High-Dimensional Gaussian Copula Regression: Adaptive Estimation and Statistical Inference

T. Tony Cai and Linjun Zhang

University of Pennsylvania

Supplementary Material

This supplementary material includes the proof of auxiliary lemmas in Section 1, and supplementary simulation results in Section 2.

S1 Proof of Auxiliary Lemmas

S1.1 Proof of Lemma 1

Recall that $\sigma^2_{g_1(\boldsymbol{u}_i)} = \mathrm{Var}(g_1(Z;\boldsymbol{u}_i)) = x(\boldsymbol{u}_i)^\top \Sigma_{h_Z} x(\boldsymbol{u}_i)$ and for $i = 1, 2, ..., p$, $H_i = \boldsymbol{u}_i^\top \hat{\Sigma} \boldsymbol{v}_0$. Taylor expansion yields

$$\hat{\Sigma} - \Sigma = \sin(\frac{\pi}{2}\hat{T}) - \sin(\frac{\pi}{2}T) = \cos(\frac{\pi}{2}T) \circ (\hat{T} - T) \cdot \frac{\pi}{2} - \frac{1}{2}\sin(\frac{\pi}{2}\check{T}) \circ \frac{\pi}{2}(\hat{T} - T) \circ \frac{\pi}{2}(\hat{T} - T).$$

It follows

$$\begin{aligned} H_i - E[H_i] =& \boldsymbol{u}_i^\top (\hat{\Sigma} - \Sigma)\boldsymbol{v}_0 \\ =& \frac{\pi}{2}\boldsymbol{u}_i^\top \cos(\frac{\pi}{2}T) \circ (\hat{T} - T)\boldsymbol{v}_0 - \frac{1}{2}\boldsymbol{u}_i^\top \sin(\frac{\pi}{2}\check{T}) \circ \frac{\pi}{2}(\hat{T} - T) \circ \frac{\pi}{2}(\hat{T} - T)\boldsymbol{v}_0. \end{aligned} \tag{S1.1}$$

Let $J_i = \boldsymbol{u}_i^\top \cos(\frac{\pi}{2}T) \circ \hat{T}\boldsymbol{v}_0 = \frac{1}{n(n-1)/2}\sum_{i<i'} g(\boldsymbol{Z}_i, \boldsymbol{Z}_{i'}; \boldsymbol{u}_i)$, which is a U-statistics of order two. Then by the Berry-Essen bound for U statisticsCallaert and Janssen (1978), we have

$$\sup_{x \in \mathbb{R}} |P(\frac{\sqrt{n}(J_i - E[J_i])}{2\sigma_{g_1(\boldsymbol{u}_i)}} \le x) - \Phi(x)| \le C\frac{\eta_g^3}{\sigma^3_{g_1(\boldsymbol{u}_i)}} \cdot \frac{1}{\sqrt{n}}, \tag{S1.2}$$

where $\eta_g^3 = E[|g(\boldsymbol{Z}, \boldsymbol{Z}'; \boldsymbol{u}_i)^3|]$.

To prove Lemma 3, it is sufficient to show that $\frac{\eta_g^3}{\sigma_{g_1}^3}$ is upper bounded by a constant. If we can show this, then apply Lemma 2 and (S1.2) to (S1.1), we would get the desired result.

Therefore we proceed to prove that $\frac{\eta_g^3}{\sigma_{g_1}^3}$ is upper bounded by a constant. Recall that

$$\sigma_{g_1}^2 = \text{Var}(g_1(\boldsymbol{Z};\boldsymbol{u}_i)) = \text{Var}(E[g(\boldsymbol{Z},\boldsymbol{Z}';\boldsymbol{u}_i)|\boldsymbol{Z}]) = \text{Var}(g(\boldsymbol{Z},\boldsymbol{Z}')) - E[\text{Var}(g(\boldsymbol{Z},\boldsymbol{Z}')|\boldsymbol{Z})],$$

$h_Z(\boldsymbol{Z}) = E[h(\boldsymbol{Z},\boldsymbol{Z}')|\boldsymbol{Z}]$, and $\Sigma_{h_Z} = \text{Var}(h_Z(\boldsymbol{Z}))$ for $h_Z(\boldsymbol{Z}) = E[\text{sgn}(\boldsymbol{Z}-\boldsymbol{Z}')\otimes\text{sgn}(\boldsymbol{Z}-\boldsymbol{Z}')|\boldsymbol{Z}] = E[\text{vec}(\hat{T}_{\boldsymbol{Z},\boldsymbol{Z}'})|\boldsymbol{Z}]$, where $\hat{T}_{\boldsymbol{Z},\boldsymbol{Z}'}$ is the Kendall's tau estimator based on only two samples $\{\boldsymbol{Z},\boldsymbol{Z}'\}$.

We start from

$$\begin{aligned}\sigma_{g_1}^2 &= \text{Var}(g_1(\boldsymbol{Z};\boldsymbol{u}_i)) = \text{vec}(\boldsymbol{u}_i\boldsymbol{v}_0^\top \circ \cos(\frac{\pi}{2}T))^\top \cdot \Sigma_{h_Z} \cdot \text{vec}(\boldsymbol{u}_i\boldsymbol{v}_0^\top \circ \cos(\frac{\pi}{2}T)) \\ &\geq M_3||(0,\boldsymbol{u}_i)(1,-\boldsymbol{\beta})^\top \circ \cos(\frac{\pi}{2}T)||_F^2 \geq \frac{M_3}{M_1^2}||\boldsymbol{u}_i||_2^2.\end{aligned}$$

The second inequality uses the fact that for any $i \neq j$,

$$\cos(\frac{\pi}{2}T_{ij}) = \sqrt{1-\Sigma_{ij}^2} = \det(\Sigma_{\{i,j\},\{i,j\}}) \geq \lambda_{\min}(\Sigma_{\{i,j\},\{i,j\}}) \geq 1/M_1.$$

Therefore, we have $\sigma_{g_1}^3 \geq \frac{M_3^{3/2}}{M_1^3}||\boldsymbol{u}||_2^3$. Then we derive the upper bound for η_g^3. Taylor expansion yields $\cos(\frac{\pi}{2}T) = \sum_{k=0}^\infty \binom{1/2}{k}(-1)^k \Sigma \circ_{2k} \Sigma$. Then

$$||\cos(\frac{\pi}{2}T)||_2 \leq \sum_{k=0}^\infty |\binom{1/2}{k}| \cdot ||\Sigma \circ_{2k} \Sigma||_2 \leq \sum_{k=0}^\infty |\binom{1/2}{k}| \cdot ||\Sigma||,$$

where the last inequality comes from Theorem 5.5.18 in Horn and Johnson (1995).

Therefore, $||\cos(\frac{\pi}{2}T)||_2 \leq \sum_{k=0}^\infty |\binom{1/2}{k}| \cdot ||\Sigma|| = 2||\Sigma|| \leq 2M_1$. In addition, since $\boldsymbol{\beta} = \Sigma_{XX}^{-1}\Sigma_{XY}$, we have

$$||\boldsymbol{\beta}||_2 = ||\Sigma_{XX}^{-1}\Sigma_{XY}||_2 \leq M_1||\Sigma_{XY}||_2 \leq M_1^2.$$

This implies

$$\begin{aligned}
\eta_{g_1}^3 &= E[|g(\boldsymbol{Z},\boldsymbol{Z}';\boldsymbol{u}_i)|^3] = E[|\mathrm{sgn}(\boldsymbol{Z}-\boldsymbol{Z}')^\top(\boldsymbol{u}_i\boldsymbol{v}_0^\top \circ \cos(\frac{\pi}{2}T))\mathrm{sgn}(\boldsymbol{Z}-\boldsymbol{Z}')|^3] \\
&= E[|\mathrm{tr}(\boldsymbol{u}_i\boldsymbol{v}_0^\top \mathrm{diag}(\mathrm{sgn}(\boldsymbol{Z}-\boldsymbol{Z}'))\cos(\frac{\pi}{2}T)\mathrm{diag}(\mathrm{sgn}(\boldsymbol{Z}-\boldsymbol{Z}')))|^3] \\
&= E[|\mathrm{tr}(\boldsymbol{v}_0^\top \mathrm{diag}(\mathrm{sgn}(\boldsymbol{Z}-\boldsymbol{Z}'))\cos(\frac{\pi}{2}T)\mathrm{diag}(\mathrm{sgn}(\boldsymbol{X}-\boldsymbol{X}'))\boldsymbol{u}_i)|^3] \\
&= E[|\boldsymbol{v}_0^\top \mathrm{diag}(\mathrm{sgn}(\boldsymbol{Z}-\boldsymbol{Z}'))\cos(\frac{\pi}{2}T)\mathrm{diag}(\mathrm{sgn}(\boldsymbol{Z}-\boldsymbol{Z}'))\boldsymbol{u}_i|^3] \\
&= E[\Big|||\boldsymbol{u}_i||_2||\boldsymbol{v}_0||_2||\cos(\frac{\pi}{2}T)||_2\Big|^3] \\
&\le E[\Big|||\boldsymbol{u}_i||_2 M_1^2 \cdot 2M_1\Big|^3] \\
&\le 8M_1^9||\boldsymbol{u}_i||_2^3.
\end{aligned}$$

The last equality due to the fact that

$$||A||_2 = \sup_{\boldsymbol{u},\boldsymbol{v}_0:||\boldsymbol{u}||=||\boldsymbol{v}_0||=1} \boldsymbol{u}^\top A\boldsymbol{v}_0 = \sup_{\boldsymbol{u},\boldsymbol{v}_0:||\boldsymbol{u}||=||\boldsymbol{v}_0||=1} (\boldsymbol{u}^\top D)A(D\boldsymbol{v}_0) = ||DAD||_2,$$

when diagonal elements of D is either 1 or -1.

It follows

$$\frac{\eta_g^3}{\sigma_{g_1}^3} \le \frac{8M_1^9||\boldsymbol{u}_i||_2^3}{\frac{M_3^{3/2}}{M_1^3}||\boldsymbol{u}_i||_2^3} = \frac{8M_1^{12}}{M_3^{3/2}}.$$

Then with (S1.2) together,

$$\lim_{n\to\infty}\sup_{x\in\mathbb{R}}|P(\frac{\sqrt{n}(H_i-E[H_i])}{2v_{g_1}} \le x) - \Phi(x)| = 0.$$

□

S1.2 Proof of Lemma 2

Since the $1/M_1 \le \lambda_{\min}(\Sigma) \le \lambda_{\max}(\Sigma) < M_1$, $||\Sigma^{-1}|| < M_2$, the third constraint in (3.6):

$$b^{-1}n^{-a} \le ||\boldsymbol{u}||_2 \le ||\boldsymbol{u}||_1 \le bn^{a/2}$$

is feasible when $\boldsymbol{u} = \Sigma_{i,\cdot}^{-1}$ and $a, b > 0$.

Then it is sufficient to show that $(0, \Sigma_{i,\cdot}^{-1})^\top$ satisfies the constraint condition when $\mu = C\sqrt{\frac{\log p}{n}}$ with high probability.

By Lemma 2, with probability at least $1 - p^{-2}$

$$||\hat{\Sigma}_{XX}\Sigma_{\cdot,i}^{-1} - e_i^{(p)}||_\infty = ||\hat{\Sigma}_{XX}\Sigma_{\cdot,i}^{-1} - \Sigma_{XX}\Sigma_{\cdot,i}^{-1}||_\infty$$
$$\leq |\hat{\Sigma}_{XX} - \Sigma_{XX}|_\infty \cdot ||\Sigma_{\cdot,i}^{-1}||_1 \lesssim \sqrt{\frac{\log p}{n}}.$$

In addition, due to the fact that $||\Sigma_{i,\cdot}^{-1}||_1 \leq ||\Sigma^{-1}||_1 \leq M_2$, and $||\Sigma_{i,\cdot}^{-1}||_2 \geq \lambda_{\min}(\Sigma^{-1}) \geq \frac{1}{M_1}$, we concludes that the constraints in the optimization problem (3.6) is feasible with probability at least $1 - p^{-2}$. □

S1.3 Proof of Lemma 3

Recall that $x(\boldsymbol{u}) = \text{vec}(\boldsymbol{u}\boldsymbol{v}_0^\top \circ \cos(\frac{\pi}{2}T))$, $\sigma_{g_1}^2 = x(\boldsymbol{u})^\top \Sigma_{h_Z} x(\boldsymbol{u})$, and

$$\hat{\Sigma}_{h_Z} = \frac{1}{n}\sum_{i=1}^n (\hat{h}_Z(\boldsymbol{Z}_i) - \frac{1}{n}\sum_{i'=1}^n \hat{h}_Z(\boldsymbol{Z}_{i'}))(\hat{h}_Z(\boldsymbol{Z}_i) - \frac{1}{n}\sum_{i'=1}^n \hat{h}_Z(\boldsymbol{Z}_{i'}))^\top,$$

with $\hat{h}_Z(\boldsymbol{Z}_i) = \frac{1}{n-1}\sum_{i \neq i'} \text{sgn}(\boldsymbol{Z}_i - \boldsymbol{Z}_{i'}) \otimes \text{sgn}(\boldsymbol{Z}_i - \boldsymbol{Z}_{i'}) \in \mathbb{R}^{d^2}$.

We would like to prove with high probability

$$|x(\boldsymbol{u})^\top (\hat{\Sigma}_{h_Z} - \Sigma_{h_Z}) x(\boldsymbol{u})| \leq \sqrt{\frac{\log p}{n^{1-2a}}}.$$

By definition, for a random vector $\boldsymbol{Z} = (Z_{(1)}, ..., Z_{(d)})$ with an independent copy $\boldsymbol{Z}'$, and any $j, k \in [d]$, let's use $h_Z(\boldsymbol{Z})_{jk}$ to denote the $[(j-1)d + k]$-th coordinate of $h_Z(\boldsymbol{Z})$

$$h_Z(\boldsymbol{Z})_{jk} = E[\text{sgn}(Z_{(j)} - Z'_{(j)})\text{sgn}(Z_{(k)} - Z'_{(k)})|\boldsymbol{Z}],$$

and

$$\hat{h}_Z(\boldsymbol{Z}_i)_{jk} = \frac{1}{n-1}\sum_{i' \neq i}^n \text{sgn}(Z_{ij} - Z_{i'j})\text{sgn}(Z_{ik} - Z_{i'k}).$$

This implies $\frac{1}{n}\sum_{i=1}^{n}\hat{h}_Z(\boldsymbol{Z}_i)_{jk}=\hat{\tau}_{jk}$. Therefore

$$\begin{aligned}\hat{\Sigma}_{h_Z(jk,\,j_1k_1)} =&\frac{1}{n}\sum_i(\hat{h}_Z(\boldsymbol{Z}_i)_{jk}-\frac{1}{n}\sum_{i'}\hat{h}_Z(\boldsymbol{Z}_{i'})_{jk})(\hat{h}_Z(\boldsymbol{Z}_i)_{j_1,k_1}-\frac{1}{n}\sum_{i'}\hat{h}_Z(\boldsymbol{Z}_{i'})_{j_1,k_1})\\ =&\frac{1}{n}\sum_i(\hat{h}_Z(\boldsymbol{Z}_i)_{jk}-\hat{\tau}_{jk})(\hat{h}_Z(\boldsymbol{Z}_i)_{j_1,k_1}-\hat{\tau}_{j_1k_1})\\ =&\frac{1}{n}\sum_i\hat{h}_Z(\boldsymbol{Z}_i)_{jk}\hat{h}_Z(\boldsymbol{Z}_i)_{j_1,k_1}-\hat{\tau}_{jk}\hat{\tau}_{j_1k_1}.\end{aligned}$$

It follows

$$\begin{aligned}x(\boldsymbol{u})^\top\hat{\Sigma}_{h_Z}x(\boldsymbol{u}) =&\frac{1}{n}\sum_i x(\boldsymbol{u})^\top\Big(\hat{h}_Z(\boldsymbol{Z}_i)-\frac{1}{n}\sum_{i'}\hat{h}_Z(\boldsymbol{Z}_{i'})\Big)\Big(\hat{h}_Z(\boldsymbol{Z}_i)-\frac{1}{n}\sum_{i'}\hat{h}_Z(\boldsymbol{Z}_{i'})\Big)^\top x(\boldsymbol{u})\\ =&\frac{1}{n}\sum_i x(\boldsymbol{u})^\top\hat{h}_Z(\boldsymbol{Z}_i)\hat{h}_Z(\boldsymbol{Z}_i)^\top x(\boldsymbol{u})-x(\boldsymbol{u})^\top\mathrm{vec}(\hat{T})\mathrm{vec}(\hat{T})^\top x(\boldsymbol{u})\\ =&\frac{1}{n}\sum_i[x(\boldsymbol{u})^\top\hat{h}_Z(\boldsymbol{Z}_i)]^2-[x(\boldsymbol{u})^\top\mathrm{vec}(\hat{T})]^2.\end{aligned}$$

Since

$$x(\boldsymbol{u})^\top\hat{h}_Z(\boldsymbol{Z}_i)=\frac{1}{n-1}\sum_{i\neq i'}x(\boldsymbol{u})^\top\Big(\mathrm{sgn}(\boldsymbol{Z}_i-\boldsymbol{Z}_{i'})\otimes\mathrm{sgn}(\boldsymbol{Z}_i-\boldsymbol{Z}_{i'})\Big).$$

Conditional on $\boldsymbol{Z}_i$, $x(\boldsymbol{u})^\top\mathrm{sgn}(\boldsymbol{Z}_i-\boldsymbol{Z}_{i'})\otimes\mathrm{sgn}(\boldsymbol{Z}_i-\boldsymbol{Z}_{i'})$ are $n-1$ *i.i.d* random vectors. In addition, similar as the proof in Lemma 3,

$$\begin{aligned}|x(\boldsymbol{u})^\top\mathrm{sgn}(\boldsymbol{Z}_i-\boldsymbol{Z}_{i'})\otimes\mathrm{sgn}(\boldsymbol{Z}_i-\boldsymbol{Z}_{i'})| =& |\mathrm{sgn}(\boldsymbol{Z}-\boldsymbol{Z}')^\top(\boldsymbol{u}\boldsymbol{v}_0^\top\circ\cos(\frac{\pi}{2}T))\mathrm{sgn}(\boldsymbol{Z}-\boldsymbol{Z}')|\\ =&|\mathrm{tr}(\boldsymbol{u}\boldsymbol{v}_0^\top\mathrm{diag}(\mathrm{sgn}(\boldsymbol{Z}-\boldsymbol{Z}'))\cos(\frac{\pi}{2}T)diag(\mathrm{sgn}(\boldsymbol{Z}-\boldsymbol{Z}')))|\\ =&|\boldsymbol{v}_0^\top\mathrm{diag}(\mathrm{sgn}(\boldsymbol{Z}-\boldsymbol{Z}'))\cos(\frac{\pi}{2}T)\mathrm{diag}(\mathrm{sgn}(\boldsymbol{Z}-\boldsymbol{Z}'))\boldsymbol{u}|\\ \leq&||\boldsymbol{v}_0||_2||\cos(\frac{\pi}{2}T)||_2||\boldsymbol{u}||_2\leq 2M_1^3||\boldsymbol{u}||_2.\end{aligned}$$

Therefore, by Hoeffding inequality,

$$P(|x(\boldsymbol{u})^\top\hat{h}_Z(\boldsymbol{Z}_i)-x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i)|>t|Z_i)\leq e^{-\frac{nt^2}{4M_1^3||\boldsymbol{u}||_2}}.$$

This implies

$$E[e^{t(x(\boldsymbol{u})^\top\hat{h}_Z(\boldsymbol{Z}_i)-x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i))}]\leq e^{\frac{4M_1^3||\boldsymbol{u}||_2t^2}{n}}.$$

Therefore the sub-gaussian norm of $x(\boldsymbol{u})^\top \hat{h}_Z(\boldsymbol{Z}_i) - x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i)$,

$$||x(\boldsymbol{u})^\top \hat{h}_Z(\boldsymbol{Z}_i)) - x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i)||_{\psi_2} \leq \frac{4M_1^3||\boldsymbol{u}||_2}{n},$$

and this implies

$$\begin{aligned}
&||\frac{1}{n}\sum_{i=1}^n x(\boldsymbol{u})^\top \hat{h}_Z(\boldsymbol{Z}_i) - x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i)||_{\psi_2} \\
\leq &\frac{1}{n}\sum_{i=1}^n ||x(\boldsymbol{u})^\top \hat{h}_Z(\boldsymbol{Z}_i) - x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i)||_{\psi_2} \leq \frac{4M_1^3||\boldsymbol{u}||_2}{n}.
\end{aligned}$$

Therefore

$$P(|\frac{1}{n}\sum_{i=1}^n x(\boldsymbol{u})^\top \hat{h}_Z(\boldsymbol{Z}_i) - x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i)| > t) \leq e^{-\frac{nt^2}{4M^3||\boldsymbol{u}||_2}}.$$

Similarly, by Hoeffding inequality,

$$\begin{aligned}
P(|x(\boldsymbol{u})^\top \mathrm{vec}(\hat{T}) - x(\boldsymbol{u})^\top \mathrm{vec}(T)| > t) \leq &e^{-\frac{nt^2}{4M_1^3||\boldsymbol{u}||_2}}, \\
P(|\frac{1}{n}\sum_{i=1}^n (x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i))^2 - E[(x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i))^2]| > t) \leq &e^{-\frac{nt^2}{8M_1^6||\boldsymbol{u}||_2^2}}.
\end{aligned}$$

It follows with probability at least $1 - 4e^{-\frac{nt^2}{4M_1^3||\boldsymbol{u}||_2}} - e^{-\frac{nt^2}{8M_1^6||\boldsymbol{u}||_2^2}}$,

$$\begin{aligned}
&|x(\boldsymbol{u})^\top (\hat{\Sigma}_{h_Z} - \Sigma_{h_Z}) x(\boldsymbol{u})| \\
=&|\frac{1}{n}\sum_i [x(\boldsymbol{u})^\top \hat{h}_Z(\boldsymbol{Z}_i)]^2 - [x(\boldsymbol{u})^\top \mathrm{vec}(\hat{T})]^2 - \frac{1}{n}\sum_i E[(x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i))^2] + [x(\boldsymbol{u})^\top \mathrm{vec}(T)]^2 \\
=&|\frac{1}{n}\sum_i [x(\boldsymbol{u})^\top \hat{h}_Z(\boldsymbol{Z}_i)]^2 - [x(\boldsymbol{u})^\top \mathrm{vec}(\hat{T})]^2 - \frac{1}{n}\sum_i (x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i))^2 + \frac{1}{n}\sum_i (x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i))^2 \\
&- \frac{1}{n}\sum_i E[(x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i))^2] + [x(\boldsymbol{u})^\top \mathrm{vec}(T)]^2| \\
\leq&|\frac{1}{n}\sum_i (x(\boldsymbol{u})^\top \hat{h}_Z(\boldsymbol{Z}_i))^2 - \frac{1}{n}\sum_i (x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i))^2| + |\frac{1}{n}\sum_i (x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i))^2 - \frac{1}{n}\sum_i E[(x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i))^2]| \\
&+ |[x(\boldsymbol{u})^\top \mathrm{vec}(\hat{T})]^2 - [x(\boldsymbol{u})^\top \mathrm{vec}(T)]^2| \\
=&|\frac{1}{n}\sum_i (x(\boldsymbol{u})^\top \hat{h}_Z(\boldsymbol{Z}_i) - x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i))(x(\boldsymbol{u})^\top \hat{h}_Z(\boldsymbol{Z}_i) + x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i))| + |[x(\boldsymbol{u})^\top \mathrm{vec}(\hat{T})]^2 - [x(\boldsymbol{u})^\top \mathrm{vec}(T)]^2| \\
&+ |\frac{1}{n}\sum_i (x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i) - E[(x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i)])((x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i) + E[(x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i)])|.
\end{aligned}$$

In addition,

$$|x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i)| \leq ||\text{vec}(\boldsymbol{u}\boldsymbol{v}_0^\top \circ \cos(\frac{\pi}{2}T))||_1 \leq ||\boldsymbol{u}||_1||\boldsymbol{v}_0||_1 \leq ||\boldsymbol{u}||_1||\boldsymbol{\beta}||_1 \lesssim \sqrt{s}||\boldsymbol{u}||_1.$$

Similarly,

$$|x(\boldsymbol{u})^\top \hat{h}_Z(\boldsymbol{Z}_i)| \lesssim \sqrt{s}||\boldsymbol{u}||_1, |x(\boldsymbol{u})^\top \text{vec}(\hat{T})| \lesssim \sqrt{s}||\boldsymbol{u}||_1, \text{ and } |x(\boldsymbol{u})^\top \text{vec}(T)| \lesssim \sqrt{s}||\boldsymbol{u}||_1.$$

Therefore

$$\begin{aligned}
&|x(\boldsymbol{u})^\top(\hat{\Sigma}_{h_Z} - \Sigma_{h_Z})x(\boldsymbol{u})| \\
\leq &\sqrt{s}||\boldsymbol{u}||_1 \cdot \Big[|\frac{1}{n}\sum_i x(\boldsymbol{u})^\top \hat{h}_Z(\boldsymbol{Z}_i) - x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i)| + |x(\boldsymbol{u})^\top \text{vec}(\hat{T}) - x(\boldsymbol{u})^\top \text{vec}(T)| \\
&+ |\frac{1}{n}\sum_i x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i) - E[(x(\boldsymbol{u})^\top h_Z(\boldsymbol{Z}_i)]|\Big] \\
\leq &3\sqrt{s}||\boldsymbol{u}||_1 t.
\end{aligned}$$

Therefore,

$$P(|x(\boldsymbol{u})^\top(\hat{\Sigma}_{h_Z} - \Sigma_{h_Z})x(\boldsymbol{u})| \gtrsim \sqrt{s}||\boldsymbol{u}||_1 t) \leq 4e^{-\frac{nt^2}{4M_1^3||\boldsymbol{u}||_2}} + e^{-\frac{nt^2}{8M_1^6||\boldsymbol{u}||_2^2}}.$$

Let $t = \sqrt{\frac{8M_1^3 \log p \cdot n^a}{n}}$, and by the fact that $||\boldsymbol{u}||_2 \leq ||\boldsymbol{u}||_1 \leq n^{a/2}$, we have for any $\epsilon > 0$,

$$P(|x(\boldsymbol{u})^\top(\hat{\Sigma}_{h_Z} - \Sigma_{h_Z})x(\boldsymbol{u})| \gtrsim \sqrt{\frac{s \log p}{n^{1-2a}}}) \leq 5p^{-2}.$$

□

S1.4 Proof of Lemma 4

Recall that $x(\boldsymbol{u}) = \text{vec}(\boldsymbol{u}\boldsymbol{v}_0^\top \circ \cos(\frac{\pi}{2}T))$, and $\hat{x}(\boldsymbol{u}) = \text{vec}(\boldsymbol{u}\hat{\boldsymbol{v}}_0{}^\top \circ \cos(\frac{\pi}{2}\hat{T}))$, therefore

$$\begin{aligned}
||x(\boldsymbol{u}) - \hat{x}(\boldsymbol{u})||_1 =& ||\text{vec}(\boldsymbol{u}\boldsymbol{v}_0^\top \circ \cos(\frac{\pi}{2}T)) - \text{vec}(\boldsymbol{u}\hat{\boldsymbol{v}}^\top \circ \cos(\frac{\pi}{2}\hat{T}))||_1 \\
=& |\boldsymbol{u}\boldsymbol{v}_0^\top \circ \cos(\frac{\pi}{2}T) - \boldsymbol{u}\hat{\boldsymbol{v}}^\top \circ \cos(\frac{\pi}{2}\hat{T})||_1 \\
\leq& ||\boldsymbol{u}(\boldsymbol{v}_0 - \hat{\boldsymbol{v}})^\top||_1 \leq ||\boldsymbol{u}||_1||\boldsymbol{v}_0 - \hat{\boldsymbol{v}}||_\infty \\
\leq& ||\boldsymbol{u}||_1||\boldsymbol{v}_0 - \hat{\boldsymbol{v}}||_2 \lesssim n^a\sqrt{\frac{s\log p}{n}}.
\end{aligned}$$

□

S1.5 Proof of Lemma 5

Since $||\boldsymbol{\beta}||_2 = ||\Sigma_{XX}^{-1}\Sigma_{XY}||_2 \geq M_1^{-2}$, then

$$\begin{aligned}
\sigma_{g_1(\boldsymbol{u})}^2 &= \text{Var}(g_1(\boldsymbol{Z};\boldsymbol{u})) = \text{vec}(\boldsymbol{u}\boldsymbol{v}_0^\top \circ \cos(\frac{\pi}{2}T))^\top \cdot \Sigma_{h_Z} \cdot \text{vec}(\boldsymbol{u}\boldsymbol{v}_0^\top \circ \cos(\frac{\pi}{2}T)) \\
&\geq \frac{M_3}{M_1}||(0, \boldsymbol{u}[2:p])(1, -\boldsymbol{\beta})^\top \circ \cos(\frac{\pi}{2}T)||_F^2 \\
&\geq \frac{M_3}{M_1^5}||\boldsymbol{u}||_2^2 \geq \frac{M_3}{M_1^5 n^{2a}}.
\end{aligned}$$

□

S1.6 Proof of Lemma 6

Let $A_1, A_2 \in \mathbb{R}^{d\times 2d}$ with each row of A_i has unit norm, and for some diagonal matrix $D = \text{diag}(m_1, m_2, ..., m_d)$, satisfy

$$\begin{cases}
A_1 P_{A_2}^\perp A_1^\top = D \\
(A_1 + A_2)(A_1 + A_2)^\top = \Sigma \\
\text{rank}(A_1) = \text{rank}(A_2) = d,
\end{cases} \tag{S1.1}$$

where $P_{A_2}^{\perp} = I_{2d} - A_2^{\top}(A_2A_2^{\top})^{-1}A_2$. Therefore, if $\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} \sim N(0, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix})$ with $\Sigma_{11} = A_1A_1^{\top}$, $\Sigma_{22} = A_2A_2^{\top}$, $\Sigma_{12} = A_1A_2^{\top}$, then

$$\Sigma_{11\cdot 2} := \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = D,$$

$$\Sigma_{11} + \Sigma_{12} + \Sigma_{21} + \Sigma_{22} = \Sigma.$$

This implies $\boldsymbol{X}|\boldsymbol{Y} \sim N(\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{Y}, D)$, $\boldsymbol{X}+\boldsymbol{Y} \sim N(0,\Sigma)$. For $\boldsymbol{v} \in \mathbb{R}^d$ with $||\boldsymbol{v}||_2 = 1$,

$$\begin{aligned} Ee^{v^{\top}\mathrm{sgn}(\boldsymbol{Z})} &= Ee^{v^{\top}\mathrm{sgn}(\boldsymbol{X}+\boldsymbol{Y})} = E[E[e^{v^{\top}\mathrm{sgn}(\boldsymbol{X}+\boldsymbol{Y})}|\boldsymbol{Y}] \\ &= E[E[e^{\sum_{i=1}^d v_i \mathrm{sgn}(X_i+Y_i)}|Y] = E[\prod_{i=1}^d E[e^{v_i \mathrm{sgn}(X_i+Y_i)}|\boldsymbol{Y}]]. \end{aligned}$$

We have

$$\mathrm{sgn}(X_i+Y_i)|\boldsymbol{Y} \sim \begin{cases} 1, & \text{with probability } \Phi(\frac{Y_i + e_i^{\top}\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{Y}}{\sqrt{m}_i}), \\ -1, & \text{with probability } 1-\Phi(\frac{Y_i + e_i^{\top}\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{Y}}{\sqrt{m}_i}). \end{cases} \tag{S1.2}$$

Let $h_i(\boldsymbol{Y}) := E[\mathrm{sgn}(X_i+Y_i)|\boldsymbol{Y}] = 2\Phi(\frac{Y_i + e_i^{\top}\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{Y}}{\sqrt{m}_i}) - 1 = 2\Phi(\frac{e_i^{\top}(I+\Sigma_{12}\Sigma_{22}^{-1})\boldsymbol{Y}}{\sqrt{m}_i}) - 1$.

Therefore

$$Ee^{v^{\top}\mathrm{sgn}(\boldsymbol{Z})} = E[\prod_{i=1}^d E[e^{v_i(\mathrm{sgn}(X_i+Y_i) - h_i(\boldsymbol{Y}))}|\boldsymbol{Y}]e^{v_i h_i(\boldsymbol{Y})}] = e^{1/2}E[\prod_{i=1}^d e^{v_i h_i(\boldsymbol{Y})}] = e^{1/2}E[e^{\sum_{i=1}^d v_i h_i(\boldsymbol{Y})}].$$

Let $g(\tilde{\boldsymbol{Y}}) = \sum_{i=1}^d v_i h_i(\Sigma_{22}^{1/2}\tilde{\boldsymbol{Y}})$ where $\tilde{\boldsymbol{Y}} = \Sigma_{22}^{-1/2}\boldsymbol{Y} \sim N_d(0, I)$, then we have

$$\begin{aligned} |g(\tilde{\boldsymbol{Y}}_1) - g(\tilde{\boldsymbol{Y}}_2)| &= |\sum_{i=1}^d v_i(h_i(\Sigma_{22}^{1/2}\tilde{\boldsymbol{Y}}_1) - h_i(\Sigma_{22}^{1/2}\tilde{\boldsymbol{Y}}_2))| \le \sqrt{\sum_{i=1}^d (h_i(\Sigma_{22}^{1/2}\tilde{\boldsymbol{Y}}_1) - h_i(\Sigma_{22}^{1/2}\tilde{\boldsymbol{Y}}_2))^2} \\ &\le \pi ||D^{-1/2}(I+\Sigma_{12}\Sigma_{22}^{-1})\Sigma_{22}^{1/2}|| \cdot ||\tilde{\boldsymbol{Y}}_1 - \tilde{\boldsymbol{Y}}_2|| \\ &\le \pi\sqrt{||D^{-1/2}(I+\Sigma_{12}\Sigma_{22}^{-1})\Sigma_{22}(I+\Sigma_{12}\Sigma_{22}^{-1})^{\top}D^{-1/2}||} \cdot ||\tilde{\boldsymbol{Y}}_1 - \tilde{\boldsymbol{Y}}_2|| \\ &= \pi\sqrt{||D^{-1/2}\Sigma - 1||} \cdot ||\tilde{\boldsymbol{Y}}_1 - \tilde{\boldsymbol{Y}}_2||. \end{aligned}$$

From (1) we know that $0 \le (A_1 + A_2)P_{A_2}^{\perp}(A_1 + A_2)^\top = \Sigma - D$. Then we have

$$|g(\tilde{\boldsymbol{Y}}_1) - g(\tilde{\boldsymbol{Y}}_2)| \le \frac{\lambda_{\max}(\Sigma) - \lambda_{\min}(\Sigma)}{\lambda_{\min}(\Sigma)}\pi||\tilde{\boldsymbol{Y}}_1 - \tilde{\boldsymbol{Y}}_2||.$$

Thus, the Lipschitz norm of $g(\cdot)$ is bounded by $\frac{\lambda_{\max}(\Sigma)-\lambda_{\min}(\Sigma)}{\lambda_{\min}(\Sigma)}\pi$. By the Gaussian concentration inequalityBorell (1975), $E[e^{\sum_{i=1}^d v_i h_i(\boldsymbol{Y})}] = e^{\pi}E[e^{g(\Sigma_{22}^{-1/2}\boldsymbol{Y})}] \le e^{\pi M/2}$ with $M = ||D^{-1}\Sigma||_2$. If we let $D = \lambda_{\min}(\Sigma)I$, then $M = \kappa(\Sigma)$. □

S1.7 Proof of Lemma 7

1. According to Taylor's expansion, $|\hat{\Sigma} - \Sigma|_\infty$ can be bounded by $|\hat{T} - T|_\infty$:

$$\begin{aligned}\hat{\Sigma} - \Sigma &= \sin(\frac{\pi}{2}\hat{T}) - \sin(\frac{\pi}{2}T)\\ &= \cos(\frac{\pi}{2}T) \circ (\hat{T} - T) \cdot \frac{\pi}{2} - \frac{1}{2}\sin(\frac{\pi}{2}) \circ \frac{\pi}{2}(\hat{T} - T) \circ \frac{\pi}{2}(\hat{T} - T).\end{aligned}$$

This implies $|\hat{\Sigma} - \Sigma|_\infty \lesssim |\hat{T} - T|_\infty + |\hat{T} - T|_\infty^2$. By Hoeffding inequality, $P(|T_{jk} - T_{jk}| > t) \le 2\exp(-nt^2/4)$. Therefore,

$$\begin{aligned}P(|\hat{T} - T|_\infty > t) &\le \sum_{j,k=1}^{p} P(|T_{jk} - T_{jk}| > t)\\ &\le 2p^2 \exp(-nt^2/4) = 2\exp(2\log p - nt^2/4).\end{aligned}$$

Let $t = 4\sqrt{\frac{\log p}{n}}$, the above inequality implies that with probability $1 - 2p^{-2}$,

$$|\hat{T} - T|_\infty \lesssim \sqrt{\frac{\log p}{n}}.$$

This shows that with probability $1 - 2p^{-2}$,

$$|\hat{\Sigma} - \Sigma|_\infty \le |\hat{T} - T|_\infty + |\hat{T} - T|_\infty^2 \lesssim \sqrt{\frac{\log p}{n}}.$$

2. Let $d = p+1$, and without loss of generality we assume n is even. For $i, i' \in \{1, 2, ..., n\}$, define $\boldsymbol{S}_{i,i'} = \text{sgn}(\boldsymbol{Z}_i - \boldsymbol{Z}_{i'}) = (\text{sgn}(Z_{i1} - Z_{i'1}), ..., \text{sgn}(Z_{id} - Z_{i'd}))^\top$, and $\hat{\Delta}_{i,i'} = \frac{1}{n(n-1)}(\boldsymbol{S}_{i,i'}\boldsymbol{S}_{i,i'}^\top -$

T). Moreover, for any permutation $\sigma \in S_n$, where S_n is the permutation group of $\{1, ..., n\}$, let $(i_1, ..., i_n) = \sigma(1, ..., n)$. For $r = 1, ..., n/2$ (without loss of generality, we assume n is even), we define $\boldsymbol{S}_r^\sigma$ and $\hat{\Delta}_r^\sigma$ to be $\boldsymbol{S}_r^\sigma = \boldsymbol{S}_{2i_r, 2i_r - 1}$, $\hat{\Delta}_r^\sigma = \frac{1}{n/2}(\boldsymbol{S}_r^\sigma \boldsymbol{S}_r^{\sigma T} - T)$. Then

$$\Delta = \hat{T} - T = \sum_{i,i'} \Delta_{i,i'} = \frac{1}{|S_n|} \sum_{\sigma \in S_n} \sum_{r=1}^{n/2} \hat{\Delta}_r^\sigma.$$

and consequently,

$$||\hat{T} - T|| \leq \frac{1}{|S_n|} \sum_{\sigma \in S_n} \sum_{r=1}^{n/2} \hat{\Delta}_r^\sigma.$$

Let N_ϵ be the largest number of ϵ-balls one can pack in the $(1 + \epsilon)$-ball centered at the origin and $\{\boldsymbol{w}_{(j)}, j \leq N_\epsilon\}$ be the centers of such ϵ-balls in one of such configurations. From straight forward volume comparison we have

$$N_\epsilon \leq (1/\epsilon + 1)^d.$$

For each $\boldsymbol{w} \in S^{d-1}$, $||\boldsymbol{w} - \boldsymbol{w}_{(j)}||_2 \leq 2\epsilon$ for some $j \leq N_\epsilon$, so that

$$\begin{aligned} |\boldsymbol{w}^\top \Delta \boldsymbol{w}| \leq & |\boldsymbol{w}_{(j)}^\top \Delta \boldsymbol{w}_{(j)}| + |(\boldsymbol{w} - \boldsymbol{w}_{(j)})^\top \Delta (\boldsymbol{w} - \boldsymbol{w}_{(j)})| \\ \leq & |\boldsymbol{w}_{(j)}^\top \Delta \boldsymbol{w}_{(j)}| + 4\epsilon^2 ||\Delta||_2. \end{aligned}$$

This implies

$$||\Delta||_2 \leq \sup_{j \leq N_\epsilon} \frac{|\boldsymbol{w}_{(j)}^\top \Delta \boldsymbol{w}_{(j)}|}{1 - \epsilon^2}, \tag{S1.1}$$

with $N_\epsilon \leq (1 + 1/\epsilon)^d$.

In addition, for any $\boldsymbol{w} \in S^{d-1}$, according to Lemma 1, we have

$$E[e^{t\boldsymbol{w}^\top (\sum_{r=1}^{n/2} \hat{\Delta}_r^\sigma) \boldsymbol{w}}] = \prod_{r=1}^{n/2} E[e^{t\boldsymbol{w}^\top \hat{\Delta}_r^\sigma \boldsymbol{w}}] = \prod_{r=1}^{n/2} E[e^{\frac{t}{n/2} \boldsymbol{w}^\top (S_r^\sigma S_r^{\sigma T} - T) \boldsymbol{w}}] \leq e^{\frac{2t^2 M^2 \pi}{n}}.$$

Then by Jensen's inequality,

$$E[e^{t\boldsymbol{w}^\top \Delta \boldsymbol{w}}] = E[e^{t\boldsymbol{w}^\top \frac{1}{|S_n|} \sum_{\sigma \in S_n} \sum_{r=1}^{n/2} \hat{\Delta}_r^\sigma \boldsymbol{w}}] \leq \frac{1}{|S_n|} \sum_{\sigma \in S_n} E[e^{t\boldsymbol{w}^\top \sum_{r=1}^{n/2} \hat{\Delta}_r^\sigma \boldsymbol{w}}] \leq e^{\frac{2t^2 M^2 \pi}{n}}.$$

Therefore, by the property of sub-gaussian random variable, for any $\boldsymbol{w} \in S^{d-1}$,

$$P(\boldsymbol{w}^\top \Delta \boldsymbol{w} > t) \leq e^{-\frac{nt^2}{2M^2\pi}}.$$

Then by (S1.1) and let $\epsilon = 1/2$, we have

$$P(||\Delta||_2 > t) \leq 3^d e^{-\frac{nt^2}{2M^2\pi}} = e^{d\log 3 - \frac{nt^2}{2M^2\pi}}.$$

Let $t = \sqrt{(2\pi \log 3M^2)\frac{d+t}{n}}$, then with probability at least $1 - e^{-t}$,

$$||\Delta|| \lesssim \sqrt{\frac{d+t}{n}}.$$

At last, according to Theorem 2.2 in Wegkamp and Zhao (2016),

$$||\Sigma - \hat{\Sigma}|| \lesssim \sqrt{\frac{d+t}{n}}.$$

3. For any $A \subset [p]$ with $|A| = s$, we have $\lambda_{\max}(\Sigma_{A,A})/\lambda_{\min}(\Sigma_{A,A}) \leq \kappa_s(\Sigma) \leq M$. Then by the result in (2), with probability at least $1 - e^{-t}$,

$$\frac{||\Delta_{A\times A}||}{||\Sigma_{A\times A}||} \lesssim \sqrt{\frac{s+t}{n}}.$$

Therefore

$$\frac{||\Delta||_{2,s}}{||\Sigma||_{2,s}} = \frac{\sup_{|A|=s} ||\Delta_{A\times A}||}{\sup_{|A|=s} ||\Sigma_{A\times A}||} \leq \sup_{|A|=s} \frac{||\Delta_{A\times A}||}{||\Sigma_{A\times A}||} \lesssim \sqrt{\frac{s+t}{n}}$$

with probability at least $1 - \binom{p}{s} \cdot e^{-t} = 1 - e^{s\log p - t}$.

This implies with probability at least $1 - e^{-t}$,

$$\frac{||\Delta||_{2,s}}{||\Sigma||_{2,s}} \lesssim \sqrt{\frac{s\log p + t}{n}} \Rightarrow ||\Delta||_{2,s} \lesssim ||\Sigma||_{2,s}\sqrt{\frac{s\log p + t}{n}}.$$

□

S1.8 Proof of Lemma 10

We first show that the error vector $\widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta} \in C(s)$. Let $S = \mathrm{supp}(\boldsymbol{\beta})$. By the definition of $\widehat{\boldsymbol{\beta}}(\lambda)$, we have

$$L(\widehat{\boldsymbol{\beta}}(\lambda)) + \lambda \|\widehat{\boldsymbol{\beta}}(\lambda)\|_1 \leq L(\boldsymbol{\beta}) + \lambda \|\boldsymbol{\beta}\|_1 \tag{S1.1}$$

Since

$$\begin{aligned} L(\widehat{\boldsymbol{\beta}}(\lambda)) - L(\boldsymbol{\beta}) =& (\widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta})^\top \hat{\Sigma}_{XX} (\widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta}) + (2\boldsymbol{\beta}^\top \hat{\Sigma}_{XX} - 2\hat{\Sigma}_{YX})(\widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta}) \\ \geq& \nabla L(\boldsymbol{\beta})^\top (\widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta}), \end{aligned}$$

then the optimization inequality (S1.1) implies that

$$\begin{aligned} 2\lambda \|\widehat{\boldsymbol{\beta}}(\lambda)\|_1 \leq& 2\nabla L(\boldsymbol{\beta})^\top (\widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta}) + 2\lambda \|\boldsymbol{\beta}\|_1 \\ \leq& \lambda \|\widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta}\|_1 + 2\lambda \|\boldsymbol{\beta}\|_1. \end{aligned}$$

On the left-hand side, using the triangle inequality,

$$\|\widehat{\boldsymbol{\beta}}(\lambda)\|_1 = \|\widehat{\boldsymbol{\beta}}_S(\lambda)\|_1 + \|\widehat{\boldsymbol{\beta}}_{S^c}(\lambda)\|_1 \geq \|\boldsymbol{\beta}_S\|_1 - \|\widehat{\boldsymbol{\beta}}_S(\lambda) - \boldsymbol{\beta}_S\|_1 + \|\widehat{\boldsymbol{\beta}}_{S^c}(\lambda)\|_1,$$

we obtain

$$\begin{aligned} 2\|\boldsymbol{\beta}_S\|_1 - 2\|\widehat{\boldsymbol{\beta}}_S(\lambda) - \boldsymbol{\beta}_S\|_1 + 2\|\widehat{\boldsymbol{\beta}}_{S^c}(\lambda)\|_1 \leq& \|\widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta}\|_1 + 2\|\boldsymbol{\beta}_S\|_1 \\ =& \|\widehat{\boldsymbol{\beta}}_S(\lambda) - \boldsymbol{\beta}_S\|_1 + \|\widehat{\boldsymbol{\beta}}_{S^c}(\lambda)\|_1 \end{aligned}$$

It follows that

$$\|\widehat{\boldsymbol{\beta}}_{S^c}(\lambda)\|_1 \leq 3\|\widehat{\boldsymbol{\beta}}_S(\lambda) - \boldsymbol{\beta}_S\|_1,$$

which is equivalent to

$$\|\widehat{\boldsymbol{\beta}}_{S^c}(\lambda) - \boldsymbol{\beta}_{S^c}\|_1 \leq 3\|\widehat{\boldsymbol{\beta}}_S(\lambda) - \boldsymbol{\beta}_S\|_1,$$

that is $\widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta} \in C(s)$.

Then we are ready to show the desired result. Let $\Delta = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$. Use (S1.1) again, we then have

$$L(\boldsymbol{\beta} + \Delta) - L(\boldsymbol{\beta}) \leq \lambda \|\boldsymbol{\beta}\|_1 - \lambda \|\widehat{\boldsymbol{\beta}}(\lambda)\|_1 \leq \lambda ||\widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta}||_1.$$

Substitute (6.1) into above inequality, and use the fact that $\Delta \in C(s)$,

$$\kappa_L \|\Delta\|_2^2 \leq \lambda ||\widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta}||_1 - \langle \nabla L(\boldsymbol{\beta}), \Delta \rangle \leq \frac{3}{2} \lambda \|\Delta\|_1 \leq 3\lambda \|\Delta_S\|_1 \leq 3\lambda \sqrt{s} \|\Delta\|_2.$$

Therefore, we have

$$\|\Delta\|_2 \lesssim \sqrt{s}\lambda,$$

and

$$\|\Delta\|_1 \leq 4\sqrt{s} \|\Delta\|_1 \lesssim s\lambda.$$

□

S2 Simulation results

Model 2									
	SNR	$\widehat{\boldsymbol{\beta}}_{\text{Copula}}(\boldsymbol{Y}, X)$		$\widehat{\boldsymbol{\beta}}_{\text{Lasso}}(\widetilde{\boldsymbol{Y}}, \widetilde{X})$		$\widehat{\boldsymbol{\beta}}_{\text{Lasso}}(\boldsymbol{Y}, X)$		$\widehat{\boldsymbol{\beta}}_{\text{enet}}(\boldsymbol{Y}, X)$	
(n, p, s)		$e_{\text{selection}}$	e_{est}	$e_{\text{selection}}$	e_{est}	$e_{\text{selection}}$	e_{est}	$e_{\text{selection}}$	e_{est}
$(100, 500, 10)_1$	0.5473	0.0247	0.7621	0.0251	0.6257	0.0200	0.9197	0.0164	0.9557
$(100, 500, 10)_2$	0.5473	0.0247	0.7621	0.0251	0.6257	0.0178	2.7970	0.0092	1.9027
$(100, 500, 20)_1$	0.5556	0.0250	0.7776	0.0250	0.5596	0.0180	1.0560	0.0159	0.9301
$(100, 500, 20)_2$	0.5556	0.0250	0.7776	0.0250	0.5596	0.0206	0.9921	0.0148	1.1096
$(100, 1000, 10)_1$	0.5278	0.0291	0.7848	0.0261	0.6814	0.0152	0.8734	0.0135	0.8459
$(100, 1000, 10)_2$	0.5278	0.0291	0.7848	0.0261	0.6814	0.0143	0.9306	0.0042	0.8827
$(100, 1000, 20)_1$	0.5394	0.0284	0.7753	0.0247	0.6272	0.0266	0.8754	0.0133	0.9171
$(100, 1000, 20)_2$	0.5394	0.0284	0.7753	0.0247	0.6272	0.0258	5.0359	0.0100	1.9455
$(200, 500, 10)_1$	0.5239	0.0366	0.5622	0.0266	0.4705	0.0278	6.4231	0.0137	1.8282
$(200, 500, 10)_2$	0.5239	0.0366	0.5622	0.0266	0.4705	0.0276	8.9832	0.0120	1.0145
$(200, 500, 20)_1$	0.5891	0.0280	0.5663	0.0275	0.4694	0.0384	9.5063	0.0109	2.3346
$(200, 500, 20)_2$	0.5891	0.0280	0.5663	0.0275	0.4694	0.0368	8.2352	0.0126	0.9153
$(200, 1000, 10)_1$	0.5661	0.0304	0.5962	0.0286	0.5292	0.0337	0.8713	0.0157	0.9709
$(200, 1000, 10)_2$	0.5661	0.0304	0.5962	0.0286	0.5292	0.0303	0.9024	0.0202	0.8055
$(200, 1000, 20)_1$	0.5912	0.0302	0.5947	0.0254	0.5208	0.0341	1.4135	0.0175	1.4315
$(200, 1000, 20)_2$	0.5912	0.0302	0.5947	0.0254	0.5208	0.0340	3.1180	0.0154	1.8348

Table 1: Simulation results for the synthetic data under Model 2 described in Section 4. The results corresponds to model selection error $e_{\text{selection}}$ and estimation error e_{est} for $\widehat{\boldsymbol{\beta}}_{\text{Copula}}(\boldsymbol{Y}, X)$, $\widehat{\boldsymbol{\beta}}_{\text{Lasso}}(\widetilde{\boldsymbol{Y}}, \widetilde{X})$, $\widehat{\boldsymbol{\beta}}_{\text{Lasso}}(\boldsymbol{Y}, X)$ and $\widehat{\boldsymbol{\beta}}_{\text{enet}}(\boldsymbol{Y}, X)$. The subscript i $(i = 1, 2)$ in $(n, p, s)_i$ denotes the i-th setting of transformations

Model 3									
	SNR	$\widehat{\boldsymbol{\beta}}_{\text{Copula}}(\boldsymbol{Y}, X)$		$\widehat{\boldsymbol{\beta}}_{\text{Lasso}}(\widetilde{\boldsymbol{Y}}, \widetilde{X})$		$\widehat{\boldsymbol{\beta}}_{\text{Lasso}}(\boldsymbol{Y}, X)$		$\widehat{\boldsymbol{\beta}}_{\text{enet}}(\boldsymbol{Y}, X)$	
(n, p, s)		$e_{\text{selection}}$	e_{est}	$e_{\text{selection}}$	e_{est}	$e_{\text{selection}}$	e_{est}	$e_{\text{selection}}$	e_{est}
$(100, 500, 10)_1$	177.2462	0.0227	0.2267	0.0318	0.2005	0.0127	1.1280	0.0161	0.7143
$(100, 500, 10)_2$	177.2462	0.0227	0.2267	0.0318	0.2005	0.0130	4.2015	0.0128	0.8058
$(100, 500, 20)_1$	65.9603	0.0196	0.1929	0.0333	0.1788	0.0156	8.3003	0.0221	2.4008
$(100, 500, 20)_2$	65.9603	0.0196	0.1929	0.0333	0.1788	0.0190	0.9023	0.0184	1.2964
$(100, 1000, 10)_1$	387.7395	0.0111	0.2027	0.0294	0.2487	0.0274	0.8152	0.0162	0.7159
$(100, 1000, 10)_2$	387.7395	0.0111	0.2027	0.0294	0.2487	0.0242	0.5549	0.0102	0.9456
$(100, 1000, 20)_1$	56.3955	0.0102	0.1861	0.0287	0.1926	0.0279	6.4805	0.0189	1.3961
$(100, 1000, 20)_2$	56.3955	0.0102	0.1861	0.0287	0.1926	0.0274	4.1451	0.0101	0.9435
$(200, 500, 10)_1$	27.9771	0.0167	0.1365	0.0302	0.1119	0.0243	0.9522	0.0122	0.7159
$(200, 500, 10)_2$	27.9771	0.0167	0.1365	0.0302	0.1119	0.0263	3.2474	0.0112	0.9879
$(200, 500, 20)_1$	6.9023	0.0150	0.1412	0.0329	0.1133	0.0402	6.9362	0.0216	0.9043
$(200, 500, 20)_2$	6.9023	0.0150	0.1412	0.0329	0.1133	0.0592	8.5190	0.0215	0.8182
$(200, 1000, 10)_1$	579.5246	0.0107	0.1337	0.0267	0.1569	0.0376	1.4837	0.0161	0.8241
$(200, 1000, 10)_2$	579.5246	0.0107	0.1337	0.0267	0.1569	0.0422	7.4434	0.0116	1.1035
$(200, 1000, 20)_1$	25.4231	0.0101	0.1490	0.0294	0.1481	0.0345	4.4740	0.0152	1.0644
$(200, 1000, 20)_2$	25.4231	0.0101	0.1490	0.0294	0.1481	0.0382	0.4315	0.0107	1.2419

Table 2: Simulation results for the synthetic data under Model 3 described in Section 4. The results corresponds to model selection error $e_{\text{selection}}$ and estimation error e_{est} for $\widehat{\boldsymbol{\beta}}_{\text{Copula}}(\boldsymbol{Y}, X)$, $\widehat{\boldsymbol{\beta}}_{\text{Lasso}}(\widetilde{\boldsymbol{Y}}, \widetilde{X})$, $\widehat{\boldsymbol{\beta}}_{\text{Lasso}}(\boldsymbol{Y}, X)$ and $\widehat{\boldsymbol{\beta}}_{\text{enet}}(\boldsymbol{Y}, X)$. The subscript i $(i = 1, 2)$ in $(n, p, s)_i$ denotes the i-th setting of transformations

Model 2						
	$\text{CI}(\widehat{\boldsymbol{\beta}}^u_{\text{Copula}}(\boldsymbol{Y}, X))$		$\text{CI}(\widehat{\boldsymbol{\beta}}^u_{\text{Lasso}}(\widetilde{\boldsymbol{Y}}, \widetilde{X}))$		$\text{CI}(\widehat{\boldsymbol{\beta}}^u_{\text{Lasso}}(\boldsymbol{Y}, X))$	
(n, p, s)	$l(\beta_1)$	$C(\beta_1)$	$l(\beta_1)$	$C(\beta_1)$	$l(\beta_1)$	$C(\beta_1)$
$(100, 500, 10)_1$	0.0482	0.938	0.0348	0.952	1.1152	0.424
$(100, 500, 10)_2$	0.0482	0.938	0.0348	0.952	0.1331	0.628
$(100, 500, 20)_1$	0.0487	0.942	0.0325	0.958	0.7472	0.294
$(100, 500, 20)_2$	0.0487	0.942	0.0325	0.958	0.1301	0.470
$(100, 1000, 10)_1$	0.0402	0.954	0.0302	0.946	0.8500	0.282
$(100, 1000, 10)_2$	0.0402	0.954	0.0302	0.946	0.1071	0.096
$(100, 1000, 20)_1$	0.0411	0.952	0.0295	0.954	1.3763	0.798
$(100, 1000, 20)_2$	0.0411	0.952	0.0295	0.954	0.1840	0.024
$(200, 500, 10)_1$	0.0535	0.944	0.0254	0.956	1.0074	0.132
$(200, 500, 10)_2$	0.0535	0.944	0.0254	0.956	0.1586	0.054
$(200, 500, 20)_1$	0.0540	0.940	0.0262	0.950	0.9223	0.336
$(200, 500, 20)_2$	0.0540	0.940	0.0262	0.950	0.1505	0.298
$(200, 1000, 10)_1$	0.0438	0.952	0.0255	0.962	0.6714	0.482
$(200, 1000, 10)_2$	0.0438	0.952	0.0255	0.962	0.0897	0.230
$(200, 1000, 20)_1$	0.0429	0.948	0.0254	0.954	0.7603	0.302
$(200, 1000, 20)_2$	0.0429	0.948	0.0254	0.954	0.1245	0.194

Table 3: Simulation results for the synthetic data under Model 2 described in Section 4. The results corresponds to 95% confidence intervals. $C(\beta_1)$ and $l(\beta_1)$ respectively stand for coverage probability and average lengths of the confidence interval for β_1. The subscript i $(i = 1, 2)$ in $(n, p, s)_i$ denotes the i-th setting of transformations.

Model 3						
	$\mathrm{CI}(\widehat{\boldsymbol{\beta}}^u_{\mathrm{Copula}}(\boldsymbol{Y},X))$		$\mathrm{CI}(\widehat{\boldsymbol{\beta}}^u_{\mathrm{Lasso}}(\widetilde{\boldsymbol{Y}},\widetilde{X}))$		$\mathrm{CI}(\widehat{\boldsymbol{\beta}}^u_{\mathrm{Lasso}}(\boldsymbol{Y},X))$	
(n,p,s)	$l(\beta_1)$	$C(\beta_1)$	$l(\beta_1)$	$C(\beta_1)$	$l(\beta_1)$	$C(\beta_1)$
$(100,500,10)_1$	0.0223	0.952	0.0455	0.958	0.8968	0.234
$(100,500,10)_2$	0.0223	0.952	0.0455	0.958	0.1227	0.142
$(100,500,20)_1$	0.0241	0.948	0.0389	0.948	0.9163	0.034
$(100,500,20)_2$	0.0241	0.948	0.0389	0.948	0.1081	0.392
$(100,1000,10)_1$	0.0203	0.958	0.0462	0.956	1.3611	0.524
$(100,1000,10)_2$	0.0203	0.958	0.0462	0.956	0.2073	0.082
$(100,1000,20)_1$	0.0224	0.962	0.0399	0.952	1.3931	0.830
$(100,1000,20)_2$	0.0224	0.962	0.0399	0.952	0.1734	0.048
$(200,500,10)_1$	0.0138	0.946	0.0264	0.944	1.2536	0.342
$(200,500,10)_2$	0.0138	0.946	0.0264	0.944	0.1623	0.062
$(200,500,20)_1$	0.0154	0.952	0.0281	0.948	0.5148	0.722
$(200,500,20)_2$	0.0154	0.952	0.0281	0.948	0.1059	0.234
$(200,1000,10)_1$	0.0121	0.958	0.0374	0.950	0.9380	0.148
$(200,1000,10)_2$	0.0121	0.958	0.0374	0.950	0.1072	0.326
$(200,1000,20)_1$	0.0140	0.962	0.0256	0.952	0.4721	0.694
$(200,1000,20)_2$	0.0140	0.962	0.0256	0.952	0.0980	0.526

Table 4: Simulation results for the synthetic data under Model 3 described in Section 4. The results corresponds to 95% confidence intervals. $C(\beta_1)$ and $l(\beta_1)$ respectively stand for coverage probability and average lengths of the confidence interval for β_1. The subscript i $(i=1,2)$ in $(n,p,s)_i$ denotes the i-th setting of transformations.

References

Callaert, H., and Janssen, P. (1978). The Berry-Esseen theorem for U-statistics. *The Annals of Statistics*, 417-421.

Horn, R. A., and Johnson, C. R. (1995). *Matrix analysis*. Cambridge University press.

Borell, C. (1975). The Brunn-Minkowski inequality in Gauss space. *Inventiones Mathematicae* **30**, 207-216.

Wegkamp, M., and Zhao, Y. (2016). Adaptive estimation of the copula correlation matrix for semiparametric elliptical copulas. *Bernoulli* **22**, 1184-1226.