Supplementary material to

Inferring Hub Nodes on Differential Gaussian Graphical Models

Xin Zhou* Kean Ming Tan† Junwei Lu‡

Abstract

This document contains the supplementary material to the paper "Inferring Hub Nodes on Differential Gaussian Graphical Models". In Appendix A, we provide the proof of Theorem 4.1 and Lemma 4.2. In Appendix B, we prove Theorem 4.4. Appendix C states some technical lemmas of Appendix A. Appendix D presents additional numeric results for Section 5.1.

Throughout the Appendix, we will use the following Landau symbols. We write $f(n,d) = O(g(n,d))$ if $f(n,d) \le Cg(n,d)$ for some positive constant C and all sufficiently large n and d. The constant C is independent of the sample size n and the dimension d, but may vary from line to line. For simplicity, we also write $f(n,d) \lesssim g(n,d)$ to indicate $f(n,d) = O(g(n,d))$. And we write $f(n,d) = o(g(n,d))$ if $f(n,d)/g(n,d) \to 0$ as $n \to \infty$ and $d \to \infty$.

A Proof of Theorem 4.1 and Lemma 4.2

In this section, we provide the proofs for Theorem 4.1 and Lemma 4.2. To this end, we start with some technical lemmas that will be helpful in the proof. The proofs of the technical lemmas are deferred to Appendix C. The following lemma shows that the term $\widehat{\mathbf{\Delta}}^d - \mathbf{\Delta}$ can be rewritten as the summation of a leading term and a remainder term.

Lemma A.1. We have $\widehat{\mathbf{\Delta}}^d - \mathbf{\Delta} = \text{Leading} + \text{Remainder}$, where

$$\text{Leading} = (\mathbf{V}_Y - \mathbf{\Delta})\big(\widehat{\mathbf{\Sigma}}_X - \mathbf{\Sigma}_X\big)\mathbf{V}_X^T - \mathbf{V}_Y\big(\widehat{\mathbf{\Sigma}}_Y - \mathbf{\Sigma}_Y\big)(\mathbf{V}_X + \mathbf{\Delta})^T;$$

$$\begin{aligned}\text{Remainder} =& \{(\widehat{\mathbf{\Delta}} - \mathbf{\Delta}) - \mathbf{V}_Y\widehat{\mathbf{\Sigma}}_Y(\widehat{\mathbf{\Delta}} - \mathbf{\Delta})(\mathbf{V}_X\widehat{\mathbf{\Sigma}}_X)^T\} - \mathbf{V}_Y(\widehat{\mathbf{\Sigma}}_Y - \mathbf{\Sigma}_Y)\mathbf{\Delta}(\widehat{\mathbf{\Sigma}}_X - \mathbf{\Sigma}_X)\mathbf{V}_X^T \\ &- \mathbf{V}_Y(\widehat{\mathbf{\Sigma}}_Y - \mathbf{\Sigma}_Y)\mathbf{\Delta}(\mathbf{V}_X\mathbf{\Sigma}_X - \mathbf{I})^T - (\mathbf{V}_Y\mathbf{\Sigma}_Y - \mathbf{I})\mathbf{\Delta}(\widehat{\mathbf{\Sigma}}_X - \mathbf{\Sigma}_X)\mathbf{V}_X^T.\end{aligned}$$

*Division of Biostatistics, University of California, Berkeley, Berkeley, CA 94720; email: xinzhou@berkeley.edu
† Department of Statistics, University of Michigan, Ann Arbor, MI 48109; Email: keanming@umich.edu, ‡Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115; Email: junweilu@hsph.harvard.edu

The next lemma summarizes the estimation rates of the estimators. The results follow directly from Theorem 3 in Zhao et al. (2014), and from Lemma E.5 in Ma et al. (2017). Thus the proof is omitted.

Lemma A.2. Assume the conditions in Theorem 4.1. We have with probability at least $1 - c/d$,

$$\|\widehat{\mathbf{\Delta}} - \mathbf{\Delta}\|_{1,1} = O(s\sqrt{\log d/n}),\ \|\widehat{\mathbf{\Sigma}}_X - \mathbf{\Sigma}_X\|_{\max} = O(\sqrt{\log d/n}),\ \text{and}\ \|\widehat{\mathbf{\Sigma}}_Y - \mathbf{\Sigma}_Y\|_{\max} = O(\sqrt{\log d/n}),$$

where c is a generic constant. Furthermore, we define the event Γ as follows:

$$\Gamma = \{(\mathbb{X}^T, \mathbb{Y}^T) \in \mathbb{R}^{2n\times d} | \|\widehat{\mathbf{\Sigma}}'_X - \mathbf{\Sigma}_X\|_{\max} \le C\sqrt{\log d/n},\ \|\widehat{\mathbf{\Sigma}}'_Y - \mathbf{\Sigma}_Y\|_{\max} \le C\sqrt{\log d/n}\}, \quad \text{(A.1)}$$

where $\widehat{\mathbf{\Sigma}}'_X, \widehat{\mathbf{\Sigma}}'_Y$ are sample covariance matrices obtained from the splitted data $\mathcal{D}_2$ as suggested in Section 3.1, and C is a sufficiently large constant. Then we have $\mathbb{P}(\mathcal{D}_2 \in \Gamma) \ge 1 - 4d^{-1}$.

The next lemma establishes properties of the bias correction matrices $\mathbf{V}_X$ and $\mathbf{V}_Y$ that will be useful for bounding the remainder term.

Lemma A.3. Assume the conditions in Theorem 4.1. Let λ' in (34) satisfies $\lambda' = C'\sqrt{\log d/n}$, where C' is a sufficiently large constant. We have with probability at least $1 - c/d$,

$$\|\mathbf{V}_X\|_\infty = O(1); \qquad \|\mathbf{V}_Y\|_\infty = O(1);$$

$$\|\mathbf{V}_X\widehat{\mathbf{\Sigma}}_X - \mathbf{I}\|_{\max} = O(\sqrt{\log d/n});\ \text{and}\ \ \|\mathbf{V}_Y\widehat{\mathbf{\Sigma}}_Y - \mathbf{I}\|_{\max} = O(\sqrt{\log d/n}),$$

where c is a generic constant.

Next, we present a lemma on an upper bound for the remainder term.

Lemma A.4. Assume the conditions in Theorem 4.1. We have with probability at least $1 - c/d$ that $\|\text{Remainder}\|_{\max} = O(s\log d/n)$, where c is a generic constant.

The next lemma provides a lower bound for the asymptotic variance ξ_{jk} defined in (43).

Lemma A.5. Under conditions in Theorem 4.1 , we have $1/\xi_{jk} = O_P(1)$.

The sixth lemma bounds of the third moment of a random variable, which can lead to the asymptotic normality of the Leading term.

Lemma A.6. (Third moment bound) Suppose the conditions of Theorem 4.1 hold, and denote

$$Z_{jk} = (\mathbf{V}_Y - \boldsymbol{\Delta})_{j\bullet}\big(\boldsymbol{X}\boldsymbol{X}^T - \boldsymbol{\Sigma}_X\big)\mathbf{V}_{X,k\bullet}^T - \mathbf{V}_{Y,j\bullet}\big(\boldsymbol{Y}\boldsymbol{Y}^T - \boldsymbol{\Sigma}_Y\big)(\mathbf{V}_X + \boldsymbol{\Delta})_{k\bullet}^T, \quad \forall j,k \in [d].$$

We have $\forall (\mathbb{X}^T, \mathbb{Y}^T)^T \in \Gamma$, $\rho := \mathbb{E}(|Z_{jk}|^3 \big| \mathcal{D}_2 = (\mathbb{X}^T, \mathbb{Y}^T)^T) = O(1)$.

With Lemmas A.1–A.6, we now provide the proof of Theorem 4.1.

A.1 Proof of Theorem 4.1

Proof. From Lemma A.1, we have

$$\sqrt{n}(\widehat{\boldsymbol{\Delta}}_{jk}^d - \boldsymbol{\Delta}_{jk})/\xi_{jk} = \sqrt{n}\text{Leading}_{jk}/\xi_{jk} + \sqrt{n}\text{Remainder}_{jk}/\xi_{jk}.$$

From Lemmas A.4–A.5 and the scaling condition $s \log d/\sqrt{n} = o(1)$, we have $\sqrt{n}\text{Remainder}_{jk}/\xi_{jk} = o_P(1)$. By Slutsky's theorem, it remains to show that $\sqrt{n}\text{Leading}_{jk}/\xi_{jk} \rightsquigarrow N(0,1)$. To this end, let

$$Z_{i,jk} = (\mathbf{V}_Y - \boldsymbol{\Delta})_{j\bullet}\big(\boldsymbol{X}_i\boldsymbol{X}_i^T - \boldsymbol{\Sigma}_X\big)\mathbf{V}_{X,k\bullet}^T - \mathbf{V}_{Y,j\bullet}\big(\boldsymbol{Y}_i\boldsymbol{Y}_i^T - \boldsymbol{\Sigma}_Y\big)(\mathbf{V}_X + \boldsymbol{\Delta})_{k\bullet}^T,$$

$S_n = \sum_{i=1}^n Z_{i,jk} = n\cdot\text{Leading}_{jk}$, and $s_n = \sqrt{n}\xi_{jk}$. Conditioning on $\mathcal{D}_2$, we know that $Z_{1,jk}, \ldots, Z_{n,jk}$ are i.i.d., so we can apply the Berry-Esseen Theorem (see Theorem 3.4.17 in Durrett (2019)) by conditioning on $\mathcal{D}_2$ to obtain an upper bound: there exists a universal constant C such that for every $(\mathbb{X}^T, \mathbb{Y}^T)^T \in \Gamma$ in (A.1) and every $x \in \mathbb{R}$,

$$\left|\mathbb{P}\left(\frac{S_n}{s_n} \le x \middle| \mathcal{D}_2 = (\mathbb{X}^T, \mathbb{Y}^T)^T\right) - \Phi(x)\right| \le \frac{C\rho}{\sqrt{n}\xi_{jk}^3} \le \frac{C}{\sqrt{n}}, \tag{A.2}$$

where ρ is the conditional third moment in Lemma A.6 and C_2 is another universal constant. Note that the last inequality holds by an application of Lemmas A.5 and A.6.

Define the event $\mathcal{E} = \{\mathcal{D}_2 \in \Gamma\}$. Using the law of total probability, we have

$$\begin{aligned}\left|\mathbb{P}\left(\frac{S_n}{s_n} \le x\right) - \Phi(x)\right| &= \left|\mathbb{P}\left(\frac{S_n}{s_n} \le x \middle| \mathcal{E}\right)\mathbb{P}(\mathcal{E}) + \mathbb{P}\left(\frac{S_n}{s_n} \le x \middle| \mathcal{E}^c\right)\mathbb{P}(\mathcal{E}^c) - \Phi(x)\right| \\ &\le \left|\mathbb{P}\left(\frac{S_n}{s_n} \le x \middle| \mathcal{E}\right) - \Phi(x)\right|\mathbb{P}(\mathcal{E}) + \left|\mathbb{P}\left(\frac{S_n}{s_n} \le x \middle| \mathcal{E}^c\right) - \Phi(x)\right|\mathbb{P}(\mathcal{E}^c) \qquad \text{(A.3)} \\ &\le \left|\mathbb{P}\left(\frac{S_n}{s_n} \le x \middle| \mathcal{E}\right) - \Phi(x)\right| + 2\mathbb{P}(\mathcal{E}^c).\end{aligned}$$

From (A.2), we have

$$\left|\mathbb{P}\left(\frac{S_n}{s_n} \le x \middle| \mathcal{E}\right) - \Phi(x)\right| = \left|\int \left[\mathbb{P}\left(\frac{S_n}{s_n} \le x \middle| \mathcal{D}_2 = (\mathbb{X}^T, \mathbb{Y}^T)^T\right) - \Phi(x)\right] d\mathbb{P}(\mathbb{X}^T, \mathbb{Y}^T)^T\right| \le \frac{C}{\sqrt{n}}.$$

From Lemma A.2, we have $\mathbb{P}(\mathcal{E}^c) \le 4d^{-1}$. From (A.3), we have

$$\left|\mathbb{P}\left(\frac{\sqrt{n}\text{Leading}_{jk}}{\xi_{jk}} \le x\right) - \Phi(x)\right| = \left|\mathbb{P}\left(\frac{S_n}{s_n} \le x\right) - \Phi(x)\right| \le \frac{C}{\sqrt{n}} + \frac{8}{d} = o(1),$$

indicating that $\sqrt{n}\text{Leading}_{jk}/\xi_{jk} \rightsquigarrow N(0,1)$, as desired. □

A.2 Proof of Lemma 4.2

Proof. It suffices to prove $|\widehat{\xi}_{jk}^2 - \xi_{jk}^2| = O_P(1)$. Recall from the definition of ξ_{jk} in (43), we have

$$\begin{aligned}\xi_{jk}^2 = &\underbrace{((\mathbf{V}_Y - \boldsymbol{\Delta})_{j\bullet}\boldsymbol{\Sigma}_X(\mathbf{V}_Y - \boldsymbol{\Delta})_{j\bullet}^T)(\mathbf{V}_{X,k\bullet}\boldsymbol{\Sigma}_X\mathbf{V}_{X,k\bullet}^T)}_{I_1} + \underbrace{((\mathbf{V}_Y - \boldsymbol{\Delta})_{j\bullet}\boldsymbol{\Sigma}_X\mathbf{V}_{X,k\bullet}^T)^2}_{I_2} \\ &+ \underbrace{(\mathbf{V}_{Y,j\bullet}\boldsymbol{\Sigma}_Y\mathbf{V}_{Y,j\bullet}^T)((\mathbf{V}_X + \boldsymbol{\Delta})_{k\bullet}\boldsymbol{\Sigma}_Y(\mathbf{V}_X + \boldsymbol{\Delta})_{k\bullet}^T)}_{I_3} + \underbrace{(\mathbf{V}_{Y,j\bullet}\boldsymbol{\Sigma}_Y(\mathbf{V}_X + \boldsymbol{\Delta})_{k\bullet}^T)^2}_{I_4}\end{aligned}$$

Similarly we use the definition of $\widehat{\xi}_{jk}$ in (45) to expand $\widehat{\xi}_{jk}^2 = \widehat{I}_1 + \widehat{I}_2 + \widehat{I}_3 + \widehat{I}_4$. Then $|\widehat{\xi}_{jk}^2 - \xi_{jk}^2| \le \sum_{t=1}^4 |I_t - \widehat{I}_t|$. We start with showing $|I_1 - \widehat{I}_1| = O_P(1)$. Note that

$$\begin{aligned}|I_1 - \widehat{I}_1| &= |((\mathbf{V}_Y - \boldsymbol{\Delta})_{j\bullet}\boldsymbol{\Sigma}_X(\mathbf{V}_Y - \boldsymbol{\Delta})_{j\bullet}^T)(\mathbf{V}_{X,k\bullet}\boldsymbol{\Sigma}_X\mathbf{V}_{X,k\bullet}^T) - ((\mathbf{V}_Y - \widehat{\boldsymbol{\Delta}})_{j\bullet}\widehat{\boldsymbol{\Sigma}}_X(\mathbf{V}_Y - \widehat{\boldsymbol{\Delta}})_{j\bullet}^T)(\mathbf{V}_{X,k\bullet}\widehat{\boldsymbol{\Sigma}}_X\mathbf{V}_{X,k\bullet}^T)| \\ &\le |((\mathbf{V}_Y - \boldsymbol{\Delta})_{j\bullet}\boldsymbol{\Sigma}_X(\mathbf{V}_Y - \boldsymbol{\Delta})_{j\bullet}^T - (\mathbf{V}_Y - \widehat{\boldsymbol{\Delta}})_{j\bullet}\widehat{\boldsymbol{\Sigma}}_X(\mathbf{V}_Y - \widehat{\boldsymbol{\Delta}})_{j\bullet}^T)(\mathbf{V}_{X,k\bullet}\boldsymbol{\Sigma}_X\mathbf{V}_{X,k\bullet}^T)| \\ &\quad + |(\mathbf{V}_Y - \widehat{\boldsymbol{\Delta}})_{j\bullet}\widehat{\boldsymbol{\Sigma}}_X(\mathbf{V}_Y - \widehat{\boldsymbol{\Delta}})_{j\bullet}^T)\{\mathbf{V}_{X,k\bullet}(\widehat{\boldsymbol{\Sigma}}_X - \boldsymbol{\Sigma}_X)\mathbf{V}_{X,k\bullet}^T\}|.\end{aligned}$$

From Assumption 1, we have $\|(\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}\|_1 = O_P(1)$, $\|(\mathbf{V}_Y - \widehat{\mathbf{\Delta}})_{j\bullet} - (\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}\|_1 = O_P(1)$, $\|\mathbf{V}^X_{k\bullet}\|_1 = O_P(1)$, $\|\mathbf{\Sigma}_X\|_{\max} = O(1)$, and $\|\widehat{\mathbf{\Sigma}}_X - \mathbf{\Sigma}_X\|_{\max} = O_P(\sqrt{\log d/n})$. Combining the above with Lemma A.2, we further have

$$|((\mathbf{V}_Y-\widehat{\mathbf{\Delta}})_{j\bullet}\widehat{\mathbf{\Sigma}}_X(\mathbf{V}_Y-\widehat{\mathbf{\Delta}})^T_{j\bullet})-((\mathbf{V}_Y-\mathbf{\Delta})_{j\bullet}\mathbf{\Sigma}_X(\mathbf{V}_Y-\mathbf{\Delta})^T_{j\bullet})| = O_P(1); |(\mathbf{V}_{X,k\bullet}\mathbf{\Sigma}_X\mathbf{V}^T_{X,k\bullet})| = O_P(1);$$

$$|(\mathbf{V}_{X,k\bullet}\widehat{\mathbf{\Sigma}}_X\mathbf{V}^T_{X,k\bullet}) - (\mathbf{V}_{X,k\bullet}\mathbf{\Sigma}_X\mathbf{V}^T_{X,k\bullet})| = O_P(1); |((\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}\mathbf{\Sigma}_X(\mathbf{V}_Y - \mathbf{\Delta})^T_{j\bullet})| = O_P(1).$$

Through the use of the inequality $|\mathbf{u}^T\mathbf{A}\mathbf{v}| \le \|\mathbf{u}\|_1\|\mathbf{A}\|_{\max}\|\mathbf{v}\|_1$ that holds for any vector $\mathbf{u}$, $\mathbf{v}$ and any matrix $\mathbf{A}$, we conclude that $|I_1 - \widehat{I}_1| = O_P(1)$, as desired.

The other quantities can be derived similarly and are hence omitted from the proof. □

B Proof of Theorem 4.4

In this proof, we will use the notation T^E to indicate the test statistic T_E and T^B to indicate T^B_E. To approximate the statistic T^E by the multiplier bootstrap process T^B, we also define two intermediate processes

$$T^E_0 := \max_{(j,k)\in E} \frac{1}{\sqrt{n}}\sum_{i=1}^n [(\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}(\boldsymbol{X}_i\boldsymbol{X}_i^T - \mathbf{\Sigma}_X)\mathbf{V}^T_{X,k\bullet} - \mathbf{V}_{Y,j\bullet}(\boldsymbol{Y}_i\boldsymbol{Y}_i^T - \mathbf{\Sigma}_Y)(\mathbf{V}_X + \mathbf{\Delta})^T_{k\bullet}],$$

$$T^B_0 := \max_{(j,k)\in E} \frac{1}{\sqrt{n}}\sum_{i=1}^n [(\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}(\boldsymbol{X}_i\boldsymbol{X}_i^T - \mathbf{\Sigma}_X)\mathbf{V}^T_{X,k\bullet} - \mathbf{V}_{Y,j\bullet}(\boldsymbol{Y}_i\boldsymbol{Y}_i^T - \mathbf{\Sigma}_Y)(\mathbf{V}_X + \mathbf{\Delta})^T_{k\bullet}]\xi_i.$$

Since T^B_0 is the multiplier bootstrap process of T^E_0, we can use T^B_0 to estimate the quantiles of T^E_0. We will prove that we can utilize T^E_0 and T^B_0 to approximate T^E and T^B, respectively. To prove this theorem, it suffices to check three conditions in Corollary 3.1 of Chernozhukov et al. (2013), stated as follows:

1. For some positive constants c and C, we have

$$\min_{(j,k)\in E} \mathbb{E}[(\mathbf{V}_Y-\mathbf{\Delta})_{j\bullet}(\boldsymbol{X}\boldsymbol{X}^T-\mathbf{\Sigma}_X)\mathbf{V}^T_{X,k\bullet}]^2 \ge c,\ \min_{(j,k)\in E} \mathbb{E}[\mathbf{V}_{Y,j\bullet}(\boldsymbol{Y}\boldsymbol{Y}^T-\mathbf{\Sigma}_Y)(\mathbf{V}_X+\mathbf{\Delta})^T_{k\bullet}]^2 \ge c;$$

$$\max_{(j,k)\in E} \mathbb{E}[(\mathbf{V}_Y-\mathbf{\Delta})_{j\bullet}(\boldsymbol{X}\boldsymbol{X}^T-\mathbf{\Sigma}_X)\mathbf{V}^T_{X,k\bullet}]^2 \le C,\ \max_{(j,k)\in E} \mathbb{E}[\mathbf{V}_{Y,j\bullet}(\boldsymbol{Y}\boldsymbol{Y}^T-\mathbf{\Sigma}_Y)(\mathbf{V}_X+\mathbf{\Delta})^T_{k\bullet}]^2 \le C;$$

2. $\mathbb{P}(|T^E - T_0^E| > \zeta_1) < \zeta_2$;

3. And $\mathbb{P}\big(\mathbb{P}\big(|T^B - T_0^B| > \zeta_1 \big| \mathcal{D}_1\big) > \zeta_2\big) < \zeta_2$ hold for $\zeta_1\sqrt{\log d} + \zeta_2 = o(1)$.

Note that since we do not emphasize on the polynomial decaying in our result, we need only $\zeta_1\sqrt{\log d} + \zeta_2 = o(1)$ and the scaling condition $(\log(dn))^7/n = o(1)$. We choose $\zeta_1 = C\sqrt{\frac{s^2(\log^3(dn))}{n}}$ and $\zeta_2 = c/d$ to check these conditions, where c and C are two constants. The scaling conditions in Theorem 4.4 imply the aforementioned two scaling conditions.

We start with checking the first condition. The first two inequalities are proven in the proof of Lemma A.5. Moreover, in the proof of Lemma A.6, we have $\forall j, k \in [d]$,

$$\|(\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}\big(\boldsymbol{X}\boldsymbol{X}^T - \mathbf{\Sigma}_X\big)\mathbf{V}_{X,k\bullet}^T\|_{\psi_1} = O(1) \text{ and } \|\mathbf{V}_{Y,j\bullet}\big(\boldsymbol{Y}\boldsymbol{Y} - \mathbf{\Sigma}_Y\big)(\mathbf{V}_X + \mathbf{\Delta})_{k\bullet}^T\|_{\psi_1} = O(1),$$

which imply the two inequalities in the first condition.

Second, we bound the difference between T^E and T_0^E. Note that T_0^E can be rewritten as $T_0^E = \max_{(j,k)\in E}\sqrt{n}\text{Leading}_{jk}$. By Lemma A.1, we have

$$|T^E - T_0^E| \le \sqrt{n}\max_{(j,k)\in E}|\text{Remainder}_{jk}| \le \sqrt{n}\|\text{Remainder}\|_{\max}.$$

From Lemma A.4, we have $|T^E - T_0^E| \le Cs\log d/\sqrt{n}$ with probability at least $1 - c/d$ for some positive constants C and c, implying the second condition.

Finally, to verify the third condition, we bound the difference between T^B and T_0^B:

$$\begin{aligned}
&|T^B - T_0^B| = \max_{(j,k)\in E}|f(\mathcal{D}_1, \xi)|, \text{ where} \\
&f(\mathcal{D}_1, \xi) = \frac{1}{\sqrt{n}}\sum_{i=1}^n[((\mathbf{V}_Y - \widehat{\mathbf{\Delta}})_{j\bullet}\big(\boldsymbol{X}_i\boldsymbol{X}_i^T - \widehat{\mathbf{\Sigma}}_X\big)\mathbf{V}_{X,k\bullet}^T - (\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}\big(\boldsymbol{X}_i\boldsymbol{X}_i^T - \mathbf{\Sigma}_X\big)\mathbf{V}_{X,k\bullet}^T) \\
&\qquad - (\mathbf{V}_{Y,j\bullet}\big(\boldsymbol{Y}_i\boldsymbol{Y}_i^T - \widehat{\mathbf{\Sigma}}_Y\big)(\mathbf{V}_X + \widehat{\mathbf{\Delta}})_{k\bullet}^T - \mathbf{V}_{Y,j\bullet}\big(\boldsymbol{Y}_i\boldsymbol{Y}_i^T - \mathbf{\Sigma}_Y\big)(\mathbf{V}_X + \mathbf{\Delta})_{k\bullet}^T)]\xi_i
\end{aligned} \tag{B.1}$$

Notice that the right hand side of the above inequality is a suprema of the Gaussian process when conditioning on the data $\mathcal{D}_1$, so we need to bound the conditional variance of $f(\mathcal{D}_1, \xi)$ as σ_ξ^2 in the following inequality.

$$
\begin{aligned}
&\frac{1}{n}\max_{(j,k)\in E}|\sum_{i=1}^{n}([(\mathbf{V}_Y-\widehat{\mathbf{\Delta}})_{j\bullet}(\boldsymbol{X}_i\boldsymbol{X}_i^T-\widehat{\mathbf{\Sigma}}_X)\mathbf{V}_{X,k\bullet}^T-(\mathbf{V}_Y-\mathbf{\Delta})_{j\bullet}(\boldsymbol{X}_i\boldsymbol{X}_i^T-\mathbf{\Sigma}_X)\mathbf{V}_{X,k\bullet}^T]\\
&\qquad\qquad-(\mathbf{V}_{Y,j\bullet}(\boldsymbol{Y}_i\boldsymbol{Y}_i^T-\widehat{\mathbf{\Sigma}}_Y)(\mathbf{V}_X+\widehat{\mathbf{\Delta}})_{k\bullet}^T-\mathbf{V}_{Y,j\bullet}(\boldsymbol{Y}_i\boldsymbol{Y}_i^T-\mathbf{\Sigma}_Y)(\mathbf{V}_X+\mathbf{\Delta})_{k\bullet}^T)]|^2\\
&\le\max_{i\in[n]}\max_{(j,k)\in E}|([(\mathbf{V}_Y-\widehat{\mathbf{\Delta}})_{j\bullet}(\boldsymbol{X}_i\boldsymbol{X}_i^T-\widehat{\mathbf{\Sigma}}_X)\mathbf{V}_{X,k\bullet}^T-(\mathbf{V}_Y-\mathbf{\Delta})_{j\bullet}(\boldsymbol{X}_i\boldsymbol{X}_i^T-\mathbf{\Sigma}_X)\mathbf{V}_{X,k\bullet}^T]\\
&\qquad\qquad-(\mathbf{V}_{Y,j\bullet}(\boldsymbol{Y}_i\boldsymbol{Y}_i^T-\widehat{\mathbf{\Sigma}}_Y)(\mathbf{V}_X+\widehat{\mathbf{\Delta}})_{k\bullet}^T-\mathbf{V}_{Y,j\bullet}(\boldsymbol{Y}_i\boldsymbol{Y}_i^T-\mathbf{\Sigma}_Y)(\mathbf{V}_X+\mathbf{\Delta})_{k\bullet}^T)]|^2\\
&\le 2\max_{i\in[n]}\max_{(j,k)\in E}|[(\mathbf{V}_Y-\widehat{\mathbf{\Delta}})_{j\bullet}(\boldsymbol{X}_i\boldsymbol{X}_i^T-\widehat{\mathbf{\Sigma}}_X)\mathbf{V}_{X,k\bullet}^T-(\mathbf{V}_Y-\mathbf{\Delta})_{j\bullet}(\boldsymbol{X}_i\boldsymbol{X}_i^T-\mathbf{\Sigma}_X)\mathbf{V}_{X,k\bullet}^T]|^2\\
&\quad+2\max_{i\in[n]}\max_{(j,k)\in E}|(\mathbf{V}_{Y,j\bullet}(\boldsymbol{Y}_i\boldsymbol{Y}_i^T-\widehat{\mathbf{\Sigma}}_Y)(\mathbf{V}_X+\widehat{\mathbf{\Delta}})_{k\bullet}^T-\mathbf{V}_{Y,j\bullet}(\boldsymbol{Y}_i\boldsymbol{Y}_i^T-\mathbf{\Sigma}_Y)(\mathbf{V}_X+\mathbf{\Delta})_{k\bullet}^T]|^2.
\end{aligned}
\tag{B.2}
$$

Denote the right hand side of (B.2) as V_M. We now bound each term of V_M. For the first term, we have

$$
\begin{aligned}
&(\mathbf{V}_Y-\widehat{\mathbf{\Delta}})_{j\bullet}(\boldsymbol{X}_i\boldsymbol{X}_i^T-\widehat{\mathbf{\Sigma}}_X)\mathbf{V}_{X,k\bullet}^T-(\mathbf{V}_Y-\mathbf{\Delta})_{j\bullet}(\boldsymbol{X}_i\boldsymbol{X}_i^T-\mathbf{\Sigma}_X)\mathbf{V}_{X,k\bullet}^T\\
=&(\mathbf{V}_Y-\widehat{\mathbf{\Delta}})_{j\bullet}(\mathbf{\Sigma}_X-\widehat{\mathbf{\Sigma}}_X)\mathbf{V}_{X,k\bullet}^T-(\widehat{\mathbf{\Delta}}-\mathbf{\Delta})_{j\bullet}(\boldsymbol{X}_i\boldsymbol{X}_i^T-\mathbf{\Sigma}_X)\mathbf{V}_{X,k\bullet}^T.
\end{aligned}
$$

Moreover, with probability at least $1-c/d$, we have

$$
\max_{(j,k)\in E}|(\mathbf{V}_Y-\widehat{\mathbf{\Delta}})_{j\bullet}|_1\le\|\mathbf{V}_Y\|_\infty+\|\mathbf{\Delta}\|_\infty+\|\widehat{\mathbf{\Delta}}-\mathbf{\Delta}\|_\infty=O(1),
$$

$$
\|\mathbf{\Sigma}_X-\widehat{\mathbf{\Sigma}}_X\|_{\max}=O(\sqrt{\log d/n}),\ \text{and}\ \max_{(j,k)\in E}|\mathbf{V}_{X,k\bullet}|_1\le\|\mathbf{V}_X\|_\infty=O(1).
$$

Thus, we have with probability at least $1-c/d$,

$$
\begin{aligned}
&\max_{i\in[n]}\max_{(j,k)\in E}|(\mathbf{V}_Y-\widehat{\mathbf{\Delta}})_{j\bullet}(\mathbf{\Sigma}_X-\widehat{\mathbf{\Sigma}}_X)\mathbf{V}_{X,k\bullet}^T|\\
\le&\max_{(j,k)\in E}|(\mathbf{V}_Y-\widehat{\mathbf{\Delta}})_{j\bullet}|_1\|\mathbf{\Sigma}_X-\widehat{\mathbf{\Sigma}}_X\|_{\max}\max_{(j,k)\in E}|\mathbf{V}_{X,k\bullet}|_1=O(\sqrt{\log d/n}).
\end{aligned}
\tag{B.3}
$$

From Lemma A.2, we have with probability at least $1-(dn)^{-1}$, $\|\boldsymbol{X}_i\boldsymbol{X}_i^T-\mathbf{\Sigma}_X\|_{\max}=O(\log(dn))$. This implies that with probability at least $1-1/d$, $\max_{i\in[n]}\|\boldsymbol{X}_i\boldsymbol{X}_i^T-\mathbf{\Sigma}_X\|_{\max}=O(\log(dn))$. In addition, we have with probability no smaller than $1-c/d$,

$$
|(\widehat{\mathbf{\Delta}}-\mathbf{\Delta})_{j\bullet}|_1\le\|\widehat{\mathbf{\Delta}}-\mathbf{\Delta}\|_\infty=O(s\sqrt{\log d/n}),\ \text{and}\ \max_{(j,k)\in E}|\mathbf{V}_{X,k\bullet}|_1\le\|\mathbf{V}_X\|_\infty=O(1),
$$

Therefore, we have with probability no smaller than $1 - c/d$,

$$\begin{aligned}
&\max_{i\in[n]} \max_{(j,k)\in E} |(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})_{j\bullet}(\boldsymbol{X}_i\boldsymbol{X}_i^T - \boldsymbol{\Sigma}_X)\mathbf{V}_{X,k\bullet}^T| \\
\le& |(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})_{j\bullet}|_1 \max_{i\in[n]} \|\boldsymbol{X}_i\boldsymbol{X}_i^T - \boldsymbol{\Sigma}_X\|_{\max} \max_{(j,k)\in E} |\mathbf{V}_{X,k\bullet}|_1 = O\Big(\frac{s\log(dn)}{\sqrt{n}}\Big).
\end{aligned} \tag{B.4}$$

The second term in (B.2) can be bounded using a similar argument.

Next, we define the event $\mathcal{E}$ about V_M as $\mathcal{E} = \{V_M \le C\frac{s^2(\log^2(dn))}{n}\}$ where C is a constant. From (B.2), (B.3) and (B.4), we have $\mathbb{P}(\mathcal{E}^c) \le \frac{c}{d}$, where c is a constant related to C. By Lemma A.1 in Chernozhukov et al. (2013), under the event $\mathcal{E}$, recall that $f(\mathcal{D}_1, \xi)$ is defined in (B.1) and we have

$$E_M := \mathbb{E}\Big[\max_{(j,k)\in E} f(\mathcal{D}_1,\xi)\Big|\mathcal{D}_1\Big] \lesssim \sqrt{nV_M}\sqrt{\log d/n} = O\Bigg(\sqrt{\frac{s^2(\log^3(dn))}{n}}\Bigg). \tag{B.5}$$

Applying Borell's inequality (see Proposition A.2.1 in Van Der Vaart and Wellner (1996)), we have $\forall t > 0$,

$$\mathbb{P}\Big(\max_{(j,k)\in E} |f(\mathcal{D}_1,\xi) - E_M| > t\Big|\mathcal{D}_1\Big) \le 2\exp\Big(-\frac{t^2}{2\sigma_\xi^2}\Big),$$

where the conditional variance σ_ξ^2 is the left side in (B.2). From (B.2) and (B.5), setting $t = \sqrt{c\log(2d)\sigma_\xi^2}$, we have under the event $\mathcal{E}$,

$$\mathbb{P}\Bigg(\max_{(j,k)\in E} f(\mathcal{D}_1,\xi) > C\sqrt{\frac{s^2(\log^3(dn))}{n}}\Bigg|\mathcal{D}_1\Bigg) \le \frac{c}{d}.$$

From (B.1), we have

$$\mathbb{P}\Bigg(\mathbb{P}\Bigg(|T^B - T_0^B| > C\sqrt{\frac{s^2(\log^3(dn))}{n}}\Bigg|\mathcal{D}_1\Bigg) > \frac{c}{d}\Bigg) \le \mathbb{P}(\mathcal{E}^c) \le \frac{c}{d},$$

as desired. Therefore, by Corolllary 3.1 of Chernozhukov et al. (2013), we have

$$\lim_{n\to\infty} |\mathbb{P}(T^E > c(\alpha, E)) - \alpha| = 0,$$

as desired.

C Proof of Lemmas A.1–A.6

In this section, we provide proofs for Lemmas A.1–A.6 provided in Section A. We start with a lemma on the the basic property of the fourth order moments of a multivariate Gaussian random variables.

Lemma C.1. For a random vector $\boldsymbol{X} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and four matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$, we have

$$\mathbb{E}[(\mathbf{A}\boldsymbol{X})(\mathbf{B}\boldsymbol{X})^T(\mathbf{C}\boldsymbol{X})(\mathbf{D}\boldsymbol{X})^T] = (\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^T)(\mathbf{C}\boldsymbol{\Sigma}\mathbf{D}^T) + (\mathbf{A}\boldsymbol{\Sigma}\mathbf{C}^T)(\mathbf{B}\boldsymbol{\Sigma}\mathbf{D}^T) + \mathrm{tr}\big(\mathbf{B}\boldsymbol{\Sigma}\mathbf{C}^T\big)(\mathbf{A}\boldsymbol{\Sigma}\mathbf{D}^T).$$

Proof. The proof can be found in Isserlis (1916). □

C.1 Proof of Lemma A.1

Proof. Using the equality $\boldsymbol{\Sigma}_Y\boldsymbol{\Delta}\boldsymbol{\Sigma}_X - \big(\boldsymbol{\Sigma}_X - \boldsymbol{\Sigma}_Y\big) = \mathbf{0}$, from (31), we have

$$\widehat{\boldsymbol{\Delta}}^d - \boldsymbol{\Delta} = -\mathbf{V}_Y(\widehat{\boldsymbol{\Sigma}}_Y\widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\Sigma}}_X - \big(\widehat{\boldsymbol{\Sigma}}_X - \widehat{\boldsymbol{\Sigma}}_Y\big))\mathbf{V}_X^T + (\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}) + \mathbf{V}_Y(\boldsymbol{\Sigma}_Y\boldsymbol{\Delta}\boldsymbol{\Sigma}_X - \big(\boldsymbol{\Sigma}_X - \boldsymbol{\Sigma}_Y\big))\mathbf{V}_X^T.$$

Note that $\widehat{\boldsymbol{\Sigma}}_Y\widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\Sigma}}_X = \widehat{\boldsymbol{\Sigma}}_Y(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})\widehat{\boldsymbol{\Sigma}}_X + \widehat{\boldsymbol{\Sigma}}_Y\boldsymbol{\Delta}\widehat{\boldsymbol{\Sigma}}_X$. Thus, we have

$$\widehat{\boldsymbol{\Sigma}}_Y\boldsymbol{\Delta}\widehat{\boldsymbol{\Sigma}}_X - \boldsymbol{\Sigma}_Y\boldsymbol{\Delta}\boldsymbol{\Sigma}_X = (\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y)\boldsymbol{\Delta}(\widehat{\boldsymbol{\Sigma}}_X - \boldsymbol{\Sigma}_X) + (\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y)\boldsymbol{\Delta}\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_Y\boldsymbol{\Delta}(\widehat{\boldsymbol{\Sigma}}_X - \boldsymbol{\Sigma}_X).$$

The result in Lemma A.1 can then be obtained after some algebraic manipulation. □

C.2 Proof of Lemma A.3

Proof. We first show that $\boldsymbol{\Theta}_X$ and $\boldsymbol{\Theta}_Y$ are feasible solutions for the optimization problems in (34), respectively. For two matrices $\mathbf{A}$ and $\mathbf{B}$, we have $\|\mathbf{A}\mathbf{B}\|_{\max} \le \|\mathbf{A}^T\|_1\|\mathbf{B}\|_{\max}$ and $\|\mathbf{A}\mathbf{B}\|_{\max} \le \|\mathbf{A}\|_\infty\|\mathbf{B}\|_{\max}$. By the aforementioned inequality, Lemma A.2, and the fact that $\boldsymbol{\Theta}_X$ is symmetric, we have

$$\|\boldsymbol{\Theta}_X\widehat{\boldsymbol{\Sigma}}'_X - \mathbf{I}\|_{\max} = \|\boldsymbol{\Theta}_X(\widehat{\boldsymbol{\Sigma}}'_X - \boldsymbol{\Sigma}_X)\|_{\max} \le \|\boldsymbol{\Theta}_X\|_1\|\widehat{\boldsymbol{\Sigma}}'_X - \boldsymbol{\Sigma}_X\|_{\max} = O(\sqrt{\log d/n}),$$

with probability at least $1-c/d$ where c is some positive constant. Since we select $\lambda' = C'\sqrt{\log d/n}$, where C' is a sufficiently large constant, we know that $\mathbf{\Theta}_X$ and $\mathbf{\Theta}_Y$ are feasible solutions for optimization problem (34) with probability at least $1-c/d$.

Since $\mathbf{V}_X$ and $\mathbf{V}_Y$ are minimizers of the optimization problem (34) and by Assumption 1, we have with probability at least $1-c/d$ that $\|\mathbf{V}_X\|_\infty \le \|\mathbf{\Theta}_X\|_\infty = \|\mathbf{\Theta}_X\|_1 = O(1)$, $\|\mathbf{V}_Y\|_\infty = O(1)$, and $\|\mathbf{V}\|_\infty = \|\mathbf{V}_X\|_\infty \|\mathbf{V}_Y\|_\infty = O(1)$, where $\mathbf{V} = \mathbf{V}_X \otimes \mathbf{V}_Y$.

Finally, by the triangle inequality and Lemma A.2, we have with probability at least $1-c/d$ that

$$\|\mathbf{V}_X\widehat{\mathbf{\Sigma}}_X - \mathbf{I}\|_{\max} \le \|\mathbf{V}_X\widehat{\mathbf{\Sigma}}'_X - \mathbf{I}\|_{\max} + \|\mathbf{V}_X(\widehat{\mathbf{\Sigma}}'_X - \mathbf{\Sigma}_X)\|_{\max} + \|\mathbf{V}_X(\widehat{\mathbf{\Sigma}}_X - \mathbf{\Sigma}_X)\|_{\max} = O(\sqrt{\log d/n}),$$

where we use the inequality $\|\mathbf{V}_X(\widehat{\mathbf{\Sigma}}_X - \mathbf{\Sigma}_X)\|_{\max} \le \|\mathbf{V}_X\|_\infty \|\widehat{\mathbf{\Sigma}}_X - \mathbf{\Sigma}_X\|_{\max}$. □

C.3 Proof of Lemma A.4

Proof. Throughout the proof, for simplicity, we will use the notation $O_P(\cdot)$ to indicate that an event occurs with probability at least $1-c/d$. The remainder term in (33) is

$$\begin{aligned} \text{Remainder} = & \underbrace{[(\widehat{\mathbf{\Delta}} - \mathbf{\Delta}) - \mathbf{V}_Y\widehat{\mathbf{\Sigma}}_Y(\widehat{\mathbf{\Delta}} - \mathbf{\Delta})(\mathbf{V}_X\widehat{\mathbf{\Sigma}}_X)^T]}_{I_1} - \underbrace{\mathbf{V}_Y(\widehat{\mathbf{\Sigma}}_Y - \mathbf{\Sigma}_Y)\mathbf{\Delta}(\widehat{\mathbf{\Sigma}}_X - \mathbf{\Sigma}_X)\mathbf{V}_X^T}_{I_2} \\ & - \underbrace{\mathbf{V}_Y(\widehat{\mathbf{\Sigma}}_Y - \mathbf{\Sigma}_Y)\mathbf{\Delta}(\mathbf{V}_X\mathbf{\Sigma}_X - \mathbf{I})^T}_{I_3} - \underbrace{(\mathbf{V}_Y\mathbf{\Sigma}_Y - \mathbf{I})\mathbf{\Delta}(\widehat{\mathbf{\Sigma}}_X - \mathbf{\Sigma}_X)\mathbf{V}_X^T}_{I_4}. \end{aligned}$$

We will provide upper bounds for I_1, I_2, I_3, and I_4 under the max norm, respectively.

We start with obtaining an upper bound for $\|I_1\|_{\max}$. First, note that

$$\begin{aligned} \|I_1\|_{\max} &= \|(\widehat{\mathbf{\Delta}} - \mathbf{\Delta}) - \mathbf{V}_Y\widehat{\mathbf{\Sigma}}_Y(\widehat{\mathbf{\Delta}} - \mathbf{\Delta})(\mathbf{V}_X\widehat{\mathbf{\Sigma}}_X)^T\|_{\max} \\ &= \|\{\mathbf{I} - \mathbf{V}(\widehat{\mathbf{\Sigma}}_X \otimes \widehat{\mathbf{\Sigma}}_Y)\}\text{vec}(\widehat{\mathbf{\Delta}} - \mathbf{\Delta})\|_\infty \\ &\le \|\mathbf{I} - \mathbf{V}(\widehat{\mathbf{\Sigma}}_X \otimes \widehat{\mathbf{\Sigma}}_Y)\|_{\max} \cdot \|\text{vec}(\widehat{\mathbf{\Delta}} - \mathbf{\Delta})\|_1, \end{aligned}$$

where the second equality holds by vectorizing the product of matrices. It suffices to obtain upper bounds for $\|\mathbf{I} - \mathbf{V}(\widehat{\mathbf{\Sigma}}_X \otimes \widehat{\mathbf{\Sigma}}_Y)\|_{\max}$ and $\|\text{vec}(\widehat{\mathbf{\Delta}} - \mathbf{\Delta})\|_1$. By Lemma A.2, $\|\text{vec}(\widehat{\mathbf{\Delta}} - \mathbf{\Delta})\|_1 = O_P(s\sqrt{\log d/n})$. Next, we bound the term $\|\mathbf{I} - \mathbf{V}(\widehat{\mathbf{\Sigma}}_X \otimes \widehat{\mathbf{\Sigma}}_Y)\|_{\max}$.

By the triangle inequality, we have

$$\|\mathbf{V}\big(\widehat{\boldsymbol{\Sigma}}_X \otimes \widehat{\boldsymbol{\Sigma}}_Y\big) - \mathbf{I}\|_{\max} = \|\mathbf{V}\big\{\big(\widehat{\boldsymbol{\Sigma}}_X \otimes \widehat{\boldsymbol{\Sigma}}_Y\big) - \big(\boldsymbol{\Sigma}_X \otimes \boldsymbol{\Sigma}_Y\big)\big\}\|_{\max} + \|\mathbf{V}\big\{\big(\boldsymbol{\Sigma}_X \otimes \boldsymbol{\Sigma}_Y\big) - \big(\widehat{\boldsymbol{\Sigma}}'_X \otimes \widehat{\boldsymbol{\Sigma}}'_Y\big)\big\}\|_{\max} + \|\mathbf{V}\big(\widehat{\boldsymbol{\Sigma}}'_X \otimes \widehat{\boldsymbol{\Sigma}}'_Y\big) - \mathbf{I}\|_{\max}.$$

To bound the above term, we start with some intermediate results obtained from Lemma A.2 and Assumption 1:

$$\begin{aligned}
\|(\widehat{\boldsymbol{\Sigma}}_X - \boldsymbol{\Sigma}_X) \otimes \boldsymbol{\Sigma}_Y\|_{\max} &= O_P(\sqrt{\log d/n});\\
\|\boldsymbol{\Sigma}_X \otimes (\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y)\|_{\max} &= O_P(\sqrt{\log d/n});\\
\|(\widehat{\boldsymbol{\Sigma}}_X - \boldsymbol{\Sigma}_X) \otimes (\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y)\|_{\max} &= O_P(\log d/n).
\end{aligned}$$

Moreover, it can be shown that

$$\|\widehat{\boldsymbol{\Sigma}}_X \otimes \widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_X \otimes \boldsymbol{\Sigma}_Y\|_{\max} = O_P(\sqrt{\log d/n}); \quad \|\widehat{\boldsymbol{\Sigma}}'_X \otimes \widehat{\boldsymbol{\Sigma}}'_Y - \boldsymbol{\Sigma}_X \otimes \boldsymbol{\Sigma}_Y\|_{\max} = O_P(\sqrt{\log d/n}).$$

The first term can then be upper bounded as

$$\|\mathbf{V}\big(\widehat{\boldsymbol{\Sigma}}_X \otimes \widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_X \otimes \boldsymbol{\Sigma}_Y\big)\|_{\max} \leq \|\mathbf{V}\|_\infty \|\big(\widehat{\boldsymbol{\Sigma}}_X \otimes \widehat{\boldsymbol{\Sigma}}_Y\big) - \big(\boldsymbol{\Sigma}_X \otimes \boldsymbol{\Sigma}_Y\big)\|_{\max} = O_P(\sqrt{\log d/n}).$$

The second term can be proven similarly. Finally, since $\lambda' = C'\sqrt{\log d/n}$ for some sufficient large constant C', from (34), we have $\|((\mathbf{V}_X\widehat{\boldsymbol{\Sigma}}'_X) \otimes (\mathbf{V}_Y\widehat{\boldsymbol{\Sigma}}'_Y)) - \mathbf{I} \otimes \mathbf{I}\|_{\max} = O_P(\sqrt{\log d/n})$, which implies $\|\mathbf{V}\big(\widehat{\boldsymbol{\Sigma}}'_X \otimes \widehat{\boldsymbol{\Sigma}}'_Y\big) - \mathbf{I}\|_{\max} = O_P(\sqrt{\log d/n})$. Combining the above, we have $\|I_1\|_{\max} = O_P(s\log d/n)$.

For I_2, using a similar argument as the upper bound for I_1, we have

$$\begin{aligned}
\|\mathbf{V}_Y(\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y)\boldsymbol{\Delta}(\widehat{\boldsymbol{\Sigma}}_X - \boldsymbol{\Sigma}_X)\mathbf{V}_X^T\|_{\max} &= \|\mathbf{V}[\big(\widehat{\boldsymbol{\Sigma}}_X - \boldsymbol{\Sigma}_X\big) \otimes \big(\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y\big)]\mathrm{vec}(\boldsymbol{\Delta})\|_\infty\\
&\leq \|\mathbf{V}[\big(\widehat{\boldsymbol{\Sigma}}_X - \boldsymbol{\Sigma}_X\big) \otimes \big(\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y\big)]\|_{\max}\|\mathrm{vec}(\boldsymbol{\Delta})\|_1\\
&\leq sM \cdot \|\mathbf{V}[\big(\widehat{\boldsymbol{\Sigma}}_X - \boldsymbol{\Sigma}_X\big) \otimes \big(\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y\big)]\|_{\max}\\
&= O_P(s\log d/n),
\end{aligned}$$

where the second inequality holds under the family of true differential graphs defined in (41).

For I_3, similar to the proof of Lemma A.3, we have $\|\mathbf{V}_X\boldsymbol{\Sigma}_X - \mathbf{I}\|_{\max} = O_P(\sqrt{\log d/n})$. Moreover,

$\|\mathbf{V}_Y(\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y)\|_{\max} \le \|\mathbf{V}_Y\|_\infty \|\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y\|_{\max} = O_P(\sqrt{\log d/n})$. Coupling the above with the fact that $\|\boldsymbol{\Delta}\|_{1,1} \le sM = O(s)$ under the condition in (41), we obtain

$$\|I_3\|_{\max} \le \|\mathbf{V}_Y(\widehat{\boldsymbol{\Sigma}}_Y - \boldsymbol{\Sigma}_Y)\|_{\max}\|\boldsymbol{\Delta}\|_{1,1}\|\mathbf{V}_X\boldsymbol{\Sigma}_X - \mathbf{I}\|_{\max} = O_P(s\log d/n).$$

The upper bound for I_4 can be derived using the same argument as the above and is omitted.

Putting the upper bounds for I_1, I_2, I_3, and I_4 together, we obtain the desired result. □

C.4 Proof of Lemma A.5

Proof. We first show that conditional on $\mathcal{D}_2$, $\mathrm{Var}(\sqrt{n}\mathrm{Leading}_{jk}) = \xi_{jk}^2$ for all $j, k \in [d]$. From the definition in (32), it can be shown that $\mathbb{E}(\sqrt{n}\mathrm{Leading}_{jk}) = 0$. From the fact that conditional on $\mathcal{D}_2$, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are i.i.d. samples and $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are i.i.d. samples, after some algebraic manipulation, we have

$$\mathrm{Var}(\sqrt{n}\mathrm{Leading}_{jk}) = \underbrace{\mathrm{Var}(\mathbf{e}_j^T\mathbf{V}_Y\big(\boldsymbol{Y}\boldsymbol{Y}^T - \boldsymbol{\Sigma}_Y\big)(\mathbf{V}_X + \boldsymbol{\Delta})^T\mathbf{e}_k)}_{I_1} + \underbrace{\mathrm{Var}(\mathbf{e}_j^T(\mathbf{V}_Y - \boldsymbol{\Delta})\big(\boldsymbol{X}\boldsymbol{X}^T - \boldsymbol{\Sigma}_X\big)\mathbf{V}_X^T\mathbf{e}_k)}_{I_2}.$$

The term I_1 can be computed as

$$\begin{aligned} I_1 &= \mathbb{E}(\mathbf{e}_j^T\mathbf{V}_Y\big(\boldsymbol{Y}\boldsymbol{Y}^T - \boldsymbol{\Sigma}_Y\big)(\mathbf{V}_X + \boldsymbol{\Delta})^T\mathbf{e}_k\mathbf{e}_j^T\mathbf{V}_Y\big(\boldsymbol{Y}\boldsymbol{Y}^T - \boldsymbol{\Sigma}_Y\big)(\mathbf{V}_X + \boldsymbol{\Delta})^T\mathbf{e}_k) \\ &= I_{11} - (\mathbf{e}_j^T\mathbf{V}_Y\boldsymbol{\Sigma}_Y(\mathbf{V}_X + \boldsymbol{\Delta})^T\mathbf{e}_k\mathbf{e}_j^T\mathbf{V}_Y\boldsymbol{\Sigma}_Y(\mathbf{V}_X + \boldsymbol{\Delta})^T\mathbf{e}_k), \end{aligned}$$

where $I_{11} = \mathbb{E}(\big(\mathbf{e}_j^T\mathbf{V}_Y\boldsymbol{Y}\big)\big(\mathbf{e}_k^T(\mathbf{V}_X + \boldsymbol{\Delta})\boldsymbol{Y}\big)^T\big(\mathbf{e}_j^T\mathbf{V}_Y\boldsymbol{Y}\big)\big(\mathbf{e}_k^T(\mathbf{V}_X + \boldsymbol{\Delta})\boldsymbol{Y}\big)^T)$. Applying Lemma C.1 to I_{11}, we obtain

$$I_1 = (\mathbf{V}_{Y,j\bullet}\boldsymbol{\Sigma}_Y\mathbf{V}_{Y,j\bullet}^T)((\mathbf{V}_X + \boldsymbol{\Delta})_{k\bullet}\boldsymbol{\Sigma}_Y(\mathbf{V}_X + \boldsymbol{\Delta})_{k\bullet}^T) + (\mathbf{V}_{Y,j\bullet}\boldsymbol{\Sigma}_Y(\mathbf{V}_X + \boldsymbol{\Delta})_{k\bullet}^T)^2.$$

The term I_2 can be computed similarly, and thus conditional on $\mathcal{D}_2$, $\mathrm{Var}(\sqrt{n}\mathrm{Leading}_{jk}) = \xi_{jk}^2$.

Next, we show that $1/\xi_{jk} = O_P(1)$. Similar to the proof of A.3, we have with probability at least $1 - c/d$ for a universal constant c, $\|\mathbf{V}_X\boldsymbol{\Sigma}_X - \mathbf{I}\|_{\max} = O(\sqrt{\log d/n})$. And under the conditions in Lemma A.5, we then know with probability at least $1 - c/d$,

$$\|\mathbf{V}_X - \mathbf{\Theta}_X\|_{\max} = \|(\mathbf{V}_X\mathbf{\Sigma}_X - \mathbf{I})\mathbf{\Theta}_X\|_{\max} \le \|\mathbf{V}_X\mathbf{\Sigma}_X - \mathbf{I}\|_{\max}\|\mathbf{\Theta}_X\|_1 = O(\sqrt{\log d/n}).$$

And we have $\min_{s\in[d]}\{\mathbf{\Theta}_{X,ss}\} \ge \lambda_{\min}\{(\mathbf{\Theta}_X)\} = 1/\lambda_{\max}\{\mathbf{\Sigma}_X\} = 1/\|\mathbf{\Theta}_X\|_2$. Therefore, with probability at least $1 - c/d$, there exists $K > 0$, such that

$$\min_{s\in[d]}\|\mathbf{V}_{X,s\bullet}\|_2 \ge \min_{s\in[d]}\{|\mathbf{V}_{X,ss}|\} \ge \min_{s\in[d]}\{\mathbf{\Theta}_{X,ss}\} - \|\mathbf{V}_X - \mathbf{\Theta}_X\|_{\max} \ge K.$$

Similarly, we have $\|\mathbf{V}_Y - \mathbf{\Delta} - \mathbf{\Theta}_X\|_{\max} = \|\mathbf{V}_Y - \mathbf{\Theta}_Y\|_{\max} = O_P(\sqrt{\log d/n})$. Combining the above with Assumption 1, with probability at least $1 - c/d$, there exists $K > 0$, such that

$$\min_{s\in[d]}\|(\mathbf{V}_Y - \mathbf{\Delta})_{s\bullet}\|_2 \ge \min_{s\in[d]}\{|(\mathbf{V}_Y - \mathbf{\Delta})_{ss}|\} \ge \min_{s\in[d]}\{\mathbf{\Theta}_{X,ss}\} - \|\mathbf{V}_Y - \mathbf{\Delta} - \mathbf{\Theta}_X\|_{\max} \ge K.$$

Under Assumption 1, there exists $\lambda_0 > 0$, such that $\lambda_{\min}(\mathbf{\Sigma}_X) = 1/\|\mathbf{\Theta}_X\|_2 > \lambda_0$. Therefore, we know that with probability at least $1 - c/d$, there exists $K_1 > 0$, such that

$$((\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}\mathbf{\Sigma}_X(\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}^T)(\mathbf{V}_{X,k\bullet}\mathbf{\Sigma}_X\mathbf{V}_{X,k\bullet}^T) > K_1.$$

Similarly, with probability at least $1 - c/d$, there exists $K_2 > 0$, such that $(\mathbf{V}_{Y,j\bullet}\mathbf{\Sigma}_Y\mathbf{V}_{Y,j\bullet}^T)((\mathbf{V}_X + \mathbf{\Delta})_{k\bullet}\mathbf{\Sigma}_Y(\mathbf{V}_X + \mathbf{\Delta})_{k\bullet}^T) > K_2$. The lemma then follows from these two inequalities. □

C.5 Proof of Lemma A.6

Proof. Define $I_{jk} = (\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}(\boldsymbol{X}\boldsymbol{X}^T - \mathbf{\Sigma}_X)\mathbf{V}_{X,k\bullet}^T$, $J_{jk} = -\mathbf{V}_{Y,j\bullet}(\boldsymbol{Y}\boldsymbol{Y}^T - \mathbf{\Sigma}_Y)(\mathbf{V}_X + \mathbf{\Delta})_{k\bullet}^T$. Since $|Z_{jk}|^3 \le 4(|I_{jk}|^3 + |J_{jk}|^3)$, it suffices to prove

$$\mathbb{E}\big(|I_{jk}|^3\big|\mathcal{D}_2 = (\mathbb{X}^T, \mathbb{Y}^T)^T\big) = O(1), \text{ and } \mathbb{E}\big(|J_{jk}|^3\big|\mathcal{D}_2 = (\mathbb{X}^T, \mathbb{Y}^T)^T\big) = O(1).$$

In the following, we will show that $\mathbb{E}\big(|I_{jk}|^3\big|\mathcal{D}_2 = (\mathbb{X}^T, \mathbb{Y}^T)^T\big) = O(1)$. A similar proof technique can be employed to prove $\mathbb{E}\big(|J_{jk}|^3\big|\mathcal{D}_2 = (\mathbb{X}^T, \mathbb{Y}^T)^T\big) = O(1)$. Conditioned on $\mathcal{D}_2 = (\mathbb{X}^T, \mathbb{Y}^T)^T$, from Lemma A.3 and Assumption 1, we have

$$\|\mathbf{V}_{X,k\bullet}\|_1 \le \|\mathbf{V}_X\|_\infty \le \|\mathbf{\Theta}_X\|_\infty = O(1), \text{ and } \|\mathbf{V}_{Y,j\bullet}\|_1 = O(1).$$

By Assumption 1, we obtain $\|\mathbf{V}_Y - \mathbf{\Delta}\|_1 = O(1)$, and $\|\mathbf{V}_X + \mathbf{\Delta}\|_1 = O(1)$. We define $I_{jk,1} = (\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}\boldsymbol{X}$, and $I_{jk,2} = \mathbf{V}_{X,k\bullet}\boldsymbol{X}$. Then $I_{jk,1} \sim N(0, (\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}\mathbf{\Sigma}_X(\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}^T)$, and $I_{jk,2} \sim N(\mathbf{V}_{X,k\bullet}\mathbf{\Sigma}_X\mathbf{V}_{X,k\bullet}^T)$. From Condition 1, we have $\|\mathbf{\Sigma}_X\|_2 = O(1)$. We also have $\|(\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}\|_2 \le \|(\mathbf{V}_Y - \mathbf{\Delta})_{j\bullet}\|_1 = O(1)$, and $\|\mathbf{V}_{X,k\bullet}\|_2 \le \|\mathbf{V}_{X,k\bullet}\|_1 = O(1)$. Therefore, both the variance of $I_{jk,1}$ and that of $I_{jk,2}$ are $O(1)$, which means that $\|I_{jk,1}\|_{\psi_2} = O(1)$ and $\|I_{jk,2}\|_{\psi_2} = O(1)$. Then from the centering property of the sub-exponential norm and the inequality $\|XY\|_{\psi_1} \le 2\|X\|_{\psi_2}\|Y\|_{\psi_2}$ which can be found in the proof of Lemma 23 in Javanmard and Montanari (2014), we know that $I_{jk} = I_{jk,1}I_{jk,2} - \mathbb{E}(I_{jk,1}I_{jk,2})$ satisfies $\|I_{jk}\|_{\psi_1} = O(1)$, which implies $\mathbb{E}\big(|I_{jk}|^3 \big| \mathcal{D}_2 = (\mathbb{X}^T, \mathbb{Y}^T)^T\big) = O(1)$ based on the definition of ψ_1-norm, as desired. □

D Additional numeric results for Section 5.1

Tables 1–4 present the numerical results for Section 5.1, when $n = \{400, 600, 800\}$, $p = \{1600, 3600\}$, and SP $= \{0.005, 0.01, 0.015, 0.02, 0.025\}$ under the sparse and dense models respectively. We see from Table 1 and 3, only our proposed estimator is able to control the type I error at around $\alpha = 0.05$ with minimal type II error. To evaluate SP, as it increases from 0.005 to 0.025, both type I error and type II error for three methods would increase. When the model switches from sparse to dense, the type I error of the naive method increases from around 0.08 to around 0.12 for various n and p, in which this influence is much more heavy than the other two methods. For the cross-fitting method, although it has smaller type II error, its type I error is always larger than our proposed algorithm. As our algorithm controls the type I error at around $\alpha = 0.05$, the cross-fitting method, however, could not provide this guarantee and lead to higher type I errors such as 0.640 in Table 3 even when the sample size is 800. For our proposed method, it not only is able to control the type I error, but also ensures the type II error approaching zero as n is increased. In addition, we see from Table 2 and 4, the average length of confidence interval over $\mathcal{S}$ and $\mathcal{S}^c$ will also increase as SP increases. Although the naive method and the cross-fitting method provides narrower confidence intervals, these intervals are not valid for the given confidence, since the naive method assumes that both the network $\mathbf{\Theta}_X$ and $\mathbf{\Theta}_Y$ are sparse while our constructed models violate this assumption, and the cross-fitting method considers two dependent debiased estimator as independent.

References

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics* **41** 2786–2819.

Durrett, R. (2019). *Probability: theory and examples*, vol. 49. Cambridge university press.

Isserlis, L. (1916). On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression. *Biometrika* **11** 185–190.

Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research* **15** 2869–2909.

Ma, C., Lu, J. and Liu, H. (2017). Inter-subject analysis: Inferring sparse interactions with dense intra-graphs. *arXiv preprint arXiv:1709.07036* .

Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer.

Zhao, S. D., Cai, T. T. and Li, H. (2014). Direct estimation of differential networks. *Biometrika* **101** 253–268.

Table 1: Type I errors and Type II errors for three methods, averaged over 100 replications, under the sparse model.

			SP=0.005		SP=0.01		SP=0.015		SP=0.02		SP=0.025	
p	n	method	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1600	400	Proposed	0.0502	0.0500	0.0520	0.1010	0.0532	0.1062	0.0540	0.1125	0.0546	0.1482
		Naive	0.0747	0.0075	0.0751	0.0060	0.0744	0.0142	0.0720	0.0144	0.0749	0.0210
		Cross-fitting	0.0557	0.0000	0.0595	0.0000	0.0611	0.0025	0.0637	0.0003	0.0649	0.0013
	600	Proposed	0.0555	0.0000	0.0519	0.0130	0.0569	0.0054	0.0542	0.0283	0.0531	0.0522
		Naive	0.0677	0.0000	0.0682	0.0000	0.0676	0.0000	0.0677	0.0006	0.0670	0.0020
		Cross-fitting	0.0539	0.0000	0.0561	0.0000	0.0577	0.0000	0.0604	0.0000	0.0617	0.0000
	800	Proposed	0.0506	0.0000	0.0521	0.0035	0.0515	0.0050	0.0540	0.0072	0.0523	0.0132
		Naive	0.0644	0.0000	0.0648	0.0000	0.0649	0.0000	0.0644	0.0000	0.0642	0.0010
		Cross-fitting	0.0530	0.0000	0.0568	0.0000	0.0554	0.0000	0.0588	0.0000	0.0595	0.0000
3600	400	Proposed	0.0520	0.0450	0.0522	0.1808	0.0579	0.0898	0.0539	0.2152	0.0555	0.2160
		Naive	0.0844	0.0267	0.0855	0.0317	0.0863	0.0383	0.0865	0.0500	0.0853	0.0480
		Cross-fitting	0.0546	0.0022	0.0587	0.0017	0.0625	0.0054	0.0641	0.0056	0.0680	0.0076
	600	Proposed	0.0496	0.0372	0.0516	0.0581	0.0530	0.0688	0.0535	0.0837	0.0543	0.0948
		Naive	0.0743	0.0011	0.0746	0.0033	0.0741	0.0050	0.0745	0.0068	0.0740	0.0102
		Cross-fitting	0.0519	0.0000	0.0557	0.0000	0.0586	0.0000	0.0616	0.0003	0.0639	0.0009
	800	Proposed	0.0498	0.0089	0.0505	0.0233	0.0510	0.0262	0.0554	0.0074	0.0544	0.0346
		Naive	0.0685	0.0000	0.0679	0.0006	0.0685	0.0013	0.0691	0.0006	0.0681	0.0015
		Cross-fitting	0.0521	0.0000	0.0536	0.0000	0.0566	0.0000	0.0586	0.0000	0.0606	0.0003

Table 2: The average length of the confidence intervals over $\mathcal{S}$ and $\mathcal{S}^c$ for three methods, averaged over 100 replications, under the sparse model.

			SP=0.005		SP=0.01		SP=0.015		SP=0.02		SP=0.025	
p	n	method	$\text{Avglen}_{\mathcal{S}}$	$\text{Avglen}_{\mathcal{S}^c}$	$\text{Avglen}_{\mathcal{S}}$	$\text{Avglen}_{\mathcal{S}^c}$	$\text{Avglen}_{\mathcal{S}}$	$\text{Avglen}_{\mathcal{S}^c}$	$\text{Avglen}_{\mathcal{S}}$	$\text{Avglen}_{\mathcal{S}^c}$	$\text{Avglen}_{\mathcal{S}}$	$\text{Avglen}_{\mathcal{S}^c}$
1600	400	Proposed	1.1034	1.1254	1.2323	1.2529	1.2709	1.2953	1.2926	1.3105	1.4076	1.4241
		Naive	0.8692	0.8618	0.9519	0.9460	0.9812	0.9763	0.9922	0.9864	1.0654	1.0624
		Cross-fitting	0.5225	0.5546	0.5861	0.6202	0.6114	0.6408	0.6193	0.6446	0.6772	0.7063
	600	Proposed	0.6401	0.6723	0.9693	0.9883	0.7440	0.7766	1.0175	1.0327	1.1042	1.1221
		Naive	0.6958	0.6881	0.7601	0.7553	0.7841	0.7793	0.7920	0.7875	0.8496	0.8480
		Cross-fitting	0.4570	0.4793	0.5107	0.5348	0.5320	0.5537	0.5388	0.5570	0.5846	0.6078
	800	Proposed	0.7447	0.7639	0.8320	0.8459	0.8650	0.8747	0.8721	0.8832	0.9422	0.9583
		Naive	0.5960	0.5901	0.6521	0.6475	0.6727	0.6681	0.6793	0.6751	0.7302	0.7269
		Cross-fitting	0.4125	0.4294	0.4600	0.4785	0.4786	0.4951	0.4843	0.4984	0.5265	0.5435
3600	400	Proposed	0.8573	0.8859	1.4971	1.5139	0.9412	0.9692	1.5821	1.5894	1.6117	1.6198
		Naive	1.0815	1.0708	1.1354	1.1297	1.1600	1.1537	1.1837	1.1782	1.2023	1.1978
		Cross-fitting	0.6136	0.6323	0.6567	0.6753	0.6706	0.6912	0.6915	0.7076	0.7075	0.7193
	600	Proposed	1.0961	1.1034	1.1621	1.1738	1.1936	1.2015	1.2233	1.2308	1.2496	1.2539
		Naive	0.8481	0.8418	0.8935	0.8879	0.9121	0.9067	0.9307	0.9260	0.9459	0.9413
		Cross-fitting	0.5362	0.5511	0.5742	0.5880	0.5871	0.6020	0.6033	0.6155	0.6165	0.6256
	800	Proposed	0.9312	0.9354	0.9884	0.9935	1.0118	1.0180	0.7676	0.7795	1.0567	1.0603
		Naive	0.7236	0.7168	0.7607	0.7560	0.7762	0.7721	0.7926	0.7884	0.8054	0.8014
		Cross-fitting	0.4859	0.4961	0.5178	0.5284	0.5293	0.5417	0.5444	0.5535	0.5554	0.5626

Table 3: Type I errors and Type II errors for three methods, averaged over 100 replications, under the dense model.

p	n	method	SP=0.005		SP=0.01		SP=0.015		SP=0.02		SP=0.025	
			Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1600	400	Proposed	0.0468	0.1388	0.0503	0.1745	0.0497	0.1779	0.0519	0.2367	0.0534	0.2778
		Naive	0.1258	0.0000	0.1484	0.0020	0.1502	0.0037	0.1643	0.0050	0.1703	0.0115
		Cross-fitting	0.0506	0.0000	0.0565	0.0025	0.0570	0.0042	0.0622	0.0067	0.0647	0.0122
	600	Proposed	0.0482	0.0037	0.0509	0.0215	0.0505	0.0579	0.0511	0.0908	0.0579	0.0532
		Naive	0.1060	0.0000	0.1284	0.0000	0.1319	0.0000	0.1420	0.0006	0.1508	0.0018
		Cross-fitting	0.0496	0.0000	0.0534	0.0000	0.0553	0.0000	0.0593	0.0000	0.0620	0.0010
	800	Proposed	0.0481	0.0000	0.0519	0.0055	0.0495	0.0192	0.0534	0.0092	0.0572	0.0205
		Naive	0.0959	0.0000	0.1177	0.0000	0.1200	0.0000	0.1287	0.0000	0.1373	0.0000
		Cross-fitting	0.0481	0.0000	0.0518	0.0000	0.0531	0.0000	0.0584	0.0000	0.0609	0.0005
3600	400	Proposed	0.0491	0.1089	0.0546	0.1839	0.0523	0.3250	0.0539	0.3169	0.0607	0.2050
		Naive	0.1368	0.0011	0.1542	0.0158	0.1621	0.0206	0.1693	0.0177	0.1755	0.0219
		Cross-fitting	0.0521	0.0061	0.0575	0.0189	0.0608	0.0323	0.0643	0.0261	0.0680	0.0294
	600	Proposed	0.0495	0.0206	0.0539	0.0750	0.0522	0.1621	0.0574	0.0771	0.0590	0.0877
		Naive	0.1163	0.0000	0.1376	0.0011	0.1435	0.0040	0.1520	0.0013	0.1572	0.0049
		Cross-fitting	0.0521	0.0061	0.0575	0.0189	0.0608	0.0323	0.0643	0.0261	0.0680	0.0294
	800	Proposed	0.0490	0.0111	0.0509	0.0683	0.0523	0.0742	0.0555	0.0310	0.0575	0.0331
		Naive	0.1054	0.0000	0.1246	0.0000	0.1354	0.0004	0.1411	0.0000	0.1479	0.0003
		Cross-fitting	0.0506	0.0000	0.0553	0.0006	0.0578	0.0008	0.0606	0.0003	0.0640	0.0006

Table 4: The average length of the confidence intervals over $\mathcal{S}$ and $\mathcal{S}^c$ for three methods, averaged over 100 replications, under the dense model.

			SP=0.005		SP=0.01		SP=0.015		SP=0.02		SP=0.025	
p	n	method	$\text{Avglen}_{\mathcal{S}}$	$\text{Avglen}_{\mathcal{S}^c}$	$\text{Avglen}_{\mathcal{S}}$	$\text{Avglen}_{\mathcal{S}^c}$	$\text{Avglen}_{\mathcal{S}}$	$\text{Avglen}_{\mathcal{S}^c}$	$\text{Avglen}_{\mathcal{S}}$	$\text{Avglen}_{\mathcal{S}^c}$	$\text{Avglen}_{\mathcal{S}}$	$\text{Avglen}_{\mathcal{S}^c}$
1600	400	Proposed	1.3266	1.3428	1.4720	1.4889	1.4792	1.4990	1.5747	1.5876	1.6793	1.6945
		Naive	0.4262	0.4653	0.5019	0.5333	0.5073	0.5344	0.5424	0.5693	0.5893	0.6199
		Cross-fitting	0.6051	0.6359	0.6833	0.7146	0.6908	0.7180	0.7365	0.7627	0.7953	0.8234
	600	Proposed	0.7491	0.7748	0.8410	0.8671	1.1714	1.1816	1.2350	1.2483	0.9662	0.9958
		Naive	0.3903	0.4216	0.4560	0.4823	0.4595	0.4838	0.4902	0.5151	0.5322	0.5598
		Cross-fitting	0.5339	0.5519	0.5970	0.6183	0.6027	0.6220	0.6378	0.6588	0.6879	0.7100
	800	Proposed	0.6772	0.6974	0.7562	0.7792	0.9951	1.0073	0.8036	0.8294	0.8677	0.8933
		Naive	0.3643	0.3895	0.4206	0.4443	0.4251	0.4458	0.4523	0.4745	0.4908	0.5151
		Cross-fitting	0.4836	0.4964	0.5374	0.5541	0.5425	0.5578	0.5737	0.5903	0.6175	0.6354
3600	400	Proposed	0.9945	1.0413	1.1077	1.1451	1.8902	1.8959	1.8827	1.8959	1.1315	1.1506
		Naive	0.5124	0.5384	0.5820	0.6031	0.5960	0.6145	0.5947	0.6082	0.5938	0.6029
		Cross-fitting	0.7132	0.7417	0.7926	0.8161	0.8070	0.8297	0.8048	0.8243	0.8067	0.8202
	600	Proposed	0.8760	0.9088	0.9653	0.9968	1.4613	1.4733	0.9838	1.0074	0.9882	1.0044
		Naive	0.4647	0.4895	0.5259	0.5468	0.5381	0.5572	0.5375	0.5518	0.5372	0.5472
		Cross-fitting	0.6258	0.6475	0.6906	0.7100	0.7019	0.7220	0.7018	0.7176	0.7043	0.7157
	800	Proposed	0.7931	0.8225	1.2210	1.2264	1.2397	1.2462	0.8910	0.9098	0.8923	0.9073
		Naive	0.4300	0.4528	0.4844	0.5041	0.4958	0.5138	0.4954	0.5091	0.4950	0.5052
		Cross-fitting	0.5656	0.5842	0.6220	0.6385	0.6324	0.6491	0.6322	0.6460	0.6339	0.6442