \\ \mathbf{w} & \mathbf{0}\end{pmatrix} + \begin{pmatrix} 0 & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{W}\end{pmatrix} \stackrel{\text{def}}{=} \widehat{\mathbf{\Lambda}}_1 + \widehat{\mathbf{\Lambda}}_2.$$

Denote the leading eigenvalue and eigenvector as λ and $(b, \mathbf{b}^{\mathrm{T}})^{\mathrm{T}}$, respectively. By definition, $b\mathbf{w} + \mathbf{W}\mathbf{b} = \lambda\mathbf{b}$. Therefore, $b^2\|\mathbf{w}\|_2^2 \le \|\mathbf{W}\mathbf{b}\|_2^2 + \lambda^2\|\mathbf{b}\|_2^2 \le (\|\mathbf{W}\|^2 + \lambda^2)\|\mathbf{b}\|_2^2 = (\|\mathbf{W}\|^2 + \lambda^2)(1 - b^2)$. Accordingly,

$$\|\mathbf{P} - \widehat{\mathbf{P}}\|_F^2 = 2(1-b^2) \ge \frac{2\|\mathbf{w}\|_2^2}{\|\mathbf{w}\|_2^2 + \|\mathbf{W}\|^2 + \lambda^2}. \tag{S2.1}$$

The matrix $\widehat{\mathbf{\Lambda}}_1$ has two non-zero eigenvalues, $(w \pm \sqrt{w^2 + 4\|\mathbf{w}\|_2^2})/2$. By Weyl's inequality,

$$\lambda \le \frac{w + \sqrt{w^2 + 4\|\mathbf{w}\|_2^2}}{2} + \|\mathbf{W}\|.$$

In other words,

$$\lambda^2 \le 2\left(\frac{w^2 + w^2 + 4\|\mathbf{w}\|_2^2}{2} + \|\mathbf{W}\|^2\right) = 2w^2 + 4\|\mathbf{w}\|_2^2 + 2\|\mathbf{W}\|^2.$$

With the above inequality, (S2.1) reduces to

$$\|\mathbf{P} - \widehat{\mathbf{P}}\|_F^2 \ge \frac{2}{3}\frac{\|\mathbf{w}\|_2^2}{w^2 + 2\|\mathbf{w}\|_2^2 + \|\mathbf{W}\|^2}. \tag{S2.2}$$

Now, we study the asymptotic distribution of $\widehat{\mathbf{\Lambda}}$. Let $\mathbf{e}_k$ be a unit-length p-vector with its k-th entry being one. We have

$$\widehat{\mathbf{\Lambda}}_{a,b} = \mathbf{e}_a^{\mathrm{T}}\widehat{\mathbf{\Lambda}}\mathbf{e}_b = n^{-3}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n \mathbf{x}_{j,a}\mathbf{x}_{k,b}I(Y_j \le Y_i)I(Y_k \le Y_i).$$

For $(a,b) = (1,1)$, by the classical result for U-statistics,

$$w = \widehat{\mathbf{\Lambda}}_{1,1} \xrightarrow{p} \mathrm{E}\mathbf{x}_{1,1}\mathbf{x}_{2,1}I(Y_1 \le Y_3)I(Y_2 \le Y_3) = \mathbf{\Lambda}_{1,1}. \tag{S2.3}$$

Define

$$c_k \overset{\text{def}}{=} n^{-3}\sum_{i=1}^n\sum_{j=1}^n \mathbf{x}_{j,1}I(Y_j \le Y_i)I(Y_k \le Y_i).$$

Then

$$\begin{aligned} n\sum_{k=1}^n c_k^2 &= n^{-5}\sum_{i,j,k,i',j'}^n \mathbf{x}_{j,1}\mathbf{x}_{j',1}I(Y_j \le Y_i)I(Y_k \le Y_i)I(Y_{j'} \le Y_{i'})I(Y_k \le Y_{i'}) \\ &\xrightarrow{p} \mathrm{E}\mathbf{x}_{2,1}\mathbf{x}_{5,1}I(Y_2 \le Y_1)I(Y_3 \le Y_1)I(Y_3 \le Y_4)I(Y_5 \le Y_4) \\ &= \mathrm{E}m(Y_1)m(Y_4)I(Y_3 \le Y_1)I(Y_3 \le Y_4) = \frac{\arctan\sqrt{5}}{4\pi^2(1+\sigma^2)}, \end{aligned}$$

and

$$\begin{aligned}\|\mathbf{w}\|_2^2 &= \sum_{b=2}^{p} \widehat{\mathbf{\Lambda}}_{1,b}^2 = \sum_{b=2}^{p} \left\{ n^{-3} \sum_{i,j,k}^{n} \mathbf{x}_{j,1}\mathbf{x}_{k,b} I(Y_j \le Y_i) I(Y_k \le Y_i) \right\}^2 \\ &= \sum_{b=2}^{p} \left\{ \sum_{k=1}^{n} c_k \mathbf{x}_{k,b} \right\}^2 \overset{d}{=} n \sum_{k=1}^{n} c_k^2 \cdot \frac{\chi_{p-1}^2}{n} \\ &\xrightarrow{p} \frac{\gamma \arctan\sqrt{5}}{4\pi^2(1+\sigma^2)}. \end{aligned} \tag{S2.4}$$

We next study the eigenvalues of $\mathbf{W}$. Write $\widetilde{\mathbf{x}}_i$ as the vector $\mathbf{x}_i$ without the first element. Define $\widetilde{\mathbf{x}}_{(i)}$ in a similar fashion. Following $\widehat{\mathbf{\Lambda}}$, we have

$$\begin{aligned}\mathbf{W} &= n^{-1}\big(\widetilde{\mathbf{x}}_{(1)}, \widetilde{\mathbf{x}}_{(2)}, \cdots, \widetilde{\mathbf{x}}_{(n)}\big)\mathbf{T}\big(\widetilde{\mathbf{x}}_{(1)}, \widetilde{\mathbf{x}}_{(2)}, \cdots, \widetilde{\mathbf{x}}_{(n)}\big)^{\mathrm{T}} \\ &\overset{d}{=} n^{-1}\big(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \cdots, \widetilde{\mathbf{x}}_n\big)\mathbf{T}\big(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \cdots, \widetilde{\mathbf{x}}_n\big)^{\mathrm{T}},\end{aligned}$$

where we used the fact that $\widetilde{\mathbf{x}}_1, \cdots, \widetilde{\mathbf{x}}_n$ are independent of the observations $\{Y_i : 1 \le i \le n\}$. Note that $\big(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \cdots, \widetilde{\mathbf{x}}_n\big)$ are composed of i.i.d. entries. By Yin et al. (1988),

$$\|\mathbf{W}\| \le \|n^{-1}\big(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \cdots, \widetilde{\mathbf{x}}_n\big)\big(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \cdots, \widetilde{\mathbf{x}}_n\big)^{\mathrm{T}}\| \xrightarrow{a.s.} (1+\sqrt{\gamma})^2. \tag{S2.5}$$

Combining the results of (S2.2)-(S2.5), in probability, we have

$$\begin{aligned}\|\mathbf{P} - \widehat{\mathbf{P}}\|_F^2 &\ge \frac{2}{3} \frac{\frac{\gamma \arctan\sqrt{5}}{4\pi^2(1+\sigma^2)}}{\frac{1}{12\pi^2(1+\sigma^2)^2} + 2\frac{\gamma \arctan\sqrt{5}}{4\pi^2(1+\sigma^2)} + (1+\sqrt{\gamma})^4} \\ &= \frac{\gamma}{\frac{1}{2\arctan\sqrt{5}(1+\sigma^2)} + 3\gamma + \frac{6\pi^2(1+\sigma^2)(1+\sqrt{\gamma})^4}{\arctan\sqrt{5}}} \\ &\ge \frac{\gamma}{6\pi^2(1+\sigma^2)(1+\gamma)^2}.\end{aligned}$$

The proof is completed. □

S3. Some Useful Lemmas

We first present some useful lemmas to study the properties of $\mathbf{m}(y)$, $\mathbf{\Lambda}$, $\widehat{\mathbf{m}}(y)$ and $\widehat{\mathbf{\Lambda}}$. These lemmas pave the road for proving Theorem 1 with $p = o(n)$ and Theorem 2 with $\log(p) = o(n)$. For notational clarity, in what follows we assume without loss of generality that $E(\mathbf{x}) = \mathbf{0}$.

LEMMA 1. For a p-dimension random vector $\mathbf{z}$ and a unit-length vector $\mathbf{e} \in \mathcal{S}^{p-1}$ where $\mathcal{S}^{p-1}$ denotes the unit Euclidean sphere in $\mathbb{R}^p$,

$$\begin{aligned} \mathrm{pr}(\|\mathbf{z}\|_\infty \geq t) &\leq p \sup_{\mathbf{e}\in\mathcal{S}^{p-1}} \mathrm{pr}(|\mathbf{e}^{\mathrm{T}}\mathbf{z}| \geq t), \text{ and} \\ \mathrm{pr}(\|\mathbf{z}\| \geq t) &\leq 5^p \sup_{\mathbf{e}\in\mathcal{S}^{p-1}} \mathrm{pr}(|\mathbf{e}^{\mathrm{T}}\mathbf{z}| \geq t/2), \text{ for any } t \geq 0. \end{aligned}$$

Proof of Lemma 1: Let $\mathbf{e}_k$ be a unit-length p-vector with its k-th entry being one. Apparently, $\|\mathbf{z}\|_\infty = \max_k |\mathbf{e}_k^{\mathrm{T}}\mathbf{z}|$, which entails that

$$\mathrm{pr}(\|\mathbf{z}\|_\infty \geq t) \leq \sum_{k=1}^{p} \mathrm{pr}(|\mathbf{e}_k^{\mathrm{T}}\mathbf{z}| \geq t) \leq p \sup_{\mathbf{e}\in\mathcal{S}^{p-1}} \mathrm{pr}(|\mathbf{e}^{\mathrm{T}}\mathbf{z}| \geq t),$$

which completes proof of the first part.

Next we turn to the second part. We use the ε-net strategy. To be precise, we construct a 1/2-net $\mathcal{N}$ of the unit sphere $\mathcal{S}^{p-1}$. By Lemma 5.2 of Vershynin (2012), the cardinality of $\mathcal{N}$ is less than 5^p. Lemma 5.3 of

Vershynin (2012) entails that $\|\mathbf{z}\| \leq 2\max_{\mathbf{e}\in\mathcal{N}}|\mathbf{e}^{\mathrm{T}}\mathbf{z}|$. Consequently,

$$\begin{aligned} \mathrm{pr}\left(\|\mathbf{z}\| \geq t\right) &\leq \mathrm{pr}\Big(\max_{\mathbf{e}\in\mathcal{N}}|\mathbf{e}^{\mathrm{T}}\mathbf{z}| \geq t/2\Big) \leq \sum_{\mathbf{e}\in\mathcal{N}} \mathrm{pr}\left(|\mathbf{e}^{\mathrm{T}}\mathbf{z}| \geq t/2\right) \\ &\leq 5^p \sup_{\mathbf{e}\in\mathcal{S}^{p-1}} \mathrm{pr}(|\mathbf{e}^{\mathrm{T}}\mathbf{z}| \geq t/2). \end{aligned}$$

The proof for the second part is completed. □

Lemma 2. Assume condition (A2). Then

$$\sup_{y\in\mathbb{R}} \|\mathbf{m}(y)\mathbf{m}^{\mathrm{T}}(y)\| \leq c_0 \text{ and } \|\mathbf{\Lambda}\| \leq c_0.$$

Proof of Lemma 2: Let $\mathbf{e}$ be any unit-length vector. By Jensen's inequality,

$$\mathbf{e}^{\mathrm{T}}\mathbf{m}(y)\mathbf{m}^{\mathrm{T}}(y)\mathbf{e} \leq E\{(\mathbf{e}^{\mathrm{T}}\mathbf{x})^2 I(Y \leq y)\} \leq E\{(\mathbf{e}^{\mathrm{T}}\mathbf{x})^2\}.$$

By definition, $\|\mathbf{m}(y)\mathbf{m}^{\mathrm{T}}(y)\| = \sup_{\mathbf{e}}\{\mathbf{e}^{\mathrm{T}}\mathbf{m}(y)\mathbf{m}^{\mathrm{T}}(y)\mathbf{e}\}$ and $\|\mathbf{\Sigma}\| = \sup_{\mathbf{e}} E\{(\mathbf{e}^{\mathrm{T}}\mathbf{x})^2\}$. The first part of Lemma 1 follows from Condition (A2) immediately.

Again, $\|\mathbf{\Lambda}\| = \sup E\{\mathbf{e}^{\mathrm{T}}\mathbf{m}(Y)\mathbf{m}^{\mathrm{T}}(Y)\mathbf{e}\} \leq \|\mathbf{\Sigma}\| \leq c_0$. The proof is completed. □

Lemma 3. Assume condition (A3). Then, for any $t \geq 0$,

$$\begin{aligned} \mathrm{pr}\left(\|\overline{\mathbf{x}}\ \overline{\mathbf{x}}^{\mathrm{T}}\|_\infty \geq t\right) &\leq p\exp(1 - Cnt), \text{ and} \\ \mathrm{pr}\left(\|\overline{\mathbf{x}}\ \overline{\mathbf{x}}^{\mathrm{T}}\| \geq t\right) &\leq 5^p \exp(1 - Cnt). \end{aligned}$$

Proof of Lemma 3: Let $\mathbf{e}$ be any unit-length vector. Note that $\{(\mathbf{e}^{\mathrm{T}}\mathbf{x}_i), i = 1, \ldots, n\}$ are independent and sub-Gaussian. By Hoeffding type inequality

(Vershynin, 2012, Proposition 5.10),

$$\mathrm{pr}(|\mathbf{e}^{\mathrm{T}}\overline{\mathbf{x}}| \geq t) \leq \exp(1 - C_0 n t^2), \tag{A.1}$$

where $C_0 > 0$ which does not depend upon $\mathbf{e}$. By Lemma 1,

$$\begin{aligned} \mathrm{pr}(\|\overline{\mathbf{x}}\|_\infty \geq t) &\leq p \ \exp(1 - C_0 n t^2), \text{ and} \\ \mathrm{pr}(\|\overline{\mathbf{x}}\| \geq t) &\leq 5^p \ \exp(1 - C_0 n t^2). \end{aligned}$$

The proof is completed by using $\|\overline{\mathbf{x}}\ \overline{\mathbf{x}}^{\mathrm{T}}\|_\infty = \|\overline{\mathbf{x}}\|_\infty^2$ and $\|\overline{\mathbf{x}}\ \overline{\mathbf{x}}^{\mathrm{T}}\| = \|\overline{\mathbf{x}}\|^2$. □

Lemma 4. Assume conditions (A2)-(A3). For any $y \in \mathbb{R}$ and $t \geq 0$,

$$\begin{aligned} \mathrm{pr}\left\{\|\widehat{\mathbf{m}}(y) - \mathbf{m}(y)\|_\infty \geq t\right\} &\leq p \cdot \exp(2 - Cnt^2), \text{ and} \\ \mathrm{pr}\left\{\|\widehat{\mathbf{m}}(y) - \mathbf{m}(y)\| \geq t\right\} &\leq 5^p \cdot \exp(2 - Cnt^2). \end{aligned}$$

Proof of Lemma 4: For notational clarity we assumed $E(\mathbf{x}) = \mathbf{0}$. By definition,

$$\widehat{\mathbf{m}}(y) - \mathbf{m}(y) = n^{-1}\sum_{i=1}^{n}\left[\mathbf{x}_i I(Y_i \leq y) - E\{\mathbf{x}I(Y \leq y)\}\right] - \overline{\mathbf{x}}\left\{n^{-1}\sum_{i=1}^{n} I(Y_i \leq y)\right\}.$$

Since $\mathbf{x}$ is sub-Gaussian and $|(\mathbf{e}^{\mathrm{T}}\mathbf{x})I(Y \leq y)| \leq |(\mathbf{e}^{\mathrm{T}}\mathbf{x})|$, $\{\mathbf{x}I(Y \leq y)\}$ must also be sub-Gaussian for any unit-length vector $\mathbf{e}$ and any fixed y. Invoking Proposition 5.10 of Vershynin (2012) again, we have

$$\mathrm{pr}\left[\left|n^{-1}\sum_{i=1}^{n}\{\mathbf{e}^{\mathrm{T}}\mathbf{x}_i I(Y_i \leq y)\} - E\left\{\mathbf{e}^{\mathrm{T}}\mathbf{x}I(Y \leq y)\right\}\right| \geq t\right] \leq \exp(1 - C_0 n t^2). \tag{A.2}$$

In addition,

$$\mathrm{pr}\left[\left|\mathbf{e}^{\mathrm{T}}\bar{\mathbf{x}}\left\{n^{-1}\sum_{i=1}^{n}I(Y_i\leq y)\right\}\right|\geq t\right]\leq \mathrm{pr}\left(|\mathbf{e}^{\mathrm{T}}\bar{\mathbf{x}}|\geq t\right)\leq \exp(1-C_0nt^2). \tag{A.3}$$

The second inequality in (A.3) follows from (A.1). Combining (A.2) and (A.3), for any unit length vector $\mathbf{e}$, we obtain

$$\mathrm{pr}\left\{|\mathbf{e}^{\mathrm{T}}\widehat{\mathbf{m}}(y)-\mathbf{e}^{\mathrm{T}}\mathbf{m}(y)|\geq t\right\}\leq 2\exp(1-Cnt^2)$$

Therefore, by using Lemma 1, for any $y\in\mathbb{R}$,

$$\begin{aligned}\mathrm{pr}\left\{\|\widehat{\mathbf{m}}(y)-\mathbf{m}(y)\|_\infty\geq t\right\} &\leq 2p\cdot\exp(1-Cnt^2),\ \text{and}\\ \mathrm{pr}\left\{\|\widehat{\mathbf{m}}(y)-\mathbf{m}(y)\|\geq t\right\} &\leq 2\cdot 5^p\cdot\exp(1-Cnt^2).\end{aligned}$$

The proof is completed. □

LEMMA 5. Assume conditions (A1)-(A3). Write $\widetilde{\mathbf{\Lambda}}=n^{-1}\sum_{i=1}^{n}\mathbf{m}(Y_i)\mathbf{m}^{\mathrm{T}}(Y_i)$, $\widehat{\mathbf{\Lambda}}=n^{-1}\sum_{i=1}^{n}\widehat{\mathbf{m}}(Y_i)\widehat{\mathbf{m}}^{\mathrm{T}}(Y_i)$, and $\mathbf{\Lambda}=E\left\{\mathbf{m}(Y)\mathbf{m}^{\mathrm{T}}(Y)\right\}$. Then

$$\begin{aligned}\mathrm{pr}(\|\widehat{\mathbf{\Lambda}}-\widetilde{\mathbf{\Lambda}}\|_\infty\geq t^2+2c_0^{1/2}t) &\leq \exp(2+\log n+\log p-Cnt^2),\ \text{and}\\ \mathrm{pr}(\|\widehat{\mathbf{\Lambda}}-\widetilde{\mathbf{\Lambda}}\|\geq t^2+2c_0^{1/2}t) &\leq \exp(2+\log n+p\log 5-Cnt^2).\end{aligned}$$

Proof of Lemma 5: We deal with the first part in details and sketch the proof for the second part briefly. Recall the definition of $\widehat{\mathbf{\Lambda}}$ and $\widetilde{\mathbf{\Lambda}}$. We have

$$\widehat{\mathbf{\Lambda}}-\widetilde{\mathbf{\Lambda}}=n^{-1}\sum_{i=1}^{n}\left\{\widehat{\mathbf{m}}(Y_i)\widehat{\mathbf{m}}^{\mathrm{T}}(Y_i)-\mathbf{m}(Y_i)\mathbf{m}^{\mathrm{T}}(Y_i)\right\}.$$

Note that

$$\begin{aligned}\|\widehat{\mathbf{m}}(y)\widehat{\mathbf{m}}^{\mathrm{T}}(y)-\mathbf{m}(y)\mathbf{m}^{\mathrm{T}}(y)\|_\infty &\le \|\widehat{\mathbf{m}}(y)-\mathbf{m}(y)\|_\infty^2+2\|\mathbf{m}(y)\|_\infty\|\widehat{\mathbf{m}}(y)-\mathbf{m}(y)\|_\infty\\ &\le \|\widehat{\mathbf{m}}(y)-\mathbf{m}(y)\|_\infty^2+2c_0^{1/2}\|\widehat{\mathbf{m}}(y)-\mathbf{m}(y)\|_\infty.\end{aligned}$$

By Lemma 4, for any $t \ge 0$, it follows that

$$\begin{aligned}\mathrm{pr}\left\{\|\widehat{\mathbf{m}}(y)\widehat{\mathbf{m}}^{\mathrm{T}}(y)-\mathbf{m}(y)\mathbf{m}^{\mathrm{T}}(y)\|_\infty \ge t^2+2c_0^{1/2}t\right\} &\le \mathrm{pr}\left\{\|\widehat{\mathbf{m}}(y)-\mathbf{m}(y)\|_\infty \ge t\right\}\\ &\le p\cdot\exp(2-Cnt^2),\end{aligned}$$

We notice that the right hand sides of the above display do not depend upon y. Thus

$$\begin{aligned}&\mathrm{pr}\left\{\|\widehat{\mathbf{m}}(Y_k)\widehat{\mathbf{m}}^{\mathrm{T}}(Y_k)-\mathbf{m}(Y_k)\mathbf{m}^{\mathrm{T}}(Y_k)\|_\infty \ge t^2+2c_0^{1/2}t\right\} \le \mathrm{pr}\left\{\|\widehat{\mathbf{m}}(Y_k)-\mathbf{m}(Y_k)\|_\infty \ge t\right\}\\ = \;& E\Big[\mathrm{pr}\left\{\|\widehat{\mathbf{m}}(Y_k)-\mathbf{m}(Y_k)\|_\infty \ge t\right\} \mid Y_k\Big] \le p\cdot\exp(2-Cnt^2).\end{aligned}$$

Consequently,

$$\begin{aligned}\mathrm{pr}(\|\widehat{\mathbf{\Lambda}}-\widetilde{\mathbf{\Lambda}}\|_\infty \ge t^2+2c_0^{1/2}t) &\le \sum_{k=1}^n \mathrm{pr}\left\{\|\widehat{\mathbf{m}}(Y_k)\widehat{\mathbf{m}}^{\mathrm{T}}(Y_k)-\mathbf{m}(Y_k)\mathbf{m}^{\mathrm{T}}(Y_k)\|_\infty \ge t^2+2c_0^{1/2}t\right\}\\ &\le np\cdot\exp(2-Cnt^2).\end{aligned}$$

The proof of the first part is completed.

Next we turn to the second part. Note that

$$\begin{aligned}\|\widehat{\mathbf{m}}(y)\widehat{\mathbf{m}}^{\mathrm{T}}(y)-\mathbf{m}(y)\mathbf{m}^{\mathrm{T}}(y)\| &\le \|\widehat{\mathbf{m}}(y)-\mathbf{m}(y)\|^2+2\|\mathbf{m}(y)\|\|\widehat{\mathbf{m}}(y)-\mathbf{m}(y)\|\\ &\le \|\widehat{\mathbf{m}}(y)-\mathbf{m}(y)\|^2+2c_0^{1/2}\|\widehat{\mathbf{m}}(y)-\mathbf{m}(y)\|\end{aligned}$$

and

$$\begin{aligned}\mathrm{pr}(\|\widehat{\mathbf{m}}(y)\widehat{\mathbf{m}}^{\mathrm{T}}(y)-\mathbf{m}(y)\mathbf{m}^{\mathrm{T}}(y)\|\geq t^2+2c_0^{1/2}t) &\leq \mathrm{pr}\left\{\|\widehat{\mathbf{m}}(y)-\mathbf{m}(y)\|\geq t\right\}\\ &\leq 5^p\cdot\exp(2-Cnt^2).\end{aligned}$$

Following similar arguments, we can show that

$$\mathrm{pr}\left\{\|\widehat{\mathbf{m}}(Y_k)\widehat{\mathbf{m}}^{\mathrm{T}}(Y_k)-\mathbf{m}(Y_k)\mathbf{m}^{\mathrm{T}}(Y_k)\|\geq t^2+2c_0^{1/2}t\right\} \leq 5^p\cdot\exp(2-Cnt^2).$$

Therefore,

$$\begin{aligned}\mathrm{pr}(\|\widehat{\mathbf{\Lambda}}-\widetilde{\mathbf{\Lambda}}\|\geq t^2+2c_0^{1/2}t) &\leq \sum_{k=1}^{n}\mathrm{pr}\left\{\|\widehat{\mathbf{m}}(Y_k)\widehat{\mathbf{m}}^{\mathrm{T}}(Y_k)-\mathbf{m}(Y_k)\mathbf{m}^{\mathrm{T}}(Y_k)\|\geq t^2+2c_0^{1/2}t\right\}\\ &\leq 5^p n\cdot\exp(2-Cnt^2).\end{aligned}$$

The proof is completed. □

S4. Proof of Theorem 1

With the spectral decomposition, we have

$$\mathbf{\Sigma}^{-1}\mathbf{\Lambda}\mathbf{\Sigma}^{-1}=\sum_{k=1}^{d}\lambda_k\mathbf{u}_k\mathbf{u}_k^{\mathrm{T}} \text{ and } \widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Sigma}}^{-1}=\sum_{k=1}^{p}\widehat{\lambda}_k\widehat{\mathbf{u}}_k\widehat{\mathbf{u}}_k^{\mathrm{T}},$$

where $c_0^3\geq\lambda_1\geq\cdots\geq\lambda_d\geq c_0^{-3}$, $\lambda_{d+1}=\cdots=\lambda_p=0$, and $\widehat{\lambda}_1\geq\cdots\geq\widehat{\lambda}_p\geq 0$. The projection matrices are $\mathbf{P}=\sum_{k=1}^{d}\mathbf{u}_k\mathbf{u}_k^{\mathrm{T}}$ and $\widehat{\mathbf{P}}=\sum_{k=1}^{d}\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k$, respectively. We know $c_0^{-1}\leq\lambda(\mathbf{\Sigma})\leq c_0$ and $c_0^{-1}\leq\lambda_d(\mathbf{\Lambda}),\lambda_1(\mathbf{\Lambda})\leq c_0$ by conditions (A2)-(A3) and Lemma 2. By Theorem 2 of Yu et al. (2015),

$$\|\mathbf{P}-\widehat{\mathbf{P}}\|\leq\|\mathbf{P}-\widehat{\mathbf{P}}\|_F\leq 4c_0^3d^{1/2}\|\mathbf{\Sigma}^{-1}\mathbf{\Lambda}\mathbf{\Sigma}^{-1}-\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Sigma}}^{-1}\|.$$

Therefore, it suffices to show

$$\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\| = O_p\left\{(p/n)^{1/2}\right\} \text{ and } \|\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\| = O_p\left\{(\max(p,\log n)/n)^{1/2}\right\}.$$

Recall

$$\widetilde{\mathbf{\Sigma}} = n^{-1}\sum_{i=1}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} \text{ and } \widetilde{\mathbf{\Lambda}} = n^{-1}\sum_{i=1}^{n}\mathbf{m}(Y_i)\mathbf{m}^{\mathrm{T}}(Y_i),$$

where $\mathbf{x}$ and $\mathbf{m}(Y)$ are sub-Gaussian. By Theorem 5.39 and Remark 5.40 of Vershynin (2012), we have

$$\begin{aligned}
\mathrm{pr}\left\{\|\widetilde{\mathbf{\Sigma}}-\mathbf{\Sigma}\| \le \max(\delta,\delta^2)\right\} &\ge 1-2\exp(-c_1t^2),\ \delta = C(p/n)^{1/2}+(t/n^{1/2}),\ \text{and}\\
\mathrm{pr}\left\{\|\widetilde{\mathbf{\Lambda}}-\mathbf{\Lambda}\| \le \max(\delta,\delta^2)\right\} &\ge 1-2\exp(-c_2t^2),\ \delta = C(p/n)^{1/2}+(t/n^{1/2}),
\end{aligned}$$

for any $t \ge 0$. Setting $t = C_1(p^{1/2})$ for sufficiently large C_1 yields $\|\widetilde{\mathbf{\Sigma}}-\mathbf{\Sigma}\| = O_p\left\{(p/n)^{1/2}\right\}$ and $\|\widetilde{\mathbf{\Lambda}}-\mathbf{\Lambda}\| = O_p\left\{(p/n)^{1/2}\right\}$. Setting $t = C_1(p/n)$ for sufficiently large C_1 in Lemma 3 yields that $\|\widehat{\mathbf{\Sigma}}-\widetilde{\mathbf{\Sigma}}\| = \|\overline{\mathbf{x}}\,\overline{\mathbf{x}}^{\mathrm{T}}\| = O_p(p/n)$. Similarly, setting $t = C_1\{(\max(p,\log n)/n)^{1/2}\}$ in Lemma 5 for sufficiently large C_1 entails that $\|\widehat{\mathbf{\Lambda}}-\widetilde{\mathbf{\Lambda}}\| = O_p\left\{(\max(p,\log n)/n)^{1/2}\right\}$. We combine the above results and use the triangle-inequality to obtain $\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\| = O_p\left\{(p/n)^{1/2}\right\}$ and $\|\widehat{\mathbf{\Lambda}}-\mathbf{\Lambda}\| = O_p\left\{\max(p,\log n)/n)^{1/2}\right\}$. The proof is now completed. $\square$

S5. Proof of Theorem 2

Write $\mathbf{\Lambda} = (\mathbf{\Lambda}_{k,l})_{p\times p}$. For the (k,l)-th element of $(\widetilde{\mathbf{\Lambda}} - \mathbf{\Lambda})$,

$$(\widetilde{\mathbf{\Lambda}} - \mathbf{\Lambda})_{k,l} = n^{-1}\sum_{i=1}^{n}\left\{m_k(Y_i)m_l(Y_i) - \mathbf{\Lambda}_{k,l}\right\}.$$

Note that $\{m_k(Y_i)m_l(Y_i) - \mathbf{\Lambda}_{k,l}\}$, for $i = 1,\ldots,n$, are independent centered sub-exponential random variables. By Bernstein inequality (Vershynin, 2012, Proposition 5.16),

$$\text{pr}\left[\left|n^{-1}\sum_{i=1}^{n}\{m_k(Y_i)m_l(Y_i) - \mathbf{\Lambda}_{k,l}\}\right| \geq t\right] \leq \exp\{1 - Cn\min(t^2, t)\}.$$

It follows immediately that

$$\begin{aligned}\text{pr}(\|\mathbf{\Lambda} - \widetilde{\mathbf{\Lambda}}\|_\infty \geq t) &\leq \sum_{k=1}^{p}\sum_{l=1}^{p}\text{pr}\left[\left|n^{-1}\sum_{i=1}^{n}\{m_k(Y_i)m_l(Y_i) - \mathbf{\Lambda}_{k,l}\}\right| \geq t\right]\\ &\leq \exp\{1 + 2\log p - Cn\min(t^2, t)\}.\end{aligned}$$

Setting $t = C(\log p/n)^{1/2}$ for sufficient large $C > 0$ yields

$$\text{pr}\left\{\|\mathbf{\Lambda} - \widetilde{\mathbf{\Lambda}}\|_\infty \geq C(\log p/n)^{1/2}\right\} \to 0. \tag{C.1}$$

It follows from Lemma 5 that

$$\text{pr}\left\{\|\widehat{\mathbf{\Lambda}} - \widetilde{\mathbf{\Lambda}}\|_\infty \geq C(\log p/n)^{1/2}\right\} \to 0, \text{for some } C > 0. \tag{C.2}$$

Combining the results of (C.1) and (C.2) entails that

$$\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\|_\infty = O_p\left\{(\log p/n)^{1/2}\right\}. \tag{C.3}$$

Since

$$\widehat{\mathbf{\Theta}} - \mathbf{\Theta} = \widehat{\mathbf{\Omega}}_s\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Omega}}_s - \mathbf{\Omega}\mathbf{\Lambda}\mathbf{\Omega} = (\widehat{\mathbf{\Omega}}_s - \mathbf{\Omega})\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Omega}}_s + \mathbf{\Omega}\widehat{\mathbf{\Lambda}}(\widehat{\mathbf{\Omega}}_s - \mathbf{\Omega}) + \mathbf{\Omega}(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda})\mathbf{\Omega},$$

we have

$$\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_\infty \le \|\widehat{\mathbf{\Omega}}_s - \mathbf{\Omega}\|_1\|\widehat{\mathbf{\Lambda}}\|_\infty(\|\widehat{\mathbf{\Omega}}_s\|_1 + \|\mathbf{\Omega}\|_1) + \|\mathbf{\Omega}\|_1^2\|\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda}\|_\infty,$$

where we used the inequality $\|\mathbf{AB}\|_\infty \le \|\mathbf{A}\|_1\|\mathbf{B}\|_\infty$ for symmetric matrix $\mathbf{A}$ and arbitrary matrix $\mathbf{B}$. By the proof of Theorem 6 of Cai et al. (2011), $\|\widehat{\mathbf{\Omega}}_s\|_1 \le \|\mathbf{\Omega}\|_1 \le c_0$, and $\|\widehat{\mathbf{\Omega}}_s - \mathbf{\Omega}\|_1 = O_p\left\{s_1(p)(\log p/n)^{(1-q)/2}\right\}$.

Therefore, we have

$$\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_\infty = O_p\left\{s_1(p)(\log p/n)^{(1-q)/2} + (\log p/n)^{1/2}\right\} = O_p\left\{s_1(p)(\log p/n)^{(1-q)/2}\right\}.$$

Write $t_n = \|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_\infty$ and set the tuning parameter $\lambda_{2n} = 2t_n$. Decompose $\mathbf{\Theta}_{k,l}$ as $\mathbf{\Theta}_{k,l} = \mathbf{\Theta}_{k,l}I(|\mathbf{\Theta}_{k,l}| \ge 2t_n) + \mathbf{\Theta}_{k,l}I(|\mathbf{\Theta}_{k,l}| \le 2t_n)$ and

$$\widehat{\mathbf{\Theta}}_{k,l}I(|\widehat{\mathbf{\Theta}}_{k,l}| \ge 2t_n) = \big(\widehat{\mathbf{\Theta}}_{k,l} - \mathbf{\Theta}_{k,l}\big)I(|\widehat{\mathbf{\Theta}}_{k,l}| \ge 2t_n) + \mathbf{\Theta}_{k,l}I(|\widehat{\mathbf{\Theta}}_{k,l}| \ge 2t_n).$$

It follows immediately that

$$\begin{aligned}
&\max_k \sum_{l=1}^p \left|\widehat{\mathbf{\Theta}}_{k,l}I(|\widehat{\mathbf{\Theta}}_{k,l}| \ge \lambda_{2n}) - \mathbf{\Theta}_{k,l}\right| \\
\le\ &\max_k \sum_{l=1}^p \left|\widehat{\mathbf{\Theta}}_{k,l}I(|\widehat{\mathbf{\Theta}}_{k,l}| \ge 2t_n) - \mathbf{\Theta}_{k,l}I(|\mathbf{\Theta}_{k,l}| \ge 2t_n)\right| + \max_k \sum_{l=1}^p |\mathbf{\Theta}_{k,l}|I(|\mathbf{\Theta}_{k,l}| \le 2t_n) \\
\le\ &\max_k \sum_{l=1}^p |\widehat{\mathbf{\Theta}}_{k,l} - \mathbf{\Theta}_{k,l}|I(|\widehat{\mathbf{\Theta}}_{k,l}| \ge 2t_n) + \max_k \sum_{l=1}^p |\mathbf{\Theta}_{k,l}|\left|I(|\widehat{\mathbf{\Theta}}_{k,l}| \ge 2t_n) - I(|\mathbf{\Theta}_{k,l}| \ge 2t_n)\right| \\
&+ \max_k \sum_{l=1}^p |\mathbf{\Theta}_{k,l}|I(|\mathbf{\Theta}_{k,l}| \le 2t_n).
\end{aligned}$$

On one hand, $|\widehat{\mathbf{\Theta}}_{k,l} - \mathbf{\Theta}_{k,l}| I(|\widehat{\mathbf{\Theta}}_{k,l}| \geq 2t_n) \leq t_n I(|\mathbf{\Theta}_{k,l}| \geq t_n)$. On the other hand,

$$\begin{aligned}
\left| I(|\widehat{\mathbf{\Theta}}_{k,l}| \geq 2t_n) - I(|\mathbf{\Theta}_{k,l}| \geq 2t_n) \right| &\leq I(|\widehat{\mathbf{\Theta}}_{k,l}| \geq 2t_n > |\mathbf{\Theta}_{k,l}|) + I(|\mathbf{\Theta}_{k,l}| \geq 2t_n > |\widehat{\mathbf{\Theta}}_{k,l}|) \\
&\leq I\left(\left| |\mathbf{\Theta}_{k,l}| - 2t_n \right| \leq \left| |\widehat{\mathbf{\Theta}}_{k,l}| - |\mathbf{\Theta}_{k,l}| \right| \right) \\
&\leq I\left(\left| |\mathbf{\Theta}_{k,l}| - 2t_n \right| \leq |\widehat{\mathbf{\Theta}}_{k,l} - \mathbf{\Theta}_{k,l}| \right) \leq I(|\mathbf{\Theta}_{k,l}| \leq 3t_n).
\end{aligned}$$

The above two results yield that

$$\begin{aligned}
& \max_k \sum_{l=1}^p \left| \widehat{\mathbf{\Theta}}_{k,l} I(|\widehat{\mathbf{\Theta}}_{k,l}| \geq \lambda_{2n}) - \mathbf{\Theta}_{k,l} \right| \\
\leq\ & \max_k \sum_{l=1}^p t_n I(|\mathbf{\Theta}_{k,l}| \geq t_n) + \max_k \sum_{l=1}^p |\mathbf{\Theta}_{k,l}| I(|\mathbf{\Theta}_{k,l}| \leq 3t_n) + \max_k \sum_{l=1}^p |\mathbf{\Theta}_{k,l}| I(|\mathbf{\Theta}_{k,l}| \leq 2t_n) \\
\leq\ & \max_k \sum_{l=1}^p t_n I(|\mathbf{\Theta}_{k,l}| \geq t_n) + 2 \max_k \sum_{l=1}^p |\mathbf{\Theta}_{k,l}| I(|\mathbf{\Theta}_{k,l}| \leq 3t_n) \\
\leq\ & \max_k \sum_{l=1}^p t_n^{1-q} |\mathbf{\Theta}_{k,l}|^q + 2 \max_k \sum_{l=1}^p |\mathbf{\Theta}_{k,l}|^q (3t_n)^{1-q} \leq 7 s_2(p) t_n^{1-q}.
\end{aligned}$$

Gershgorin's circle theorem states that, for a symmetric $p \times p$ matrix $\mathbf{A} = (\mathbf{A}_{k,l})_{p \times p}$, its i-th largest principal eigenvalue λ_k satisfies

$$\max_k |\lambda_k(\mathbf{A})| \leq \max_k \sum_{l=1}^p |\mathbf{A}_{k,l}|.$$

Invoking the above Gershgorin's circle theorem, we have,

$$\|\widehat{\mathbf{\Theta}}_s - \mathbf{\Theta}\| \leq \max_k \sum_{l=1}^p |\widehat{\mathbf{\Theta}}_{k,l} I(|\widehat{\mathbf{\Theta}}_{k,l}| \geq \lambda_{2n}) - \mathbf{\Theta}_{k,l}| = O_p\left\{ s_1^{1-q}(p) s_2(p) (\log p/n)^{(1-q)^2/2} \right\}.$$

The proof is completed. $\square$

Acknowledgements

We thank the Editor, an Associate Editor, and two anonymous reviewers for their insightful comments. Wang's research is supported by Shanghai Sailing Program 16YF1405700 and National Natural Science Foundation of China 11701367. Yu's research is supported by the National Natural Science Foundation of China 11571111, the 111 project B14019, the Program of Shanghai Subject Chief Scientist 14XD1401600, and the Shanghai Rising Star Program 16QA1401700. Zhu is the corresponding author and his research is supported by National Natural Science Foundation of China (11731011), Chinese Ministry of Education Project of Key Research Institute of Humanities and Social Sciences at Universities (16JJD910002) and National Youth Top-notch Talent Support Program of China.

References

Cai, T., W. Liu, and X. Luo (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association 106*(494), 594–607.

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing*, pp. 210–268. Cambridge Univ. Press, Cambridge.

Yin, Y., Z. Bai, and P. R. Krishnaiah (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields 78*(4), 509–521.

Yu, Y., T. Wang, and R. Samworth (2015). A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika 102*(2), 315–323.

School of Mathematical Sciences, MOE-LSC, Shanghai Jiao Tong University, Shanghai 200240, China.

E-mail: chengwang@sjtu.edu.cn

Key Laboratory of Advanced Theory and Application in Statistics and Data Science, Ministry of Education and School of Statistics, East China Normal University, Shanghai 200241, China.

E-mail: zyu@stat.ecnu.edu.cn

Institute of Statistics and Big Data and Center for Applied Statistics, Renmin University of China, Beijing 100872, China.

E-mail: zhu.liping@ruc.edu.cn