REPRODUCIBLE LEARNING IN LARGE-SCALE MULTIPLE GRAPHICAL MODELS

Jia Zhou[1], Guangming Pan[2], Zeming Zheng[3] and Changchun Tan[1]

[1]*Hefei University of Technology,* [2]*Nanyang Technological University and* [3]*University of Science and Technology of China*

Supplementary Material

This Supplementary Material contains four auxiliary lemmas in Section S1 and their proofs in Section S6-S9, as well as the proofs of Lemmas 1-2 in Sections S2-S3 and the proof of Theorems 1-2 in Sections S4-S5. Additionally, two figures about the real data analysis are provided in Section S10. For convenience, in the following, we use C and c to denote some positive constants which can vary from expression to expression.

S1 Auxiliary Lemmas

Lemma S1.1. *For any given node* $a \in (1, \ldots, p)$*, take any subset* $S \subset [-a]$*. Then* $(\mathcal{X}_{-a}, \mathcal{Y}_a) \stackrel{d}{=} \left((\mathcal{X}_{-a})_{\mathrm{swap}(S)}, \mathcal{Y}_a\right)$*. Here,* $(\mathcal{X}_{-a})_{\mathrm{swap}(S)}$ *is obtained from* $\mathcal{X}_{-a}$ *by swapping the* $\mathbf{X}_j^{(t)}$ *and* $\widetilde{\mathbf{X}}_j^{(t)}$ *for all* $1 \le t \le k$ *with* $j \in S$*.*

Lemma S1.2. *For any* $a \in \{1, \ldots, p\}$*, conditional on* $(|W_{a,j}|, j \in [-a])$*, the signs of the null* $W_{a,j}$*,* $j \in [-a] \backslash NE_a$ *are i.i.d. coin flips.*

Lemma S1.3. *For any target FDR level* $q \in (0,1)$, *the estimated edge set* $\widehat{E}$ *obtained by the multiple graphical knockoff filter with the true precision matrices* $\{\mathbf{\Omega}^{(t)}\}_{t=1}^{k}$ *satisfies the exact FDR control, that is* $\mathrm{FDR}(\widehat{E}) = \mathbb{E}\left[(|\widehat{E} \cap E^c|)/(|\widehat{E}| \vee 1)\right] \leq q$.

Lemma S1.4. *Assume that Conditions 1-2 hold, the smallest eigenvalue of* $2diag(\mathbf{s}_a^{(t)}) - diag(\mathbf{s}_a^{(t)})\mathbf{\Omega}_{-a}^{(t)}diag(\mathbf{s}_a^{(t)})$ *is uniformly bounded from below by some positive constant for all* $1 \leq a \leq p$, $1 \leq t \leq k$. *Let* $\widehat{\boldsymbol{\beta}}_a^{aug}(\theta)$ *be the solution of (2.5) with the knockoff matrices constricting based on* θ, $\lambda = \frac{\xi+1}{\xi-1}\left[\frac{k+2\upsilon\log(p)+2\sqrt{\upsilon k\log(p)}}{n(1-\tau)}\right]^{1/2}$, $\tau^2 = 8(\upsilon\log(p)+\log(k))/n = o(1)$, *and* $\upsilon > 2$ *for some constant. Define* $\mathcal{T}_a = \{1,\ldots,2p\}\backslash\{a,(a+p)\}$, $l_a = |NE_a|$, *and* $\mathbf{\Theta} = \left\{\{\mathbf{\Gamma}_{-a}^{(t)}\}_{1\leq a\leq p,1\leq t\leq k} : \max_{1\leq a\leq p,1\leq t\leq k}\|\mathbf{\Gamma}_{-a}^{(t)} - \mathbf{\Omega}_{-a}^{(t)}\|_2 \leq Cb_n\right\}$. *If* $l_m = o(n/\log(p))$, *then*

$$\sup_{\theta\in\mathbf{\Theta}}\sum_{j\in\mathcal{T}_a}\|\widehat{\boldsymbol{\beta}}_{a(j)}^{aug}(\theta) - \boldsymbol{\beta}_{a(j)}^{aug}\|_2 = O(l_a\lambda) \tag{S1.1}$$

holds simultaneously for all $1 \leq a \leq p$ *with probability at least* $1 - p^{-c_\upsilon}$ *for some positive constant* c_υ, *where* $\boldsymbol{\beta}_a^{aug} = \left((\boldsymbol{\beta}_a^{(1)})^\top, (\mathbf{0})^\top, \ldots, (\boldsymbol{\beta}_a^{(k)})^\top, (\mathbf{0})^\top\right)^\top$.

Remark 1. Slightly different from the main body, for convenience, hereafter, we use $\widehat{\boldsymbol{\beta}}_{a(j)}^{\mathrm{aug}}(\theta)$ to uniformly represent the regression coefficients corresponding to the original variables and the knockoff variables. We divide the whole regression coefficient vector $\widehat{\boldsymbol{\beta}}_a^{\mathrm{aug}}(\theta)$ into $2(p-1)$ groups, indexed by $\mathcal{T}_a = \{1, \ldots, 2p\} \backslash \{a, (a+p)\}$. For $1 \le j \le p$, $\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta) = \widehat{\boldsymbol{\beta}}_{a(j)}(\theta)$ and for $p \le j \le 2p$, $\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta) = \widetilde{\boldsymbol{\beta}}_{a(j-p)}(\theta)$, which are defined in (2.6).

S2 Proof of Lemma 1

In view of (2.6), we can see that the statistic $W_{a,j}(\theta)$ is only determined by the estimator $\widehat{\boldsymbol{\beta}}^{\text{aug}}_a(\theta)$. Note that the HGSL solution $\widehat{\boldsymbol{\beta}}^{\text{aug}}_a(\theta)$ is a global minimizer of the objective function (2.5) with the knockoff matrices constricting based on θ, which can be equivalently formulated as

$$\widehat{\boldsymbol{\beta}}^{\text{aug}}_a(\theta) = \arg\min_{\mathbf{b}^{\text{aug}} \in \mathbb{R}^{2(p-1)k}} \left\{ \sum_{t=1}^{k} \sqrt{(\ddot{\mathbf{b}}^{(t)\prime} \mathbf{U}^{(t)}_{-a}(\theta) \ddot{\mathbf{b}}^{(t)} - 2\mathbf{V}^{(t)}_{-a}(\theta) \ddot{\mathbf{b}}^{(t)})} + \lambda \left(\sum_{j \in \mathcal{T}_a} \| \mathcal{D}^{1/2}_{a(j)}(\theta) \mathbf{b}^{\text{aug}}_{(j)} \|_2 \right) \right\},$$

where $\mathcal{T}_a = \{1, \ldots, 2p\} \backslash \{a, (a+p)\}$, $\mathbf{U}^{(t)}_{-a}(\theta) = [\mathbf{X}^{(t)}_{-a}, \widetilde{\mathbf{X}}^{(t)}_{-a}(\theta)]^\top [\mathbf{X}^{(t)}_{-a}, \widetilde{\mathbf{X}}^{(t)}_{-a}(\theta)]/n$, and $\mathbf{V}^{(t)}_{-a}(\theta) = [\mathbf{X}^{(t)}_{-a}, \widetilde{\mathbf{X}}^{(t)}_{-a}(\theta)]^\top \mathbf{X}^{(t)}_a / n$. For simplicity, we use notation $\ddot{\mathbf{b}}^{(t)}$ denotes $((\mathbf{b}^{(t)})^\top, (\widetilde{\mathbf{b}}^{(t)})^\top)^\top$ defined in (2.5). For $j \le p$, $\mathcal{D}_{a(j)}(\theta) = \mathbf{D}_{a(j)}$ defined in equation (2.5), and for $j > p$, $\mathcal{D}_{a(j)}(\theta) = \widetilde{\mathbf{D}}_{a(l)}(\theta)$ with $l = j - p$ defined in equation (2.5). Note that the tth entry of $\mathcal{D}_{a(j)}(\theta)$ is equal to $\frac{n}{n^{(t)}} \mathbf{U}^{(t)}_j(\theta)$ for all $j \in \mathcal{T}_a$, $1 \le t \le k$. Thus, $\mathcal{D}_{a(j)}(\theta)$ can be completely represented by $\mathbf{U}_{-a}(\theta) = (\mathbf{U}^{(1)}_{-a}(\theta), \ldots, \mathbf{U}^{(k)}_{-a}(\theta))$ since $n/n^{(t)}$ is a fixed positive constant.

Obviously, we can see that $\widehat{\boldsymbol{\beta}}^{\text{aug}}_a(\theta)$ depends on the data $(\mathcal{X}_{-a}(\theta), \mathcal{Y}_a)$ only through $\mathbf{U}_{-a}(\theta)$ and $\mathbf{V}_{-a}(\theta) = \left(\mathbf{V}^{(1)}_{-a}(\theta), \ldots, \mathbf{V}^{(k)}_{-a}(\theta)\right)$. Recall that

$\mathbf{H}_a(\theta) = (\mathbf{U}_{-a}(\theta), \mathbf{V}_{-a}(\theta))$. From the above analysis, we can conclude that for the statistic vector $\mathbf{W}_a(\theta)$, it only depends on $\mathbf{H}_a(\theta)$. Considering the whole statistics $\{\mathbf{W}_a(\theta), 1 \leq a \leq p\}$, since the zeros has no influence in the knockoff procedure, we can write the statistics matrix as

$$\begin{pmatrix} \mathbf{W}_1(\theta) & & & \\ & \mathbf{W}_2(\theta) & & \\ & & \ddots & \\ & & & \mathbf{W}_p(\theta) \end{pmatrix} = g(\mathbf{H}(\theta)).$$

Further, note that the threshold vector $\mathbf{T}(\theta)$ is completely determined by the whole statistics $\{\mathbf{W}_a(\theta), 1 \leq a \leq p\}$. By the process of obtaining the estimated edge set, it's obvious to see that the estimated $\widehat{E}(\theta)$ identified by our procedure with the given sequence of precision matrices θ depends only on $\mathbf{H}(\theta)$. Thus, we have completed the proof of Lemma 1.

S3 Proof of Lemma 2

Recall that $\mathcal{A}_a^*(\theta)$ is the support of knockoff statistics $\mathbf{W}_a(\theta)$. Thus, the threshold vector $\left(\widehat{T}_1(\theta), \ldots, \widehat{T}_p(\theta)\right)$ depends only on $\{W_{a,j}(\theta), j \in \mathcal{A}_a\}_{a=1}^p$ as $\mathcal{A}_a^*(\theta) \subset \mathcal{A}_a$. By equation of (2.6), it shows that

$$W_{a,j}(\theta) = \|\widehat{\boldsymbol{\beta}}_{a(j)}(\theta)\|_2 - \|\widetilde{\boldsymbol{\beta}}_{a(j)}(\theta)\|_2,$$

with

$$\widehat{\boldsymbol{\beta}}_a^{\text{aug}}(\theta) = \arg \min_{\mathbf{c} \in \mathbb{R}^{2(p-1)k}} \left\{ \sum_{t=1}^{k} Q_{t\theta}(\mathbf{c}^{(t)}) + \lambda \left(\sum_{j \in \mathcal{T}_a} \| \mathcal{D}_{a(j)}^{1/2}(\theta) \mathbf{c}_{(j)} \|_2 \right) \right\} \quad \text{(S3.1)}$$

where

$$Q_{\theta t}(\mathbf{c}^{(t)}) = \frac{\| \mathbf{X}_a^{(t)} - [\mathbf{X}_{-a}^{(t)}, \widetilde{\mathbf{X}}_{-a}^{(t)}(\theta)] \mathbf{c}^{(t)} \|_2}{\sqrt{n}}.$$

Here, we use the notation $\mathbf{c}^{(t)}$ denotes the subvector corresponding to the tth class and $\mathbf{c}_{(l)}$ represents its subvector corresponding to the lth group. The notation of $\mathcal{D}_{a(j)}^{1/2}(\theta)$ is the same as that in the proof of Lemma 1.

When we restrict ourselves to a sequence of subset $\{\mathcal{A}_a(\theta)\}_{a=1}^{p}$ that satisfies $\mathcal{A}_a \supset \mathcal{A}_a^*(\theta)$ for all $1 \leq a \leq p$, let $\widehat{\boldsymbol{\beta}}_{a(\mathcal{J}_a)}^{\text{aug}}(\theta)$ be the subvector of $\widehat{\boldsymbol{\beta}}_a^{\text{aug}}(\theta)$ consists of components indexed by $\mathcal{J}_a$ with $\mathcal{J}_a = \mathcal{A}_a \cup \{\mathcal{A}_a + p\}$. To prove this lemma it's sufficient to prove that $\widehat{\boldsymbol{\beta}}_{a(\mathcal{J}_a)}^{\text{aug}}(\theta) = \widehat{\boldsymbol{\zeta}}_a^{\text{aug}}(\theta)$ since $\widehat{\boldsymbol{\beta}}_{a(\mathcal{J}_a^c)}^{\text{aug}}(\theta) = \mathbf{0}$ that doesn't work for the knockoff procedure, where

$$\widehat{\boldsymbol{\zeta}}_a^{\text{aug}}(\theta) = \arg \min_{\mathbf{b} \in \mathbb{R}^{2|\mathcal{A}_a|k}} \left\{ \sum_{t=1}^{k} Q_{\theta t}^{\mathcal{A}_a}(\mathbf{b}^{(t)}) + \lambda \left(\sum_{j \in \mathcal{J}_a} \| \mathcal{D}_{a(j)}^{1/2}(\theta) \mathbf{b}_{(j)} \|_2 \right) \right\},$$

with the same λ and θ as in (S3.1). Here

$$Q_{\theta t}^{\mathcal{A}_a}(\mathbf{b}^{(t)}) = \frac{\| \mathbf{X}_a^{(t)} - [\mathbf{X}_{\mathcal{A}_a}, \widetilde{\mathbf{X}}_{\mathcal{A}_a}(\theta)] \mathbf{b}^{(t)} \|_2}{\sqrt{n}}.$$

We can see that by equation (S3.1) it shows that for any $\mathbf{b} \in \mathbb{R}^{2(p-1)k}$

and $\mathbf{b} \neq \widehat{\boldsymbol{\beta}}_a^{\text{aug}}(\theta)$,

$$\sum_{t=1}^{k} \frac{\|\mathbf{X}_a^{(t)} - [\mathbf{X}_{-a}^{(t)}, \widetilde{\mathbf{X}}_{-a}^{(t)}(\theta)][(\widehat{\boldsymbol{\beta}}_a^{(t)}(\theta))^\top, (\widetilde{\boldsymbol{\beta}}_a^{(t)}(\theta))^\top]^\top\|_2}{\sqrt{n}} + \lambda \left(\sum_{j \in \mathcal{T}_a} \|\mathcal{D}_{a(j)}^{1/2}(\theta) \widehat{\boldsymbol{\beta}}_{a(j)}^{\text{aug}}(\theta)\|_2 \right)$$
$$< \sum_{t=1}^{k} \frac{\|\mathbf{X}_a^{(t)} - [\mathbf{X}_{-a}^{(t)}, \widetilde{\mathbf{X}}_{-a}^{(t)}(\theta)]\mathbf{b}^{(t)}\|_2}{\sqrt{n}} + \lambda \left(\sum_{j \in \mathcal{T}_a} \|\mathcal{D}_{a(j)}^{1/2}(\theta) \mathbf{b}_{(j)}\|_2 \right). \quad \text{(S3.2)}$$

Then we claim that for any $\mathbf{c} \in \mathbb{R}^{|\mathcal{J}_a|k}$,

$$\sum_{t=1}^{k} \frac{\|\mathbf{X}_a^{(t)} - [\mathbf{X}_{\mathcal{A}_a}^{(t)}, \widetilde{\mathbf{X}}_{\mathcal{A}_a}^{(t)}(\theta)][(\widehat{\boldsymbol{\beta}}_{a,\mathcal{A}_a}^{(t)}(\theta))^\top, (\widetilde{\boldsymbol{\beta}}_{a,\mathcal{A}_a}^{(t)}(\theta))^\top]^\top\|_2}{\sqrt{n}} + \lambda \left(\sum_{j \in \mathcal{J}_a} \|\mathcal{D}_{a(j)}^{1/2}(\theta) \widehat{\boldsymbol{\beta}}_{a(j)}^{\text{aug}}(\theta)\|_2 \right)$$
$$< \sum_{t=1}^{k} \frac{\|\mathbf{X}_a^{(t)} - [\mathbf{X}_{\mathcal{A}_a}^{(t)}, \widetilde{\mathbf{X}}_{\mathcal{A}_a}^{(t)}(\theta)]\mathbf{c}^{(t)}\|_2}{\sqrt{n}} + \lambda \left(\sum_{j \in \mathcal{J}_a} \|\mathcal{D}_{a(j)}^{1/2}(\theta) \mathbf{c}_{(k)}\|_2 \right),$$

where k is the order of j in $\mathcal{J}_a$. Otherwise, if there exists $\widehat{\mathbf{C}}_{a(\mathcal{J}_a)} \in \mathbb{R}^{|\mathcal{J}_a|k}$

$$\sum_{t=1}^{k} \frac{\|\mathbf{X}_a^{(t)} - [\mathbf{X}_{-a}^{(t)}, \widetilde{\mathbf{X}}_{-a}^{(t)}(\theta)]\widehat{\mathbf{C}}_{a(\mathcal{J}_a)}^{(t)}(\theta)\|_2}{\sqrt{n}} + \lambda \left(\sum_{j \in \mathcal{J}_a} \|\mathcal{D}_{a(j)}^{1/2}(\theta) \widehat{\mathbf{C}}_{a(j)}(\theta)\|_2 \right)$$
$$< \sum_{t=1}^{k} \frac{\|\mathbf{X}_a^{(t)} - [\mathbf{X}_{-a}^{(t)}, \widetilde{\mathbf{X}}_{-a}^{(t)}(\theta)][(\widehat{\boldsymbol{\beta}}_{a,\mathcal{A}_a}^{(t)}(\theta))^\top, (\widetilde{\boldsymbol{\beta}}_{a,\mathcal{A}_a}^{(t)}(\theta))^\top]^\top\|_2}{\sqrt{n}} + \lambda \left(\sum_{j \in \mathcal{J}_a} \|\mathcal{D}_{a(j)}^{1/2}(\theta) \widehat{\boldsymbol{\beta}}_{a(j)}^{\text{aug}}(\theta)\|_2 \right)$$

Then, let $\widehat{\mathbf{C}}_{a(\mathcal{T}_a \setminus \mathcal{J}_a)} = \mathbf{0}$ it holds that

$$\sum_{t=1}^{k} \frac{\|\mathbf{X}_a^{(t)} - [\mathbf{X}_{-a}^{(t)}, \widetilde{\mathbf{X}}_{-a}^{(t)}(\theta)]\widehat{\mathbf{C}}_a^{(t)}(\theta)\|_2}{\sqrt{n}} + \lambda \left(\sum_{j \in \mathcal{T}_a} \|\mathcal{D}_{a(j)}^{1/2}(\theta) \widehat{\mathbf{C}}_{a(j)}(\theta)\|_2 \right)$$
$$< \sum_{t=1}^{k} \frac{\|\mathbf{X}_a^{(t)} - [\mathbf{X}_{-a}^{(t)}, \widetilde{\mathbf{X}}_{-a}^{(t)}(\theta)][(\widehat{\boldsymbol{\beta}}_a^{(t)}(\theta))^\top, (\widetilde{\boldsymbol{\beta}}_a^{(t)}(\theta))^\top]^\top\|_2}{\sqrt{n}} + \lambda \left(\sum_{j \in \mathcal{T}_a} \|\mathcal{D}_{a(j)}^{1/2}(\theta) \widehat{\boldsymbol{\beta}}_{a(j)}^{\text{aug}}(\theta)\|_2 \right),$$

which contradicts equation (S3.2). Thus, we can conclude that $(\widehat{\boldsymbol{\beta}}_{a(\mathcal{J}_a)}^{\text{aug}}(\theta)) = \widehat{\boldsymbol{\zeta}}_a^{\text{aug}}(\theta)$. This guarantees that $E_{\mathcal{F}}(\mathbf{H}(\theta))$ and $E_{\mathcal{A}}(\mathbf{H}_{\mathcal{A}}(\theta))$ are identical and thus concludes the proof of Lemma 2.

S4 Proof of Theorem 1

Let $\mathcal{A}_a^*(\widehat{\theta})$ be the support of knockoff statistic vector $\mathbf{W}_a(\widehat{\theta})$. Define set $\widehat{\mathcal{A}}_a(\widehat{\theta}) := \mathcal{A}_a^*(\widehat{\theta}) \cup \mathcal{A}_a^*(\theta_0)$. It follows from Condition 4 that the cardinality of $\widehat{\mathcal{A}}_a(\widehat{\theta})$ is bounded by d. Hereafter we write $\widehat{\mathcal{A}}_a(\widehat{\theta})$ as $\widehat{\mathcal{A}}_a$ for notational simplicity. Additionally, we use $\widehat{\mathcal{A}}$ to denote the set of $\{\widehat{\mathcal{A}}_a\}_{a=1}^p$.

Recall the definitions of $\mathbb{I}_{\mathcal{A}}$ as in (3.14). Define the event

$$\mathcal{E}_{np} = \left\{\mathbf{H}_{\widehat{\mathcal{A}}}(\widehat{\theta}) \in \mathbb{I}_{\widehat{\mathcal{A}}}\right\} \cap \left\{\mathbf{H}_{\widehat{\mathcal{A}}}(\theta_0) \in \mathbb{I}_{\widehat{\mathcal{A}}}\right\}.$$

It's obvious that $\mathbf{H}_{\widehat{\mathcal{A}}}(\theta_0) \in \mathbb{I}_{\widehat{\mathcal{A}}}$. We only need to prove that with the probability at least $1 - p^{-c_\delta}$

$$\max_{1\le a\le p} \|\mathbf{H}_{\widehat{\mathcal{A}}_a}(\widehat{\theta}) - \mathbf{H}_{\widehat{\mathcal{A}}_a}(\theta_0)\|_2 = o(1), \tag{S4.1}$$

which entails that $\mathbb{P}(\mathcal{E}_{np}^c) \le p^{-c_\delta}$. Then, according to Lemma S1.3, we have $\mathrm{FDR}(\mathbf{H}_{\widehat{\mathcal{A}}}(\theta_0)) = \mathrm{FDR}(\mathbf{H}) \le q$. Combining with the Condition 5, the rest of the proof is similar to the proof of theorem 1 in Fan et al. (2020) with $\pi_{np} = p^{-c_\delta}$, so we have omitted it.

Now we will focus on the proof of equation (S4.1). For convenience, in the following, we will abbreviate $\mathbf{H}_{\widehat{\mathcal{A}}_a}(\theta_0)$ to $\mathbf{H}_{\widehat{\mathcal{A}}_a}$, $\mathbf{U}^{(t)}_{\widehat{\mathcal{A}}_a}(\theta_0)$ to $\mathbf{U}^{(t)}_{\widehat{\mathcal{A}}_a}$ and $\mathbf{V}^{(t)}_{\widehat{\mathcal{A}}_a}(\theta_0)$ to $\mathbf{V}^{(t)}_{\widehat{\mathcal{A}}_a}$. Similarly, we abbreviate $\mathbf{H}_{\widehat{\mathcal{A}}_a}(\widehat{\theta})$ to $\widehat{\mathbf{H}}_{\widehat{\mathcal{A}}_a}$, $\mathbf{U}^{(t)}_{\widehat{\mathcal{A}}_a}(\widehat{\theta})$ to $\widehat{\mathbf{U}}^{(t)}_{\widehat{\mathcal{A}}_a}$

and $\mathbf{V}^{(t)}_{\widehat{\mathcal{A}}_a}(\widehat{\theta})$ to $\widehat{\mathbf{V}}^{(t)}_{\widehat{\mathcal{A}}_a}$. By the definition of $\mathbf{H}_{\widehat{\mathcal{A}}_a}(\widehat{\theta})$, we have

$$\begin{aligned}
&\max_{1\le a\le p}\|\widehat{\mathbf{H}}_{\widehat{\mathcal{A}}_a}-\mathbf{H}_{\widehat{\mathcal{A}}_a}\|_2\\
=&\max_{1\le a\le p}\left\|\left(\widehat{\mathbf{U}}^{(1)}_{\widehat{\mathcal{A}}_a},\dots,\widehat{\mathbf{U}}^{(k)}_{\widehat{\mathcal{A}}_a},\widehat{\mathbf{V}}^{(1)}_{\widehat{\mathcal{A}}_a},\dots,\widehat{\mathbf{V}}^{(k)}_{\widehat{\mathcal{A}}_a}\right)-\left(\mathbf{U}^{(1)}_{\widehat{\mathcal{A}}_a},\dots,\mathbf{U}^{(k)}_{\widehat{\mathcal{A}}_a},\mathbf{V}^{(1)}_{\widehat{\mathcal{A}}_a},\dots,\mathbf{V}^{(k)}_{\widehat{\mathcal{A}}_a}\right)\right\|_2\\
\le&\max_{1\le a\le p}\left(\sum_{t=1}^{k}\|\widehat{\mathbf{U}}^{(t)}_{\widehat{\mathcal{A}}_a}-\mathbf{U}^{(t)}_{\widehat{\mathcal{A}}_a}\|_2+\sum_{t=1}^{k}\|\widehat{\mathbf{V}}^{(t)}_{\widehat{\mathcal{A}}_a}-\mathbf{V}^{(t)}_{\widehat{\mathcal{A}}_a}\|_2\right)\\
\le& k\left(\max_{1\le a\le p,1\le t\le k}\|\widehat{\mathbf{U}}^{(t)}_{\widehat{\mathcal{A}}_a}-\mathbf{U}^{(t)}_{\widehat{\mathcal{A}}_a}\|_2+\max_{1\le a\le p,1\le t\le k}\|\widehat{\mathbf{V}}^{(t)}_{\widehat{\mathcal{A}}_a}-\mathbf{V}^{(t)}_{\widehat{\mathcal{A}}_a}\|_2\right) \qquad \text{(S4.2)}
\end{aligned}$$

Thus, we can see that it is sufficient to deduce the bounds of $\max_{1\le a\le p,1\le t\le k}\|\widehat{\mathbf{U}}^{(t)}_{\widehat{\mathcal{A}}_a}-\mathbf{U}^{(t)}_{\widehat{\mathcal{A}}_a}\|_2$ and $\max_{1\le a\le p,1\le t\le k}\|\widehat{\mathbf{V}}^{(t)}_{\widehat{\mathcal{A}}_a}-\mathbf{V}^{(t)}_{\widehat{\mathcal{A}}_a}\|_2$.

The error bounds of $\max_{1\le a\le p,1\le t\le k}\|\widehat{\mathbf{U}}^{(t)}_{\widehat{\mathcal{A}}_a}-\mathbf{U}^{(t)}_{\widehat{\mathcal{A}}_a}\|_2$. Recalling the definition of $\mathbf{U}^{(t)}_{\widehat{\mathcal{A}}_a}$, we have

$$\begin{aligned}
&\left\|\widehat{\mathbf{U}}^{(t)}_{\widehat{\mathcal{A}}_a}-\mathbf{U}^{(t)}_{\widehat{\mathcal{A}}_a}\right\|_2=\left\|[\mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a},\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a}]^\top[\mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a},\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a}]/n-[\mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a},\widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a}]^\top[\mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a},\widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a}]/n\right\|_2\\
=&\left\|\begin{pmatrix}\mathbf{0} & (\mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a})^\top(\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a}-\widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a})/n\\ (\mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a})^\top(\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a}-\widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a})/n & [(\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a})^\top\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a}-(\widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a})^\top\widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a}]/n\end{pmatrix}\right\|_2\\
\le&2\left\|(\mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a})^\top(\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a}-\widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a})/n\right\|_2+\left\|[(\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a})^\top\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a}-(\widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a})^\top\widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a}]/n\right\|_2\\
=:&2A_{at}+B_{at} \qquad \text{(S4.3)}
\end{aligned}$$

where

$$A_{at}=\|(\mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a})^\top(\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a}-\widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a})/n\|_2, \text{ and } B_{at}=\left\|\left[(\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a})^\top\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a}-(\widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a})^\top\widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a}\right]/n\right\|_2.$$

Here, the notation $\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a}$ is the abbreviation of $\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a}(\widehat{\theta})$.

Recalling the knockoff generation process (2.5), for the original data matrix $\mathbf{X}_{-a}^{(t)}$, the ideal knockoff matrix and the approximate knockoff matrix constructed using the estimated precision matrices can be represented as

$$\widetilde{\mathbf{X}}_{-a}^{(t)} = \mathbf{X}_{-a}^{(t)}\mathbf{C}_a^{(t)} + \mathbf{Z}_a^{(t)}\mathbf{B}_a^{(t)} \quad \text{and} \quad \widehat{\widetilde{\mathbf{X}}}_{-a}^{(t)} = \mathbf{X}_{-a}^{(t)}\widehat{\mathbf{C}}_a^{(t)} + \mathbf{Z}_a^{(t)}\widehat{\mathbf{B}}_a^{(t)}.$$

respectively, where $\mathbf{Z}_a^{(t)}$ is a random matrix whose rows are i.i.d copies of $N(\mathbf{0}, \mathbf{I_{p-1}})$ and independent of $\mathbf{X}_{-a}^{(t)}$, $\mathbf{C}_a^{(t)} = \mathbf{I}_{p-1} - \text{diag}\{\mathbf{s}_a^{(t)}\}\mathbf{\Omega}_{-a}^{(t)}$ and $\widehat{\mathbf{C}}_a^{(t)} = \mathbf{I}_{p-1} - \text{diag}\{\mathbf{s}_a^{(t)}\}\widehat{\mathbf{\Omega}}_{-a}^{(t)}$, $\mathbf{B}_a^{(t)} = \Big(2\text{diag}\{\mathbf{s}_a^{(t)}\} - \text{diag}\{\mathbf{s}_a^{(t)}\}\mathbf{\Omega}_{-a}^{(t)}\text{diag}\{\mathbf{s}_a^{(t)}\}\Big)^{1/2}$, and $\widehat{\mathbf{B}}_a^{(t)} = \Big(2\text{diag}\{\mathbf{s}_a^{(t)}\} - \text{diag}\{\mathbf{s}_a^{(t)}\}\widehat{\mathbf{\Omega}}_{-a}^{(t)}\text{diag}\{\mathbf{s}_a^{(t)}\}\Big)^{1/2}$.

For convenience, denote by

$$G_1 = \max_{1\le a\le p, 1\le t\le k} \|\frac{1}{n}(\mathbf{X}_{\widehat{\mathcal{A}}_a}^{(t)})^\top \mathbf{X}_{\widehat{\mathcal{A}}_a}^{(t)}\|_2, \qquad G_2 = \max_{1\le a\le p, 1\le t\le k} \|\widehat{\mathbf{C}}_a^{(t)} - \mathbf{C}_a^{(t)}\|_2,$$

$$G_3 = \max_{1\le a\le p, 1\le t\le k} \|\frac{1}{n}(\mathbf{X}_{\widehat{\mathcal{A}}_a}^{(t)})^\top \mathbf{Z}_{\widehat{\mathcal{A}}_a}^{(t)}\|_2, \qquad G_4 = \max_{1\le a\le p, 1\le t\le k} \|\widehat{\mathbf{B}}_a^{(t)} - \mathbf{B}_a^{(t)}\|_2,$$

and $G_5 = \max_{1\le a\le p, 1\le t\le k} \|\frac{1}{n}(\mathbf{Z}_{\widehat{\mathcal{A}}_a}^{(t)})^\top \mathbf{Z}_{\widehat{\mathcal{A}}_a}^{(t)}\|_2$. Here $\mathbf{Z}_{\widehat{\mathcal{A}}_a}^{(t)}$ is the submatrix of $\mathbf{Z}_a^{(t)}$ with the columns in $\widehat{\mathcal{A}}_a$. By a more elaborate derivation, we have

$$\max_{1\le a\le p, 1\le t\le k} \left\| \frac{1}{n}(\mathbf{X}_{\widehat{\mathcal{A}}_a}^{(t)})^\top (\widehat{\widetilde{\mathbf{X}}}_{\widehat{\mathcal{A}}_a}^{(t)} - \widetilde{\mathbf{X}}_{\widehat{\mathcal{A}}_a}^{(t)}) \right\|_2 \le G_1G_2 + G_3G_4,$$

$$\max_{1\le a\le p, 1\le t\le k} \frac{1}{n} \left\| (\widehat{\widetilde{\mathbf{X}}}_{\widehat{\mathcal{A}}_a}^{(t)})^\top \widehat{\widetilde{\mathbf{X}}}_{\widehat{\mathcal{A}}_a}^{(t)} - (\widetilde{\mathbf{X}}_{\widehat{\mathcal{A}}_a}^{(t)})^\top \widetilde{\mathbf{X}}_{\widehat{\mathcal{A}}_a}^{(t)} \right\|_2$$
$$\le G_1G_2^2 + 2G_3G_2G_4 + G_5G_4^2 + 2G_1G_2 + 2G_3G_4,$$

Combining with formula (S4.3), we can conclude that

$$\max_{1\le a\le p, 1\le t\le k} \left\| \widehat{\mathbf{U}}_{\widehat{\mathcal{A}}_a}^{(t)} - \mathbf{U}_{\widehat{\mathcal{A}}_a}^{(t)} \right\|_2 \le 4G_1G_2 + 4G_3G_4 + G_1G_2^2 + G_5G_4^2 + 2G_3G_2G_4.$$

Thus, in the following, we will deduce the upper bounds of G_1, G_2, G_3, G_4 and G_5 respectively.

The upper bounds of G_1, G_3, G_5.

It follows from Cauchy's interlace theorem and Condition 1 that

$$1/M_1 \leq \lambda_{\min}(\mathbf{\Sigma}^{(t)}) \leq \lambda_{\min}(\mathbf{\Sigma}^{(t)}_{\widehat{\mathcal{A}}_a,\widehat{\mathcal{A}}_a}) \leq \lambda_{\max}(\mathbf{\Sigma}^{(t)}_{\widehat{\mathcal{A}}_a,\widehat{\mathcal{A}}_a}) \leq \lambda_{\max}(\mathbf{\Sigma}^{(t)}) \leq M_1$$

holds uniformly for any $1 \leq a \leq p$, $1 \leq t \leq k$. By Remark 5.40 in Vershynin (2010), it yields that with probability at least $1 - 2pk\exp(-cn)$

$$\begin{aligned} G_1 &= \max_{1\leq a\leq p, 1\leq t\leq k} \frac{n^{(t)}}{n} \left\| \frac{1}{n^{(t)}} (\mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a})^\top \mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a} \right\|_2 \\ &\leq \frac{n^{(t)}}{n} \max_{1\leq a\leq p, 1\leq t\leq k} \lambda_{\max}(\mathbf{\Sigma}^{(t)}_{\widehat{\mathcal{A}}_a,\widehat{\mathcal{A}}_a}) + \frac{n^{(t)}}{n} \max\{\eta_1, \eta_1^2\} \leq C \end{aligned} \tag{S4.4}$$

where $\eta_1 = C_1\sqrt{d/n^{(t)}} + 1$ with C_1 is a positive constant. Under Condition 4 and Condition 2 that $d < n$ and $n \asymp n^{(t)}$, we have $\eta_1 = O(1)$, which entails the last inequality.

Correspondingly, with probability at least $1 - 2pk\exp(-cn)$

$$G_3 = \max_{1\leq a\leq p, 1\leq t\leq k} \frac{n^{(t)}}{n} \left\| \frac{1}{n^{(t)}} (\mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a})^\top \mathbf{Z}^{(t)}_{\widehat{\mathcal{A}}_a} \right\|_2 \leq \frac{n^{(t)}}{n} \max\{\eta_1, \eta_1^2\} \leq C \tag{S4.5}$$

since $\mathbb{E}[\frac{1}{n^{(t)}} (\mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a})^\top \mathbf{Z}^{(t)}_{\widehat{\mathcal{A}}_a}] = 0$. Similarly, with probability at least $1-2pk\exp(-cn)$

$$G_5 = \max_{1\leq a\leq p, 1\leq t\leq k} \frac{n^{(t)}}{n} \left\| \frac{1}{n^{(t)}} (\mathbf{Z}^{(t)}_{\widehat{\mathcal{A}}_a})^\top \mathbf{Z}^{(t)}_{\widehat{\mathcal{A}}_a} \right\|_2 \leq \frac{n^{(t)}}{n} \left(1 + \max\{\eta_1, \eta_1^2\}\right) \leq C, \tag{S4.6}$$

since $\mathbb{E}[\frac{1}{n^{(t)}} (\mathbf{Z}^{(t)})^\top \mathbf{Z}^{(t)}] = \mathbf{I}$.

The upper bounds of G_2 and G_4.

Since the vector $\mathbf{s}_a^{(t)}$ is chosen such that $\boldsymbol{\Sigma}_{-a,-a} - 2^{-1}\mathrm{diag}\{\mathbf{s}_a^{(t)}\}$ is positive semidefinite, we have

$$\max_{1\le a\le p, 1\le t\le k} \|\mathbf{s}_a^{(t)}\|_\infty \le 2 \max_{1\le a\le p, 1\le t\le k} \lambda_{\max}(\boldsymbol{\Sigma}_{-a,-a}^{(t)}) \le 2\lambda_{\max}(\boldsymbol{\Sigma}^{(t)}) \le 2M_1$$

where the second inequality follows form Cauchy's interlace theorem and the last inequality is due to Condition 1. Therefore, under Conditions 1 and 3, we have with probability at least $1 - p^{-\delta}$

$$\begin{aligned} G_2 &= \max_{1\le a\le p, 1\le t\le k} \|(\widehat{\mathbf{C}}_a^{(t)} - \mathbf{C}_a^{(t)})\|_2 \\ &\le \max_{1\le a\le p, 1\le t\le k} \|\mathbf{s}_a^{(t)}\|_\infty \max_{1\le a\le p, 1\le t\le k} \|\widehat{\boldsymbol{\Omega}}_{-a}^{(t)} - \boldsymbol{\Omega}_{-a}^{(t)}\|_2 \le Cb_n \end{aligned} \tag{S4.7}$$

It follows from the definitions of $\mathbf{B}_a^{(t)}$ and $\widehat{\mathbf{B}}_a^{(t)}$ that $(\mathbf{B}_a^{(t)})^2 - (\widehat{\mathbf{B}}_a^{(t)})^2 = \mathrm{diag}\{\mathbf{s}_a^{(t)}\}(\boldsymbol{\Omega}_{-a}^{(t)} - \widehat{\boldsymbol{\Omega}}_{-a}^{(t)})\mathrm{diag}\{\mathbf{s}_a^{(t)}\}$ and thus by Condition 3 it holds that with probability at least $1 - p^{-\delta}$

$$\begin{aligned} &\max_{1\le a\le p, 1\le t\le k} \|(\widehat{\mathbf{B}}_a^{(t)})^2 - (\mathbf{B}_a^{(t)})^2\|_2 \\ &\le \max_{1\le a\le p, 1\le t\le k} \|\mathbf{s}_a^{(t)}\|_\infty^2 \cdot \max_{1\le a\le p, 1\le t\le k} \|\boldsymbol{\Omega}_{-a}^{(t)} - \widehat{\boldsymbol{\Omega}}_{-a}^{(t)}\|_2 \le Cb_n. \end{aligned} \tag{S4.8}$$

This, together with the fact that $(\widehat{\mathbf{B}}_a^{(t)})^2 = (\mathbf{B}_a^{(t)})^2 + [(\widehat{\mathbf{B}}_a^{(t)})^2 - (\mathbf{B}_a^{(t)})^2]$, yields that

$$\lambda_{\min}\{(\widehat{\mathbf{B}}_a^{(t)})^2\} \geq \lambda_{\min}\{(\mathbf{B}_a^{(t)})^2\} + \lambda_{\min}\{(\widehat{\mathbf{B}}_a^{(t)})^2 - (\mathbf{B}_a^{(t)})^2\}$$
$$\geq \lambda_{\min}\{(\mathbf{B}_a^{(t)})^2\} - \lambda_{\max}\{(\mathbf{B}_a^{(t)})^2 - (\widehat{\mathbf{B}}_a^{(t)})^2\} \geq \lambda_{\min}\{(\mathbf{B}_a^{(t)})^2\} - Cb_n \tag{S4.9}$$

holds uniformly for all $1 \leq a \leq p$, $1 \leq t \leq k$.

Recall that $(\mathbf{B}_a^{(t)})^2 = 2\text{diag}\{\mathbf{s}_a^{(t)}\} - \text{diag}\{\mathbf{s}_a^{(t)}\}\boldsymbol{\Omega}_{-a}^{(t)}\text{diag}\{\mathbf{s}_a^{(t)}\}$. The assumption that $\lambda_{\min}(2\text{diag}\{\mathbf{s}_{\mathrm{a}}^{(\mathrm{t})}\} - \text{diag}\{\mathbf{s}_{\mathrm{a}}^{(\mathrm{t})}\}\boldsymbol{\Omega}_{-\mathrm{a}}^{(\mathrm{t})}\text{diag}\{\mathbf{s}_{\mathrm{a}}^{(\mathrm{t})}\}) \geq \mathrm{C}_3$ uniformly for all $1 \leq a \leq p$, $1 \leq t \leq k$ with some constant $C_3 > 0$ implies that

$$\lambda_{\min}\{(\mathbf{B}_a^{(t)})^2\} \geq C_3$$

holds uniformly over $1 \leq a \leq p$, $1 \leq t \leq k$. This, together with (S4.9) and the assumption that $b_n \to 0$ as $n \to \infty$, yields that $\lambda_{\min}\{(\widehat{\mathbf{B}}_a^{(t)})^2\} \geq C_3/2$ holds uniformly over $1 \leq a \leq p$, $1 \leq t \leq k$ for all n large enough. Thus, it follows from Lemma 2.2 in Schmitt (1992) that with probability at least $1 - p^{-\delta}$

$$G_4 = \max_{1\leq a\leq p, 1\leq t\leq k} \|\widehat{\mathbf{B}}_a^{(t)} - \mathbf{B}_a^{(t)}\|_2 \leq C \max_{1\leq a\leq p, 1\leq t\leq k} \|(\widehat{\mathbf{B}}_a^{(t)})^2 - (\mathbf{B}_a^{(t)})^2\|_2 \leq C_4 b_n. \tag{S4.10}$$

The last inequality is given by (S4.8). Therefore, in view of these results of (S4.4),(S4.5),(S4.6),(S4.7), and (S4.10), we can conclude that with probability at least $1 - 6pk\exp(-cn) - 2p^{-\delta}$

$$\max_{1\le a\le p, 1\le t\le k} \left\| \widehat{\mathbf{U}}^{(t)}_{\widehat{\mathcal{A}}_a} - \mathbf{U}^{(t)}_{\widehat{\mathcal{A}}_a} \right\|_2 \le Cb_n \qquad \text{(S4.11)}$$

The error bounds of $\max_{1\le a\le p,1\le t\le k} \|\widehat{\mathbf{V}}^{(t)}_{\widehat{\mathcal{A}}_a} - \mathbf{V}^{(t)}_{\widehat{\mathcal{A}}_a}\|_2$**.** Recalling the definition of $\mathbf{V}_{\widehat{\mathcal{A}}_a}$, we have that

$$\max_{1\le a\le p,1\le t\le k} \|\widehat{\mathbf{V}}^{(t)}_{\widehat{\mathcal{A}}_a} - \mathbf{V}^{(t)}_{\widehat{\mathcal{A}}_a}\|_2 = \max_{1\le a\le p,1\le t\le k} \left\| \frac{1}{n}[\mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a}, \widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a}]^\top \mathbf{X}^{(t)}_a - \frac{1}{n}[\mathbf{X}^{(t)}_{\widehat{\mathcal{A}}_a}, \widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a}]^\top \mathbf{X}^{(t)}_a \right\|_2$$
$$= \max_{1\le a\le p,1\le t\le k} \left\| \frac{1}{n}(\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a} - \widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a})^\top \mathbf{X}^{(t)}_a \right\|_2 \le \max_{1\le a\le p,1\le t\le k} \frac{1}{n} \left\| (\widehat{\widetilde{\mathbf{X}}}^{(t)}_{\widehat{\mathcal{A}}_a} - \widetilde{\mathbf{X}}^{(t)}_{\widehat{\mathcal{A}}_a}) \right\|_2 \left\| \mathbf{X}^{(t)}_a \right\|_2$$
$$\le \frac{1}{\sqrt{n}} \max_{1\le a\le p,1\le t\le k} \|\mathbf{X}^{(t)}_a\|_2 \sqrt{G_1} G_2 + \frac{1}{\sqrt{n}} \max_{1\le a\le p,1\le t\le k} \|\mathbf{X}^{(t)}_a\|_2 \sqrt{G_5} G_4.$$

Since we have already deduced bounds of G_1, G_2, G_4, and G_5, we only need to deduce the bounds of $\max_{1\le a\le p,1\le t\le k} \frac{1}{\sqrt{n}}\|\mathbf{X}^{(t)}_a\|_2$ in the following.

Under Condition 1, since $\|\mathbf{X}^{(t)}_a\|_2/\sigma_{aa} \sim \chi^2_{(n^{(t)})}$ for any $1 \le a \le p$ with $\sigma^{(t)}_{aa}$ denoting the (a,a)th entry of $\mathbf{\Sigma}^{(t)}$, applying the following tail probability bound with $t = 1/2$ for the chi-squared distribution with $n^{(t)}$ degrees of freedom

$$\mathbb{P}\{|\frac{\chi^2_{(n^{(t)})}}{n^{(t)}} - 1| \ge 1/2\} \le 2\exp(-n^{(t)}/32)$$

gives that

$$\sqrt{\frac{1}{2}}\sigma^{(t)}_{aa} \le \|\mathbf{X}^{(t)}_a\|_2/\sqrt{n^{(t)}} \le \sqrt{\frac{3}{2}}\sigma^{(t)}_{aa}$$

holds with probability at least $1 - 2\exp(-n^{(t)}/32)$.

By Condition 1 that the eigenvalues of $\mathbf{\Sigma}^{(t)}$ are within the interval $[1/M_1, M_1]$, we have $1/M_1 \leq \sigma_{aa}^{(t)} \leq M_1$ for any $1 \leq a \leq p$. Since $n = O(n^{(t)})$ by Condition 2, it follows that for sufficiently large n, with probability at least $1 - 2\exp(-n/32)$,

$$\frac{\|\mathbf{X}_a^{(t)}\|_2}{\sqrt{n}} = \frac{\|\mathbf{X}_a^{(t)}\|_2}{\sqrt{n^{(t)}}}\sqrt{\frac{n^{(t)}}{n}} \leq \sqrt{\frac{n^{(t)}}{n}}\sqrt{\frac{3}{2}\sigma_{aa}^{(t)}} \leq C.$$

Thus, we conclude that with the probability at least $1 - 2pk\exp(-n/32)$

$$\max_{1\leq a\leq p, 1\leq t\leq k} \frac{\|\mathbf{X}_a^{(t)}\|_2}{\sqrt{n}} \leq C. \tag{S4.12}$$

Combining these results (S4.4), (S4.6), (S4.7),(S4.10), and (S4.12), we can deduce that with the probability at least $1 - 4pk\exp(-cn) - 2p^{-\delta} - 2pk\exp(-n/32)$

$$\max_{1\leq a\leq p, 1\leq t\leq k} \|\widehat{\mathbf{V}}_{\widehat{\mathcal{A}}_a}^{(t)} - \mathbf{V}_{\widehat{\mathcal{A}}_a}^{(t)}\|_2 \leq Cb_n. \tag{S4.13}$$

By (S4.2), (S4.11) and (S4.13), we can conclude that with probability at least $1 - 10pk\exp(-cn) - 4p^{-\delta} - 2pk\exp(-n/32)$

$$\max_{1\leq a\leq p} \|\widehat{\mathbf{H}}_{\widehat{\mathcal{A}}_a} - \mathbf{H}_{\widehat{\mathcal{A}}_a}\|_2 \leq Ckb_n,$$

for some positive constants C. Since $\log(p) = o(n)$, there exits some positive constant c_δ such that $p^{-c_\delta} > 10pk\exp(-cn) + 4p^{-\delta} + 2pk\exp(-n/32)$. Further, as $kb_n = o(1)$ assumed in the theorem, it holds that with the probability at least $1 - p^{-c_\delta}$

$$\max_{1\le a\le p} \|\widehat{\mathbf{H}}_{\widehat{\mathcal{A}}_a} - \mathbf{H}_{\widehat{\mathcal{A}}_a}\|_2 = o(1),$$

which completes the proof of Theorem 1.

S5 Proof of Theorem 2

By Lemma S1.4, we have that with the probability at least $1 - p^{-c_\upsilon}$

$$\sup_{\theta\in\mathbf{\Theta}} \sum_{j\in\mathcal{T}_a} \|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta) - \boldsymbol{\beta}^{\text{aug}}_{a(j)}\|_2 \le C_l l_a \lambda, \tag{S5.1}$$

holds simultaneously for all $1 \le a \le p$, where $\lambda = C_\lambda[\frac{k+\log(p)}{n}]^{1/2}$ with $C_\lambda > 0$ some constant and C_l is some positive constant. For the notational simplicity, hereafter we write $W_{a,j}(\theta)$ which are constructed based on $\widehat{\boldsymbol{\beta}}^{\text{aug}}_a(\theta)$ as $\breve{W}_{a,j}$.

Let $|\breve{W}_{a,(1)}| \ge \ldots \ge |\breve{W}_{a,(p-1)}|$ be the ordered knockoff statistics according to magnitude. Denote by j^* the index such that $|\breve{W}_{a,(j^*)}| = T_a(\theta)$. Then, by the definition of $T_a(\theta)$, it holds that $-T_a(\theta) < \breve{W}_{a,(j^*+1)} \le 0$. As in the proof of Theorem 3 in Fan et al. (2020), it's sufficient to consider the two case of $\breve{W}_{a,(j^*+1)} = 0$ and $-T_a(\theta) < \breve{W}_{a,(j^*+1)} < 0$ separately.

Case 1. Consider the case of $-T_a(\theta) < \breve{W}_{a,(j^*+1)} < 0$. In this case, from the definition of threshold $T_a(\theta)$, we have

$$\frac{\gamma + 1 + |\{j \in \{1,\ldots,p\}, \breve{W}_{a,j} \le -T_a(\theta)\}|}{|\widehat{E}(\theta)| \vee 1} > \frac{q}{c_\gamma p}.$$

Using the same argument as in Lemma 6 of Fan et al. (2020) together with Lemma S1.4, we can prove from Condition 7 that $|\widehat{E}(\theta)| \geq M_4|E|$ with asymptotic probability one. This leads to $|\{j \in \{1,\ldots,p\}, \breve{W}_{a,j} \leq -T_a(\theta)\}| > (M_4|E|q)/(c_\gamma p) - \gamma - 1$ with the same probability. Moreover, when $\breve{W}_{a,j} \leq -T_a(\theta)$, we have $\|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta)\|_2 - \|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j+p)}(\theta)\|_2 \leq -T_a(\theta)$, and $\|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j+p)}(\theta)\|_2 \geq T_a(\theta)$. Using equation (S5.1), we obtain

$$C_l l_a \lambda \geq \sum_{j\in\mathcal{T}_a} \|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta) - \boldsymbol{\beta}^{\text{aug}}_{a(j)}\|_2 \geq \sum_{\{j:\breve{W}_{a,j}\leq -T_a\}} \|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j+p)}(\theta)\|_2$$
$$\geq T_a|\{j : \breve{W}_{a,j} \leq -T_a(\theta)\}|.$$

Combining these results leads to $C_l l_a\lambda \geq T_a(\theta)((qM_4|E|)/(c_\gamma p) - 1 - \gamma)$. According to Condition 8 that $|E| \geq \alpha p l_m$, we have

$$T_a(\theta) \leq \frac{C_l l_a \lambda c_\gamma p}{qM_4|E| - pc_\gamma - p\gamma c_\gamma} \leq \frac{C_l l_m \lambda c_\gamma}{qM_4\alpha l_m - c_\gamma - \gamma c_\gamma} \leq \nu_n\lambda/(2C_\lambda) \quad \text{(S5.2)}$$

for large enough n since $\nu_n \to \infty$ as $n \to \infty$.

In light of Equation (S5.1), we derive

$$C_l l_a \lambda \geq \sum_{j\in\mathcal{T}_a} \|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta) - \boldsymbol{\beta}^{\text{aug}}_{a(j)}\|_2 = \sum_{j\in[-a]} \left[\|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta) - \boldsymbol{\beta}^{\text{aug}}_{a(j)}\|_2 + \|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j+p)}(\theta)\|_2\right]$$
$$\geq \sum_{j\in\text{NE}_a\cap(\widehat{\text{NE}}_a(\theta))^c} \left[\|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta) - \boldsymbol{\beta}^{\text{aug}}_{a(j)}\|_2 + \|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j+p)}(\theta)\|_2\right]$$
$$\geq \sum_{j\in\text{NE}_a\cap(\widehat{\text{NE}}_a(\theta))^c} \left[\|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta) - \boldsymbol{\beta}^{\text{aug}}_{a(j)}\|_2 + \|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta)\|_2 - T_a(\theta)\right]$$

since $\|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j+p)}(\theta)\|_2 \geq \|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta)\|_2 - T_a(\theta)$ when $j \in (\widehat{\text{NE}}_a(\theta))^c$. Using the triangle inequality and noting that $\|\boldsymbol{\beta}^{\text{aug}}_{a(j)}\|_2 \geq \nu_n\lambda/C_\lambda$ for $j \in \text{NE}_a$, we can conclude that

$$C_l l_a \lambda \geq \sum_{j\in \text{NE}_a \cap (\widehat{\text{NE}}_a(\theta))^c} (\|\boldsymbol{\beta}^{\text{aug}}_{a(j)}\|_2 - T_a(\theta)) \geq (\nu_n\lambda/C_\lambda - T_a(\theta))|\text{NE}_a \cap (\widehat{\text{NE}}_a(\theta))^c|.$$

Thus, it follows that

$$\frac{|\text{NE}_a \cap \widehat{\text{NE}}_a(\theta)|}{l_a} = 1 - \frac{|\text{NE}_a \cap (\widehat{\text{NE}}_a(\theta))^c|}{l_a} \geq 1 - \frac{C_l\lambda}{\lambda\nu_n/C_\lambda - T_a(\theta)} \geq 1 - \frac{2C_l}{\nu_n C_\lambda}$$

uniformly over all $\theta \in \boldsymbol{\Theta}$ since $T_a(\theta) \leq \nu_n\lambda/(2C_\lambda)$

Case 2. Consider the case of $\breve{W}_{a,(j*+1)} = 0$. In this case, by the definition of $T_a(\theta)$

$$\frac{\gamma + |j : \breve{W}_{a,j} < 0|}{\widehat{E}(\theta)} \leq q/c_\gamma p.$$

If $|j : \breve{W}_{a,j} < 0| > 2C_l C_\lambda c_\gamma l_a \nu_n^{-1}$, then using the same argument as in (S5.2), we can also obtain that $T_a(\theta) \leq \nu_n\lambda/(2C_\lambda)$, and the rest proof is the same as in Case 1. On the other hand, if $|j : \breve{W}_{a,j} < 0| \leq 2C_l C_\lambda c_\gamma l_a \nu_n^{-1}$ we have

$$\begin{aligned} |\widehat{\text{NE}}_a(\theta) \cap \text{NE}_a| &= |\text{supp}(\breve{\mathbf{W}}_a) \cap \text{NE}_a| - |\{j : \breve{W}_{a,j} < 0\} \cap \text{NE}_a| \\ &\geq |\text{supp}(\breve{\mathbf{W}}_a) \cap \text{NE}_a| - 2C_l C_\lambda c_\gamma l_a \nu_n^{-1} \quad \text{(S5.3)} \end{aligned}$$

since $\widehat{\text{NE}}_a(\theta) = \text{supp}(\breve{\mathbf{W}}_a) \backslash \{j : \breve{W}_{a,j} < 0\}$. Let us now focus on $|\text{supp}(\breve{\mathbf{W}}_a) \cap \text{NE}_a|$. We observe that

$$\text{supp}(\breve{\mathbf{W}}_a) \supset \{1, \ldots, p\} \backslash \mathcal{L}_a \quad \text{(S5.4)}$$

where $\mathcal{L}_a = \{1 \le j \le p : \widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta) = 0\}$. Meanwhile, note that in view of Equation (S5.1) we have with probability at least $1 - p^{-c_v}$

$$C_l l_a \lambda \ge \sup_{\theta \in \boldsymbol{\Theta}} \sum_{j \in \mathcal{T}_a} \|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta) - \boldsymbol{\beta}^{\text{aug}}_{a(j)}\|_2 \ge \sup_{\theta \in \boldsymbol{\Theta}} \sum_{j \in \mathcal{L}_a \cap \text{NE}_a} \|\widehat{\boldsymbol{\beta}}^{\text{aug}}_{a(j)}(\theta) - \boldsymbol{\beta}^{\text{aug}}_{a(j)}\|_2$$
$$= \sum_{j \in \mathcal{L}_a \cap \text{NE}_a} \|\boldsymbol{\beta}^{\text{aug}}_{a(j)}\|_2 \ge |\mathcal{L}_a \cap \text{NE}_a| \min_{j \in \text{NE}_a} \|\boldsymbol{\beta}^{\text{aug}}_{a(j)}\|_2.$$

By Condition 6 , we can further deduce from the above inequality that

$$|\mathcal{L}_a \cap \text{NE}_a| \le C_l C_\lambda \nu_n^{-1} l_a,$$

which together with $|\text{NE}_a| = l_a$ entails that

$$|(\{1, \ldots, p\} \backslash \mathcal{L}_a) \cap \text{NE}_a| \ge (1 - C_l C_\lambda \nu_n^{-1}) l_a.$$

Combining this result with Equations (S5.4) yields

$$|\text{supp}(\breve{\mathbf{W}}_a) \cap \text{NE}_a| \ge |(\{1, \ldots, p\} \backslash \mathcal{L}_a) \cap \text{NE}_a| \ge (1 - C_l C_\lambda \nu_n^{-1}) l_a. \quad \text{(S5.5)}$$

Thus, in view of inequalities Equations (S5.3) and (S5.5), with probability at least $1 - p^{-c_v}$ it holds uniformly over all $1 \le a \le p$ that

$$\frac{|\widehat{\text{NE}}_a(\theta) \cap \text{NE}_a|}{l_a} \ge 1 - C_l C_\lambda \nu_n^{-1} - 2 C_l C_\lambda c_\gamma \nu_n^{-1}$$

for all $\theta \in \boldsymbol{\Theta}$.

Combining the above two scenarios, we have shown that with asymptotic probability one, uniformly over all $1 \le a \le p$ it holds that with probability at least $1 - p^{-c_v}$,

$$\frac{|\widehat{\mathrm{NE}}_a(\theta) \cap \mathrm{NE}_a|}{l_a} \geq 1 - C\nu_n^{-1}$$

for all $\theta \in \boldsymbol{\Theta}$, where C is some positive constant. Since

$$|\widehat{E}(\theta) \cap E| \geq \frac{1}{2}\sum_{a=1}^{p} |\mathrm{NE}_a \cap \widehat{\mathrm{NE}}_a(\theta)| \quad \text{and} \quad |E| = \frac{1}{2}\sum_{a=1}^{p} |\mathrm{NE}_a|,$$

then it entails that with probability at least $1 - p^{-c_v}$

$$\frac{|\widehat{E}(\theta) \cap E|}{|E|} \geq \frac{\sum_{a=1}^{p} |\mathrm{NE}_a \cap \widehat{\mathrm{NE}}_a(\theta)|}{\sum_{a=1}^{p} |\mathrm{NE}_a|} \geq \frac{\sum_{a=1}^{p} l_a(1 - C/\nu_n)}{\sum_{a=1}^{p} l_a} \geq 1 - \frac{2C}{\nu_n}.$$

This along with the assumption $\mathbb{P}\{\widehat{\theta} \in \boldsymbol{\Theta}\} \geq 1 - p^{-\delta}$ in Condition 3 gives that

$$\mathrm{Power}(\widehat{E}(\widehat{\theta})) = \mathbb{E}\left[\frac{E \cap \widehat{E}(\widehat{\theta})}{|E|}\right] \geq \mathbb{E}\left[\frac{E \cap \widehat{E}(\widehat{\theta})}{|E|} | \widehat{\theta} \in \boldsymbol{\Theta}\right] \mathbb{P}\{\widehat{\theta} \in \boldsymbol{\Theta}\}$$

$$\geq \left[1 - C\nu_n^{-1}\right](1 - p^{-c_v})(1 - p^{-\delta}) \geq 1 - C\nu_n^{-1} - p^{-\tilde{c}_\delta} + o(\nu_n^{-1}) \to 1$$

for some positive constant $\tilde{c}_\delta$, which concludes the proof of Theorem 2.

S6 Proof of Lemma S1.1

Recalling the expressions of $\mathcal{X}_{-a}$ and $\mathcal{Y}_a$, it yields that

$$(\mathcal{X}_{-a}, \mathcal{Y}_a) = \begin{pmatrix} \mathbf{X}_{-a}^{(1)} & \widetilde{\mathbf{X}}_{-a}^{(1)} & & & & & & \mathbf{X}_a^{(1)} \\ & & \mathbf{X}_{-a}^{(2)} & \widetilde{\mathbf{X}}_{-a}^{(2)} & & & & \mathbf{X}_a^{(2)} \\ & & & & \ddots & \ddots & & \vdots \\ & & & & & \mathbf{X}_{-a}^{(k)} & \widetilde{\mathbf{X}}_{-a}^{(k)} & \mathbf{X}_a^{(k)} \end{pmatrix},$$

and

$$((\mathcal{X}_{-a})_{\text{swap}(S)}, \mathcal{Y}_a) = \begin{pmatrix} [\mathbf{X}_{-a}^{(1)}, \widetilde{\mathbf{X}}_{-a}^{(1)}]_{\text{swap}(S)} & & & & \mathbf{X}_a^{(1)} \\ & [\mathbf{X}_{-a}^{(2)}, \widetilde{\mathbf{X}}_{-a}^{(2)}]_{\text{swap}(S)} & & & \mathbf{X}_a^{(2)} \\ & & \ddots & & \vdots \\ & & & [\mathbf{X}_{-a}^{(k)}, \widetilde{\mathbf{X}}_{-a}^{(k)}]_{\text{swap}(S)} & \mathbf{X}_a^{(k)} \end{pmatrix}.$$

Since $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(k)}$ are independent, it implies that $\left([\mathbf{X}_{-a}^{(t)}, \widetilde{\mathbf{X}}_{-a}^{(t)}], \mathbf{X}_a^{(t)}\right)$ are independent over t from 1 to k. Thus, to prove the claim, it suffices to show that $\left([\mathbf{X}_{-a}^{(t)}, \widetilde{\mathbf{X}}_{-a}^{(t)}], \mathbf{X}_a^{(t)}\right) \overset{d}{=} \left([\mathbf{X}_{-a}^{(t)}, \widetilde{\mathbf{X}}_{-a}^{(t)}]_{\text{swap}(S)}, \mathbf{X}_a^{(t)}\right)$, which trivially follows from the proof of Lemma 3.2 in Candès et al. (2018). Thus, we omitted it here.

S7 Proof of Lemma S1.2

For any given $a \in (1, \ldots, p)$, we can write the statistics $\mathbf{W}_a = f(\mathcal{X}_{-a}, \mathcal{Y}_a)$ for some function f. Let $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_p)$ be a sequence of independent random variables such that $\epsilon_j = \pm 1$ with probability 1/2 if $j \in [-a] \setminus \text{NE}_a$, and $\epsilon_j = 1$ otherwise. To prove the claim, it suffices to establish that

$$\mathbf{W}_a \overset{d}{=} \boldsymbol{\epsilon} \odot \mathbf{W}_a,$$

where $\odot$ denotes pointwise multiplication, i.e. $\boldsymbol{\epsilon} \odot \mathbf{W}_a = (\epsilon_1 W_{a,1}, \ldots, \epsilon_p W_{a,p})$.

Now, let $S = \{j : \epsilon_j = -1\}$. Obviously, $S \subset [-a] \setminus \text{NE}_a$. In view of (2.6), it's easy to see the statistic $W_{a,j}$ satisfies the flip-sign property, that is, when we swap the columns $\mathbf{X}_j^{(t)}$ and $\widetilde{\mathbf{X}}_j^{(t)}$ for all $1 \leq t \leq k$ simultaneously in the matrix $\mathcal{X}_{-a}$, the sign of $W_{a,j}$ will switch. Then we have

$$\boldsymbol{\epsilon} \odot \mathbf{W}_a = f\left((\mathcal{X}_{-a})_{\mathrm{swap}(S)}, \mathcal{Y}_a\right).$$

According to Lemma S1.1, it implies that

$$\mathbf{W}_a = f\left(\mathcal{X}_{-a}, \mathcal{Y}_a\right) \stackrel{d}{=} f\left((\mathcal{X}_{-a})_{\mathrm{swap}(S)}, \mathcal{Y}_a\right) = \boldsymbol{\epsilon} \odot \mathbf{W}_a,$$

which completes the proof of Lemma S1.2.

S8 Proof of Lemma S1.3

It follows from our Lemma S1.2 that $\mathbf{W}_a = (W_{a,j},\, j \in [-a])$ enjoys the sign-flip property on NE_a^c for each $a \in \{1, \ldots, p\}$. Thus, we can use Lemma C.4 in Li and Maathuis (2021) to prove these pairs of $(\gamma, c_\gamma) = (1, 1.93)$ and $(\gamma, c_\gamma) = (0.01, 102)$ satisfying

$$\mathbb{E}\left[\frac{\#\{j \notin \mathrm{NE}_a, W_{a,j} \geq \widehat{T}_a\}}{\gamma + \#\{j \notin \mathrm{NE}_a, W_{a,j} \leq -\widehat{T}_a\}}\right] \leq c_\gamma. \tag{S8.1}$$

In addition, note that the global thresholds is obtained by solving the formula (2.8). If feasible thresholds do not exist, we set $\widehat{\mathbf{T}} = (+\infty, \ldots, +\infty)$. Then the FDR is 0 because no edges can be selected. Thus the inequality

FDR $\leq q$ holds. If feasible thresholds exist, we have

$$\begin{aligned}
&\text{FDR} = \mathbb{E}\left[\frac{|\widehat{E}\cap E^c|}{|\widehat{E}|\vee 1}\right] \leq \mathbb{E}\left[\frac{\sum_{a=1}^{p}\#\{j\notin \text{NE}_a, W_{a,j}\geq \widehat{T}_a\}}{|\widehat{E}|\vee 1}\right]\\
&=\sum_{a=1}^{p}\mathbb{E}\left[\frac{\gamma+\#\{j\in[-a], W_{a,j}\leq -\widehat{T}_a\}}{|\widehat{E}|\vee 1}\frac{\#\{j\notin \text{NE}_a, W_{a,j}\geq \widehat{T}_a\}}{\gamma+\#\{j\in[-a], W_{a,j}\leq -\widehat{T}_a\}}\right]\\
&\leq\sum_{a=1}^{p}\mathbb{E}\left[\frac{\gamma+\#\{j\in[-a], W_{a,j}\leq -\widehat{T}_a\}}{|\widehat{E}|\vee 1}\frac{\#\{j\notin \text{NE}_a, W_{a,j}\geq \widehat{T}_a\}}{\gamma+\#\{j\notin \text{NE}_a, W_{a,j}\leq -\widehat{T}_a\}}\right]\\
&\leq\frac{q}{c_\gamma p}\sum_{a=1}^{p}\mathbb{E}\left[\frac{\#\{j\notin \text{NE}_a, W_{a,j}\geq \widehat{T}_a\}}{\gamma+\#\{j\notin \text{NE}_a, W_{a,j}\leq -\widehat{T}_a\}}\right]\overset{\text{(S8.1)}}{\leq} q,
\end{aligned}$$

where the penultimate inequality follows from the property of the threshold vector and the last inequality is due to (S8.1). Therefore, the FDR of our procedure is controlled, which completes the proof of Lemma S1.3.

S9 Proof of Lemma S1.4

Recalling equation (2.5), the estimated regression coefficient by HGSL with the knockoff matrix based on θ can be written as

$$\widehat{\boldsymbol{\beta}}_a^{\text{aug}}(\theta) = \arg\min_{\mathbf{c}\in\mathbb{R}^{2(p-1)k}}\left\{\sum_{t=1}^{k}Q_{t\theta}(\mathbf{c}^{(t)})+\lambda\left(\sum_{j\in\mathcal{T}_a}\|\mathcal{D}_{a(j)}^{1/2}(\theta)\mathbf{c}_{(j)}\|_2\right)\right\}$$

where

$$Q_{\theta t}(\mathbf{c}^{(t)}) = \frac{\|\mathbf{X}_a^{(t)}-[\mathbf{X}_{-a}^{(t)},\widetilde{\mathbf{X}}_{-a}^{(t)}(\theta)]\mathbf{c}^{(t)}\|_2}{\sqrt{n}}.$$

Before proving (S1.1), we first define an event $\mathcal{B}_a$, that

$$\mathcal{B}_a = \left\{\frac{\max_{j\in\mathcal{T}_a}\|\mathbf{D}_{Ea}^{-1/2}\mathcal{D}_{a(j)}^{-1/2}(\theta)\mathcal{X}_{(j)}^{\top}(\theta)\mathbf{E}_a\|_2}{\sqrt{n}}\leq\lambda\frac{\xi-1}{\xi+1}\right\},$$

where $\mathcal{T}_a = \{1, \ldots, 2p\} \backslash \{a, (a+p)\}$, $\mathbf{D}_{Ea}$ be the $k \times k$ diagonal matrix with tth diagonal entry the squared l_2 norm of the error vector $\mathbf{E}_a^{(t)}$ for $1 \leq t \leq k$, and $\mathcal{X}_{(j)}(\theta)$ is an $N \times k$ submatrix of $\mathcal{X}_{-a}(\theta)$ given by columns corresponding to the jth group.

Note that combined with knockoff matrices, there are $2(p-1)$ group in $\mathcal{X}_{-a}(\theta)$. For convenience, we index them by $\mathcal{T}_a = \{1, \ldots, 2p\} \backslash \{a, (a+p)\}$ according to the corresponding index of variables and knockoff variables. Specifically, for the index $1 \leq j \leq p$, it corresponds to the original variable X_j. For the index $j \in (p+1, \ldots, 2p)$, it corresponds to the knockoff variable $\widetilde{X}_{j-p}$. Thus, for $j \leq p$, $\mathcal{D}_{a(j)}(\theta) = \mathbf{D}_{a(j)}$, and for $j > p$, $\mathcal{D}_{a(j)}(\theta) = \widetilde{\mathbf{D}}_{a(l)}(\theta)$ with $l = j - p$, as defined in equation (2.5).

Using the same proof technique as Lemma D.6 in Ren et al. (2019), we can prove that this event $\mathcal{B}_a$ holds with probability at least $1 - 3p^{1-\upsilon}$. Conditional on the events $\{\bigcap_{1 \leq a \leq p} \mathcal{B}_a\}$ and $\theta \in \mathbf{\Theta}$, the proof of Lemma S1.4 is similar to that of Theorem 3.1 in Ren et al. (2019). We will simplify the same parts as that in the proof of Theorem 3.1 in Ren et al. (2019), and emphasize the additional proof techniques and steps that needed to deal with the barriers causing by the estimated precision matrices.

Refer to the proof of Theorem 3.1 in Ren et al. (2019). Denote by $\bar{\boldsymbol{\beta}}_a^{\text{aug}} = \mathcal{D}_a^{1/2}(\theta)\boldsymbol{\beta}_a^{\text{aug}}$, $\hat{\bar{\boldsymbol{\beta}}}_a^{\text{aug}}(\theta) = \mathcal{D}_a^{1/2}(\theta)\widehat{\boldsymbol{\beta}}_a^{\text{aug}}(\theta)$, and $\bar{\triangle}_a = \hat{\bar{\boldsymbol{\beta}}}_a^{\text{aug}}(\theta) - \bar{\boldsymbol{\beta}}_a^{\text{aug}}$. In what follows, we establish all results in terms of $\bar{\triangle}_a$. As pointed in Ren et al. (2019) this does not affect our result much since our Condition 1 and the fact of $(\mathbf{X}_l^{(t)})^\top \mathbf{X}_l^{(t)}/(\mathbf{\Sigma}_0^{(t)})_{l,l} \sim \chi^2(n^{(t)})$ and $(\widetilde{\mathbf{X}}_l^{(t)})^\top \widetilde{\mathbf{X}}_l^{(t)}/(\mathbf{\Sigma}_0^{(t)})_{l,l} \sim \chi^2(n^{(t)})$, together with an application of Lemma E.1 of Ren et al. (2019) and the union bound, entail that with probability at least $1 - 2pk\exp(-n/32)$,

$$M_1/2 \leq (\mathbf{X}_l^{(t)})^\top \mathbf{X}_l^{(t)}/n^{(t)} \leq 3M_1/2 \text{ and } M_1/2 \leq (\widetilde{\mathbf{X}}_l^{(t)})^\top \widetilde{\mathbf{X}}_l^{(t)}/n^{(t)} \leq 3M_1/2$$

holds simultaneously for all $1 \leq l \leq p$, $1 \leq t \leq k$.

Recalling the knockoff generation process (2.3), for the original data matrix $\mathbf{X}_{-a}^{(t)}$, the ideal knockoff matrix and the approximate knockoff matrix constructed using some given precision matrices θ can be represented as

$$\widetilde{\mathbf{X}}_{-a}^{(t)} = \mathbf{X}_{-a}^{(t)}\mathbf{C}_a^{(t)} + \mathbf{Z}_a^{(t)}\mathbf{B}_a^{(t)} \quad \text{and} \quad \widetilde{\mathbf{X}}_{-a}^{(t)}(\theta) = \mathbf{X}_{-a}^{(t)}\mathbf{C}_a^{(t)}(\theta) + \mathbf{Z}_a^{(t)}\mathbf{B}_a^{(t)}(\theta).$$

respectively, where $\mathbf{Z}_a^{(t)}$ is a random matrix whose rows are i.i.d copies of $N(\mathbf{0}, \mathbf{I_{p-1}})$ and independent of $\mathbf{X}_{-a}^{(t)}$, $\mathbf{C}_a^{(t)} = \mathbf{I}_{p-1} - \text{diag}\{\mathbf{s}_a^{(t)}\}\mathbf{\Omega}_{-a}^{(t)}$ and $\mathbf{C}_a^{(t)}(\theta) = \mathbf{I}_{p-1} - \text{diag}\{\mathbf{s}_a^{(t)}\}\mathbf{\Gamma}_{-a}^{(t)}$, $\mathbf{B}_a^{(t)} = \Big(2\text{diag}\{\mathbf{s}_a^{(t)}\} - \text{diag}\{\mathbf{s}_a^{(t)}\}\mathbf{\Omega}_{-a}^{(t)}\text{diag}\{\mathbf{s}_a^{(t)}\}\Big)^{1/2}$, and $\mathbf{B}_a^{(t)}(\theta) = \Big(2\text{diag}\{\mathbf{s}_a^{(t)}\} - \text{diag}\{\mathbf{s}_a^{(t)}\}\mathbf{\Gamma}_{-a}^{(t)}\text{diag}\{\mathbf{s}_a^{(t)}\}\Big)^{1/2}$.

According to Lemma B.1 in Kaul et al. (2019), we have with the probability at least $1 - p^{-c}$ for some positive constant c

$$\begin{aligned}
&\max_{1\leq a\leq p, 1\leq t\leq k} \|(\mathbf{C}_a^{(t)}(\theta) - \mathbf{C}_a^{(t)})^\top (\mathbf{X}_{-a}^{(t)})^\top \mathbf{X}_{-a}^{(t)}(\mathbf{C}_a^{(t)}(\theta) - \mathbf{C}_a^{(t)})/n^{(t)}\|_\infty \\
&\leq \max_{1\leq a\leq p, 1\leq t\leq k} \|(\mathbf{C}_a^{(t)}(\theta) - \mathbf{C}_a^{(t)})^\top \mathbf{\Sigma}_{-a,-a}(\mathbf{C}_a^{(t)}(\theta) - \mathbf{C}_a^{(t)})\|_\infty + C\sqrt{\log(p)/n^{(t)}}
\end{aligned}$$

Since $\theta \in \Theta$, by Condition 1 and equations (S4.7), we have

$$\max_{1\le a\le p, 1\le t\le k} \|(\mathbf{C}_a^{(t)}(\theta) - \mathbf{C}_a^{(t)})^\top \boldsymbol{\Sigma}_{-a,-a}(\mathbf{C}_a^{(t)}(\theta) - \mathbf{C}_a^{(t)})\|_\infty$$

$$\le \max_{1\le a\le p, 1\le t\le k} \|\mathbf{C}_a^{(t)}(\theta) - \mathbf{C}_a^{(t)}\|_2^2 \|\Sigma\|_2 \le Cb_n,$$

which shows that with the probability at least $1 - p^{-c}$

$$\max_{1\le a\le p, 1\le t\le k} \|(\mathbf{C}_a^{(t)}(\theta) - \mathbf{C}_a^{(t)})^\top (\mathbf{X}_{-a}^{(t)})^\top \mathbf{X}_{-a}^{(t)}(\mathbf{C}_a^{(t)}(\theta) - \mathbf{C}_a^{(t)})/n^{(t)}\|_\infty = O(b_n + d_n),$$

where $d_n = \sqrt{\log(p)/n^{(t)}}$.

Using the same technique, combining with the result of (S4.10), we can prove that with the probability at least $1 - p^{-c}$

$$\max_{1\le a\le p, 1\le t\le k} \|(\mathbf{C}_a^{(t)}(\theta) - \mathbf{C}_a^{(t)})^\top (\mathbf{X}_{-a}^{(t)})^\top \mathbf{Z}_{-a}^{(t)}(\mathbf{B}_a^{(t)}(\theta) - \mathbf{B}_a^{(t)})/n^{(t)}\|_\infty = O(d_n),$$

and with the probability at least $1 - p^{-c}$

$$\max_{1\le a\le p, 1\le t\le k} \|(\mathbf{B}_a^{(t)}(\theta) - \mathbf{B}_a^{(t)})^\top (\mathbf{Z}_{-a}^{(t)})^\top \mathbf{Z}_{-a}^{(t)}(\mathbf{B}_a^{(t)}(\theta) - \mathbf{B}_a^{(t)})/n^{(t)}\|_\infty = O(b_n + d_n).$$

By the expression of triangle inequality, we can derive

$$\max_{1\le a\le p, 1\le t\le k} \|(\widetilde{\mathbf{X}}_{-a}^{(t)}(\theta))^\top \widetilde{\mathbf{X}}_{-a}^{(t)}(\theta)/n^{(t)} - (\widetilde{\mathbf{X}}_{-a}^{(t)})^\top \widetilde{\mathbf{X}}_{-a}^{(t)}/n^{(t)}\|_\infty$$

$$\le \max_{1\le a\le p, 1\le t\le k} \|(\mathbf{C}_a^{(t)}(\theta) - \mathbf{C}_a^{(t)})^\top (\mathbf{X}_{-a}^{(t)})^\top \mathbf{X}_{-a}^{(t)}(\mathbf{C}_a^{(t)}(\theta) - \mathbf{C}_a^{(t)})/n^{(t)}\|_\infty$$

$$+ 2 \max_{1\le a\le p, 1\le t\le k} \|(\mathbf{C}_a^{(t)}(\theta) - \mathbf{C}_a^{(t)})^\top (\mathbf{X}_{-a}^{(t)})^\top \mathbf{Z}_{-a}^{(t)}(\theta) - \mathbf{B}_a^{(t)})/n^{(t)}\|_\infty$$

$$+ \max_{1\le a\le p, 1\le t\le k} \|(\mathbf{B}_a^{(t)}(\theta) - \mathbf{B}_a^{(t)})^\top (\mathbf{Z}_{-a}^{(t)})^\top \mathbf{Z}_{-a}^{(t)}(\mathbf{B}_a^{(t)}(\theta) - \mathbf{B}_a^{(t)})/n^{(t)}\|_\infty$$

Then, we can conclude that

$$\max_{1\le l\le p,1\le t\le k} |(\widetilde{\mathbf{X}}_l^{(t)}(\theta))^\top \widetilde{\mathbf{X}}_l^{(t)}(\theta)/n^{(t)} - (\widetilde{\mathbf{X}}_l^{(t)})^\top \widetilde{\mathbf{X}}_l^{(t)}/n^{(t)}| = O(b_n + \sqrt{\log(p)/n^{(t)}})$$

holds with probability at least $1-3p^{-c}$. It entails that for sufficiently large n there is a positive constant $\widetilde{M}_1$ that

$$\widetilde{M}_1/2 \le (\widetilde{\mathbf{X}}_l^{(t)}(\theta))^\top \widetilde{\mathbf{X}}_l^{(t)}(\theta)/n^{(t)} \le 3\widetilde{M}_1/2 \tag{S9.1}$$

holds uniformly over $1 \le l \le p$,$1 \le t \le k$ with probability at least $1 - 3p^{-c} - 2pk\exp(-n/32)$ since $b_n = o(1)$ and $\log(p) = o(n)$. Therefore, $\triangle$ and $\bar{\triangle}$ are of the same order componentwise and globally.

Using the same techniques as the *Step 1* of the proof of Theorem 3.1 in Ren et al. (2019), we can obtain that for all $1 \le a \le p$,

$$\sum_{l\in\mathcal{T}_a/\mathrm{NE}_a} \|\bar{\triangle}_{a(l)}\|_2 \le \xi \sum_{l\in\mathrm{NE}_a} \|\bar{\triangle}_{a(l)}\|_2. \tag{S9.2}$$

Further, under Conditions 1-2 and the fact that $\mathbf{X}_a^{(t)\prime}\mathbf{X}_a^{(t)}/(\mathbf{\Sigma}_0^{(t)})_{a,a} \sim \chi^2(n^{(t)})$ and $\mathbf{E}_a^{(t)\prime}\mathbf{E}_a^{(t)}/\omega_{aa}^{(t)} \sim \chi^2(n^{(t)})$, by Lemma D.7 in Ren et al. (2019) it holds that with the probability at least $1-4k\exp(-n/32)$,

$$\sum_{t=1}^{k} \frac{\|\bar{\mathcal{X}}_{-a}^{(t)}(\theta)\bar{\triangle}_a^{(t)}\|^2}{n\zeta_t} \ge \frac{1}{\sqrt{6M_1M_2}} \sum_{t=1}^{k} \frac{\|\bar{\mathcal{X}}_{-a}^{(t)}(\theta)\bar{\triangle}_a^{(t)}\|^2}{n} \tag{S9.3}$$

where $\bar{\mathcal{X}}_{-a}^{(t)}(\theta) = \mathcal{X}_{-a}^{(t)}(\theta)(\mathcal{D}_a^{(t)}(\theta))^{-1/2}$ and $\zeta_t = Q_t\left(\hat{\bar{\boldsymbol{\beta}}}_a^{\mathrm{aug}(t)}(\theta)\right) + Q_t\left(\bar{\boldsymbol{\beta}}_a^{\mathrm{aug}(t)}\right)$.

In addition, since the facts of (A.27), (A.28), and (A.30) in the supplementary materials of Ren et al. (2019) all hold here, using the same techniques of (A.31) in the supplementary materials of Ren et al. (2019), we can prove that

$$\sum_{t=1}^{k}\frac{\|\bar{\mathcal{X}}_{-a}^{(t)}(\theta)\bar{\triangle}_a^{(t)}\|^2}{n\zeta_t}\le\lambda\left(\frac{3\xi-1}{\xi+1}+\xi\frac{(\xi-3)_+}{\xi+1}\right)\sum_{l\in\mathrm{NE}_a}\|\bar{\triangle}_{a(l)}\|_2, \tag{S9.4}$$

Combining with the result of (S9.3), we can conclude that

$$\sum_{t=1}^{k}\frac{\|\bar{\mathcal{X}}_{-a}^{(t)}(\theta)\bar{\triangle}_a^{(t)}\|^2}{n}\le\frac{\lambda}{\sqrt{6M_1M_2}}\left(\frac{3\xi-1}{\xi+1}+\xi\frac{(\xi-3)_+}{\xi+1}\right)\sum_{l\in\mathrm{NE}_a}\|\bar{\triangle}_{a(l)}\|_2 \tag{S9.5}$$

holds with the probability at least $1-4k\exp(-n/32)$.

For convenience, let

$$C_\xi=\frac{1}{\sqrt{6M_1M_2}}\left(\frac{3\xi-1}{\xi+1}+\xi\frac{(\xi-3)_+}{\xi+1}\right).$$

which is shown in equation (S9.4). The inequality (S9.5) implies that with the probability at least $1-4k\exp(-n/32)$

$$\sum_{t=1}^{k}\frac{(\bar{\triangle}_a^{(t)})^\top(\bar{\mathcal{X}}_{-a}^{(t)}(\theta))^\top\bar{\mathcal{X}}_{-a}^{(t)}(\theta)\bar{\triangle}_a^{(t)}}{n}\le C_\xi\lambda\sum_{l\in\mathrm{NE}_a}\|\bar{\triangle}_{a(l)}\|_2. \tag{S9.6}$$

Following the same definition as that in Ren et al. (2019), let

$$\kappa(\xi,\mathrm{NE}_a)=\inf_{\mathbf{u}\neq\mathbf{0}}\left\{\frac{\sqrt{|\mathrm{NE}_a|}\|\bar{\mathcal{X}}_{-a}(\theta)\mathbf{u}\|_2}{\sqrt{n}\sum_{j\in\mathrm{NE}_a}\|\mathbf{u}_{(j)}\|_2}:\mathbf{u}\in\Psi(\xi,\mathrm{NE}_a)\right\},$$

where $\Psi(\xi,\mathrm{NE}_a)=\{\mathbf{u}\in\mathbb{R}^{2(p-1)k}:\sum_{j\in\mathcal{T}_a/\mathrm{NE}_a}\|\mathbf{u}_{(j)}\|_2\le\xi\sum_{j\in\mathrm{NE}_a}\|\mathbf{u}_{(j)}\|_2\}$.

Since $\bar{\triangle}_a\in\Phi(\xi,\mathrm{NE}_a)$ by (S9.2), we have

$$\sum_{t=1}^{k}\frac{(\bar{\triangle}_a^{(t)})^\top(\bar{\mathcal{X}}_{-a}^{(t)}(\theta))^\top\bar{\mathcal{X}}_{-a}^{(t)}(\theta)\bar{\triangle}_a^{(t)}}{n}\ge\frac{\kappa^2(\xi,\mathrm{NE}_a)\left(\sum_{j\in\mathrm{NE}_a}\|\bar{\triangle}_{a(j)}\|_2\right)^2}{|\mathrm{NE}_a|} \tag{S9.7}$$

Recall that for each given θ, the rows of $\mathcal{X}_{-a}^{(t)}(\theta)$ follow the Gaussian distribution $N(\mathbf{0}, \mathcal{G}(\theta))$, where

$$\mathcal{G}_a(\theta) \begin{pmatrix} \boldsymbol{\Sigma}_{-a,-a}^{(t)} & \mathbf{C}_a^{(t)}(\theta)\boldsymbol{\Sigma}_{-a,-a}^{(t)} \\ \mathbf{C}_a^{(t)}(\theta)\boldsymbol{\Sigma}_{-a,-a}^{(t)} & \mathbf{C}_a^{(t)}(\theta)\boldsymbol{\Sigma}_{-a,-a}^{(t)}\mathbf{C}_a^{(t)}(\theta) + \mathbf{B}_a^{(t)}(\theta) \end{pmatrix}.$$

According to Condition 1 and the assumption in the theorem that the smallest eigenvalue of $2\text{diag}\{\mathbf{s}_a^{(t)}\} - \text{diag}\{\mathbf{s}_a^{(t)}\}\boldsymbol{\Omega}_{-a}^{(t)}\text{diag}\{\mathbf{s}_a^{(t)}\}$ is uniformly bounded from below by some positive constant for all $1 \le a \le p$, $1 \le t \le k$, we have

$$1/M_6 \le \lambda_{\min}(\mathcal{G}_a^{(t)}(\theta_0)) \le \lambda_{\max}(\mathcal{G}_a^{(t)}(\theta_0)) \le M_6$$

holds uniformly over $1 \le a \le p$, $1 \le t \le k$ for some constant $M_6 > 1$.

Moreover, by the triangle inequality we have

$$\begin{aligned} &\lambda_{\min}(\mathcal{G}_a^{(t)}(\theta)) \ge \lambda_{\min}(\mathcal{G}_a^{(t)}(\theta_0)) + \lambda_{\min}(\mathcal{G}_a^{(t)}(\theta) - \mathcal{G}_a^{(t)}(\theta_0)) \\ &\ge \lambda_{\min}(\mathcal{G}_a^{(t)}(\theta_0)) - \lambda_{\max}(\mathcal{G}_a^{(t)}(\theta) - \mathcal{G}_a^{(t)}(\theta_0)). \end{aligned}$$

Since $\theta \in \boldsymbol{\Theta}$, according to the equation (S4.7) and (S4.10) , we have $\lambda_{\max}(\mathcal{G}_a^{(t)}(\theta) - \mathcal{G}_a^{(t)}(\theta_0)) \le Cb_n$ holds uniformly over $1 \le a \le p$, $1 \le t \le k$. Then for sufficiently large n, we can conclude that there exist some positive constant $\widetilde{M}_6$ that

$$1/\widetilde{M}_6 \le \lambda_{\min}(\mathcal{G}_a^{(t)}(\theta)) \le \lambda_{\max}(\mathcal{G}_a^{(t)}(\theta)) \le \widetilde{M}_6.$$

Thus, according to the Lemma D.5 in Ren et al. (2019) shows that

$$\kappa(\xi, \mathrm{NE}_a) > \min_{l,t} \left\{ \left(\frac{n^{(t)}}{(\mathbf{X}_l^{(t)}(\theta))^\top \mathbf{X}_l^{(t)}(\theta)} \right)^{1/2} \wedge \left(\frac{n^{(t)}}{(\widetilde{\mathbf{X}}_l^{(t)}(\theta)^\top \widetilde{\mathbf{X}}_l^{(t)}(\theta)} \right)^{1/2} \right\} /(2\widetilde{M}_6)^{1/2},$$

holds with probability at least $1 - 2k\exp(-cn)$. Combining with result of (S9.1), yields that with the probability at least $1 - 2pk\exp(-n/32) - 2k\exp(-cn) - 3p^{-c}$, $\kappa(\xi, \mathrm{NE}_a) > C_\kappa$, where C_κ is some positive constant. This together with (S9.6) and (S9.7) yields that with the probability at least $1 - 2pk\exp(-n/32) - 2k\exp(-cn) - 4k\exp(-n/32) - 3p^{-c}$

$$\frac{\left(\sum_{j\in \mathrm{NE}_a} \|\bar{\triangle}_{a(j)}\|_2\right)^2}{|\mathrm{NE}_a|} \leq C\lambda \sum_{l\in \mathrm{NE}_a} \|\bar{\triangle}_{a(l)}\|_2 \tag{S9.8}$$

Thus it holds simultaneously for all $1 \leq a \leq p$ that

$$\sup_{\theta\in\mathbf{\Theta}} \sum_{j\in \mathrm{NE}_a} \|\bar{\triangle}_{a(j)}\|_2 = O(l_a\lambda)$$

with the probability at least $1-2pk\exp(-n/32)-2k\exp(-cn)-4k\exp(-n/32)-3p^{-c}$. This together with (S9.1) and (S9.2) yields that

$$\begin{aligned}\sup_{\theta\in\mathbf{\Theta}} \sum_{j\in \mathcal{T}_a} \|\triangle_{a(j)}\|_2 \leq C \sup_{\theta\in\mathbf{\Theta}} \sum_{j\in \mathcal{T}_a} \|\bar{\triangle}_{a(j)}\| \\ \leq (1+\xi) C \sup_{\theta\in\mathbf{\Theta}} \sum_{j\in \mathrm{NE}_a} \|\bar{\triangle}_{a(j)}\| = O(l_a\lambda)\end{aligned}$$

holds simultaneously over $1 \leq a \leq p$ with the probability at least $1 - 2pk\exp(-n/32) - 2k\exp(-cn) - 4k\exp(-n/32) - 3p^{-c}$.

Note that the above analysis is conditional on the event $\bigcap_{1\leq a\leq p} \mathcal{B}_a$.

Since we have proven that $\mathbb{P}(\mathcal{B}_a) > 1 - 3p^{-\upsilon+1}$, it easy to prove that

$$\mathbb{P}(\bigcap_{1\le a\le p} \mathcal{B}_a) \ge 1 - 3pp^{1-\upsilon} \ge 1 - p^{2-\upsilon}$$

Then, we can conclude that with probability at least $1-2pk\exp(-n/32) - 2k\exp(-cn) - 4k\exp(-n/32) - p^{2-\upsilon} - 3p^{-c}$

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{j\in\mathcal{T}_a} \|\triangle_{a(j)}\|_2 = O(l_a\lambda),$$

holds simultaneously for all $1 \le a \le p$. Since $\log(p) = O(n)$, and $\upsilon > 2$ defined in the lemma, then there exits some positive constant c_υ such that $p^{-c_\upsilon} > 2pk\exp(-n/32) + 2k\exp(-cn) + 4k\exp(-n/32) + p^{2-\upsilon} + 3p^{-c}$, which completes the proof of Lemma S1.4.

S10 The figures of real data analysis

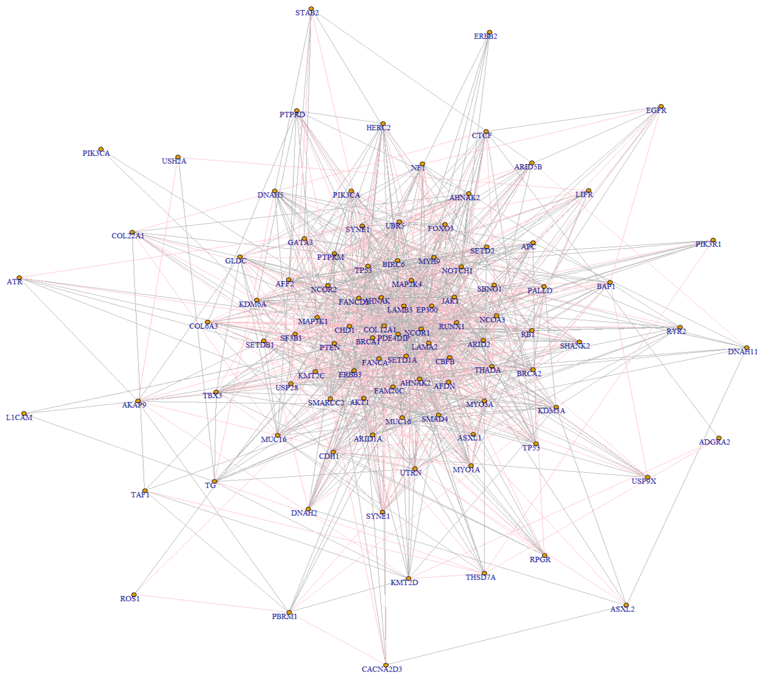

Figure 1: Network structure of genes recovered by our method. The pink lines are the edges identified by both our method and GFC method. Grey lines are identified only by our method.

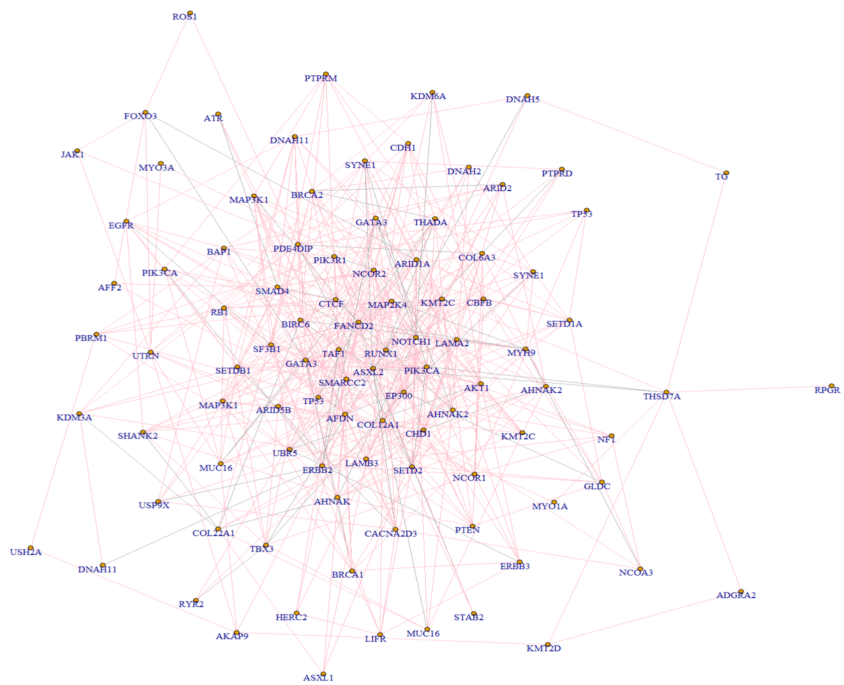

Figure 2: Network structure of genes recovered by GFC method. The pink lines are the edges identified by both GFC method and our method. Grey lines are identified only by GFC method.

Bibliography

Candès, E., Y. Fan, L. Janson, and J. Lv (2018). Panning for gold: 'model-x'knockoffs for high dimensional controlled variable selection. *J. R. Stat. Soc. Ser.* B *80*(3), 551–577.

Fan, Y., E. Demirkaya, G. Li, and J. Lv (2020). RANK: large-scale inference with graphical nonlinear knockoffs. *J. Amer. Statist. Assoc. 115*(529), 362–379.

Fan, Y., J. Lv, M. Sharifvaghefib, and Y. Uematsua (2020). Ipad: Stable interpretable forecasting with knockoffs inference. *J. Amer. Statist. Assoc. 115*(532), 1822–1834.

Kaul, A., V. K. Jandhyala, and S. B. Fotopoulos (2019). An efficient two step algorithm for high dimensional change point regression models without grid search. *J. Mach. Learn. Res. 20*(1), 1–40.

Li, J. and M. H. Maathuis (2021). GGM knockoff filter: False discovery rate control for Gaussian graphical models. *J. R. Stat. Soc. Ser.* B *83*(3), 534–558.

Ren, Z., Y. Kang, Y. Fan, and J. Lv (2019). Tuning-free heterogeneous inference in massive network. *J. Amer. Statist. Assoc. 114*(528), 1908–1925.

Schmitt, B. A. (1992). Perturbation bounds for matrix square roots and pythagorean sums. *Linear Algebra Appl. 174*, 215–227.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*.