Supplementary Material to "RANK: Large-Scale Inference with Graphical Nonlinear Knockoffs"

Yingying Fan, Emre Demirkaya, Gaorong Li and Jinchi Lv

This Supplementary Material contains additional technical details for the proofs of Lemmas 3–8. All the notation is the same as in the main body of the paper.

B Additional technical details

B.1 Lemma 3 and its proof

Lemma 3. *Assume that* $\mathbf{X} = (X_{ij}) \in \mathbb{R}^{n\times p}$ *has independent rows with distribution* $N(\mathbf{0}, \boldsymbol{\Sigma}_0)$, $\Lambda_{\max}(\boldsymbol{\Sigma}_0) \leq M$, *and* $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)^T$ *has i.i.d. components with* $\mathbb{P}\{|\varepsilon_i| > t\} \leq C_1 \exp(-C_1^{-1}t^2)$ *for* $t > 0$ *and some constants* $M, C_1 > 0$. *Then we have*

$$\mathbb{P}\left\{\left\|\frac{1}{n}\mathbf{X}^T\boldsymbol{\varepsilon}\right\|_\infty \leq C\sqrt{(\log p)/n}\right\} \geq 1 - p^{-c}$$

for some constant $c > 0$ *and large enough constant* $C > 0$.

Proof. First observe that $\mathbb{P}(|X_{ij}| > t) \leq 2\exp\{-(2M)^{-1}t^2\}$ for $t > 0$, since $X_{ij} \sim N(0, \boldsymbol{\Sigma}_{0,jj})$ and $\boldsymbol{\Sigma}_{0,jj} \leq \Lambda_{\max}(\boldsymbol{\Sigma}_0) \leq M$, where $\boldsymbol{\Sigma}_{0,jj}$ denotes the jth diagonal entry of matrix $\boldsymbol{\Sigma}_0$. By assumption, we also have $\mathbb{P}(|\varepsilon_i| > t) \leq C_1 \exp\{-C_1^{-1}t^2\}$. Combining these two inequalities yields

$$\begin{aligned}
\mathbb{P}(|\varepsilon_i X_{ij}| > t) &\leq \mathbb{P}(|\varepsilon_i| > \sqrt{t}) + \mathbb{P}(|X_{ij}| > \sqrt{t}) \\
&\leq C_1 \exp\{-C_1^{-1}t\} + 2\exp\{-(2M)^{-1}t\} \\
&\leq C_2 \exp\{-C_2^{-1}t\},
\end{aligned}$$

where $C_2 > 0$ is some constant that depends only on constants C_1 and M. Thus by Lemma 6 in [28], there exists some constant $\widetilde{C}_1 > 0$ such that

$$\mathbb{P}(|n^{-1}\sum_{i=1}^n \varepsilon_i X_{ij}| > z) \leq \widetilde{C}_1 \exp\{-\widetilde{C}_1 n z^2) \tag{A.1}$$

for all $0 < z < 1$.

Denote by $\mathbf{X}_j$ the jth column of matrix $\mathbf{X}$. Then by (A.1), the union bound leads to

$$\begin{aligned}
1 - \mathbb{P}\left(\left\|n^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}\right\|_\infty \leq z\right) &= \mathbb{P}\left(\left\|n^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}\right\|_\infty > z\right) \\
&= \mathbb{P}\Big(\max_{1\leq j\leq p} |n^{-1}\boldsymbol{\varepsilon}^T\mathbf{X}_j| > z\Big) \\
&\leq \sum_{j=1}^p \mathbb{P}(|n^{-1}\sum_{i=1}^n \varepsilon_i X_{ij}| > z) \\
&\leq p\widetilde{C}_1 \exp\{-\widetilde{C}_1 n z^2).
\end{aligned}$$

Letting $z = C\sqrt{(\log p)/n}$ in the above inequality, we obtain

$$\mathbb{P}\Big(\Big\|n^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}\Big\|_\infty \le C\sqrt{(\log p)/n}\Big) \ge 1 - \widetilde{C}_1 p^{-(\widetilde{C}_1 C^2 - 1)}.$$

Taking large enough positive constant C completes the proof of Lemma 3.

B.2 Lemma 4 and its proof

Lemma 4. *Assume that all the conditions of Proposition 2 hold and* $a_n[(L_p+L_p')^{1/2}+K_n^{1/2}] = o(1)$*. Then we have*

$$P\left\{\sup_{\boldsymbol{\Omega}\in\mathcal{A},\,|\mathcal{S}|\le K_n}\Big\|\widetilde{\boldsymbol{\rho}}_{\mathcal{S}} - \widetilde{\mathbf{G}}_{\mathcal{S},\mathcal{S}}\boldsymbol{\beta}_{\mathbb{T},\mathcal{S}}\Big\|_\infty \le C_4\sqrt{(\log p)/n}\right\} = 1 - O(p^{-c_4})$$

for some constants $c_4, C_4 > 0$.

Proof. In this proof, we use c and C to denote generic positive constants and use the same notation as in the proof of Proposition 2 in Section A.6. Since $\boldsymbol{\beta}_{\mathbb{T}} = (\boldsymbol{\beta}_0^T, 0, \ldots, 0)^T$ with $\boldsymbol{\beta}_0$ the true regression coefficient vector, it is easy to check that $\widetilde{\mathbf{X}}_{\text{KO}}\boldsymbol{\beta}_{\mathbb{T}} = \mathbf{X}\boldsymbol{\beta}_0$. In view of $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}$, it follows from the definitions of $\widetilde{\boldsymbol{\rho}}$ and $\widetilde{\mathbf{G}}$ that

$$\begin{aligned}
\widetilde{\boldsymbol{\rho}}_{\mathcal{S}} - \widetilde{\mathbf{G}}_{\mathcal{S},\mathcal{S}}\boldsymbol{\beta}_{\mathbb{T},\mathcal{S}} &= \frac{1}{n}\widetilde{\mathbf{X}}_{\text{KO},\mathcal{S}}^T\mathbf{X}\boldsymbol{\beta}_0 + \frac{1}{n}\widetilde{\mathbf{X}}_{\text{KO},\mathcal{S}}^T\boldsymbol{\varepsilon} - \frac{1}{n}\widetilde{\mathbf{X}}_{\text{KO},\mathcal{S}}^T\widetilde{\mathbf{X}}_{\text{KO},\mathcal{S}}\boldsymbol{\beta}_{\mathbb{T},\mathcal{S}} \\
&= \frac{1}{n}\mathbf{X}_{\text{KO},\mathcal{S}}^T\boldsymbol{\varepsilon} + \frac{1}{n}(\widetilde{\mathbf{X}}_{\text{KO},\mathcal{S}} - \mathbf{X}_{\text{KO},\mathcal{S}})^T\boldsymbol{\varepsilon}.
\end{aligned}$$

Using the triangle inequality, we deduce

$$\|\widetilde{\boldsymbol{\rho}}_{\mathcal{S}} - \widetilde{\mathbf{G}}_{\mathcal{S},\mathcal{S}}\boldsymbol{\beta}_{\mathbb{T},\mathcal{S}}\|_\infty \le \Big\|\frac{1}{n}\mathbf{X}_{\text{KO},\mathcal{S}}^T\boldsymbol{\varepsilon}\Big\|_\infty + \Big\|\frac{1}{n}(\widetilde{\mathbf{X}}_{\text{KO},\mathcal{S}} - \mathbf{X}_{\text{KO},\mathcal{S}})^T\boldsymbol{\varepsilon}\Big\|_\infty.$$

We will bound both terms on the right hand side of the above inequality.

By Lemma 3, we can show that for the first term,

$$\Big\|\frac{1}{n}\mathbf{X}_{\text{KO},\mathcal{S}}^T\boldsymbol{\varepsilon}\Big\|_\infty \le \Big\|\frac{1}{n}\mathbf{X}_{\text{KO}}^T\boldsymbol{\varepsilon}\Big\|_\infty \le C\sqrt{(\log p)/n}$$

with probability at least $1 - p^{-c}$ for some constants $C, c > 0$. We will prove that with probability at least $1 - o(p^{-c})$,

$$\Big\|\frac{1}{n}(\widetilde{\mathbf{X}}_{\text{KO},\mathcal{S}} - \mathbf{X}_{\text{KO},\mathcal{S}})^T\boldsymbol{\varepsilon}\Big\|_\infty \le Ca_n(L_p + L_p')^{1/2}\sqrt{(\log p)/n} + Ca_n\sqrt{n^{-1}K_n(\log p)}. \quad \text{(A.2)}$$

Then the desired result in this lemma can be shown by noting that $a_n[(L_p+L_p')^{1/2}+K_n^{1/2}] \to 0$.

It remains to prove (A.2). Recall that matrices $\breve{\mathbf{X}}_{\mathcal{S}}$ and $\breve{\mathbf{X}}_{0,\mathcal{S}}$ can be written as

$$\breve{\mathbf{X}}_{\mathcal{S}} = \mathbf{X}(\mathbf{I} - \mathbf{\Omega}\text{diag}\{\mathbf{s}\})_{\mathcal{S}} + \mathbf{Z}\mathbf{B}_{0,\mathcal{S}}(\mathbf{B}_{0,\mathcal{S}}^T\mathbf{B}_{0,\mathcal{S}})^{-1/2}\Big((\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}})^T\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}}\Big)^{1/2},$$
$$\breve{\mathbf{X}}_{0,\mathcal{S}} = \mathbf{X}(\mathbf{I} - \mathbf{\Omega}_0\text{diag}\{\mathbf{s}\})_{\mathcal{S}} + \mathbf{Z}\mathbf{B}_{0,\mathcal{S}},$$

where the notation is the same as in the proof of Proposition 2 in Section A.6. By the definitions of $\widetilde{\mathbf{X}}_{\text{KO}}$ and $\mathbf{X}_{\text{KO}}$, it holds that

$$\Big\|\frac{1}{n}(\widetilde{\mathbf{X}}_{\text{KO},\mathcal{S}} - \mathbf{X}_{\text{KO},\mathcal{S}})^T\boldsymbol{\varepsilon}\Big\|_\infty = \Big\|\frac{1}{n}(\breve{\mathbf{X}}_{\mathcal{S}} - \breve{\mathbf{X}}_{0,\mathcal{S}})^T\boldsymbol{\varepsilon}\Big\|_\infty, \tag{A.3}$$

where $\breve{\mathbf{X}}_{\mathcal{S}}$ and $\breve{\mathbf{X}}_{0,\mathcal{S}}$ represent the submatrices formed by columns in $\mathcal{S}$. We now turn to analyzing the term $n^{-1}(\breve{\mathbf{X}}_{\mathcal{S}} - \breve{\mathbf{X}}_{0,\mathcal{S}})^T\boldsymbol{\varepsilon}$. Some routine calculations give

$$\begin{aligned}\frac{1}{n}(\breve{\mathbf{X}}_{\mathcal{S}} - \breve{\mathbf{X}}_{0,\mathcal{S}})^T\boldsymbol{\varepsilon} \ &= \ \frac{1}{n}\Big(\big((\mathbf{\Omega}_0 - \mathbf{\Omega})\text{diag}\{\mathbf{s}\}\big)_{\mathcal{S}}\Big)^T\mathbf{X}^T\boldsymbol{\varepsilon}\\ &\quad + \frac{1}{n}\Big(((\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}})^T\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}})^{1/2}(\mathbf{B}_{0,\mathcal{S}}^T\mathbf{B}_{0,\mathcal{S}})^{-1/2} - \mathbf{I}\Big)\mathbf{B}_{0,\mathcal{S}}^T\mathbf{Z}^T\boldsymbol{\varepsilon}.\end{aligned}$$

Thus it follows from $s_j \leq 2\Lambda_{\max}(\mathbf{\Sigma}_0)$ for all $1 \leq j \leq p$ and the triangle inequality that

$$\begin{aligned}\Big\|\frac{1}{n}(\breve{\mathbf{X}}_{\mathcal{S}} - \breve{\mathbf{X}}_{0,\mathcal{S}})^T\boldsymbol{\varepsilon}\Big\|_\infty \ &\leq \ 2\Lambda_{\max}(\mathbf{\Sigma}_0)\Big\|\frac{1}{n}(\mathbf{\Omega}_{0,\mathcal{S}} - \mathbf{\Omega}_{\mathcal{S}})^T\mathbf{X}^T\boldsymbol{\varepsilon}\Big\|_\infty\\ &\quad + \ \Big\|\frac{1}{n}\Big(((\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}})^T\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}})^{1/2}(\mathbf{B}_{0,\mathcal{S}}^T\mathbf{B}_{0,\mathcal{S}})^{-1/2} - \mathbf{I}\Big)\mathbf{B}_{0,\mathcal{S}}^T\mathbf{Z}^T\boldsymbol{\varepsilon}\Big\|_\infty.\end{aligned} \tag{A.4}$$

We first examine the upper bound for $\Big\|\frac{1}{n}(\mathbf{\Omega}_{0,\mathcal{S}} - \mathbf{\Omega}_{\mathcal{S}})^T\mathbf{X}^T\boldsymbol{\varepsilon}\Big\|_\infty$ in (A.4). Since $\mathbf{\Omega} \in \mathcal{A}$ and $\mathbf{\Omega}_0$ is L_p-sparse, by Lemma 3 we deduce

$$\begin{aligned}\Big\|\frac{1}{n}(\mathbf{\Omega}_{0,\mathcal{S}} - \mathbf{\Omega}_{\mathcal{S}})^T\mathbf{X}^T\boldsymbol{\varepsilon}\Big\|_\infty \ &\leq \ \Big\|\frac{1}{n}(\mathbf{\Omega}_0 - \mathbf{\Omega})\mathbf{X}^T\boldsymbol{\varepsilon}\Big\|_\infty\\ &\leq \ \|\mathbf{\Omega}_0 - \mathbf{\Omega}\|_1\Big\|\frac{1}{n}\mathbf{X}^T\boldsymbol{\varepsilon}\Big\|_\infty\\ &\leq \ \sqrt{L_p + L_p'}\|\mathbf{\Omega} - \mathbf{\Omega}_0\|_2 \cdot C\sqrt{(\log p)/n}\\ &\leq \ Ca_n(L_p + L_p')^{1/2}\sqrt{(\log p)/n}.\end{aligned} \tag{A.5}$$

We can also bound the second term on the right hand side of (A.4) as

$$\begin{aligned}&\Big\|\frac{1}{n}\Big(((\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}})^T\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}})^{1/2}(\mathbf{B}_{0,\mathcal{S}}^T\mathbf{B}_{0,\mathcal{S}})^{-1/2} - \mathbf{I}\Big)\mathbf{B}_{0,\mathcal{S}}^T\mathbf{Z}^T\boldsymbol{\varepsilon}\Big\|_\infty\\ &\leq \Big\|((\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}})^T\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}})^{1/2}(\mathbf{B}_{0,\mathcal{S}}^T\mathbf{B}_{0,\mathcal{S}})^{-1/2} - \mathbf{I}\Big\|_1\Big\|\frac{1}{n}\mathbf{B}_{0,\mathcal{S}}^T\mathbf{Z}^T\boldsymbol{\varepsilon}\Big\|_\infty\\ &\leq \sqrt{2|\mathcal{S}|}\Big\|((\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}})^T\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}})^{1/2}(\mathbf{B}_{0,\mathcal{S}}^T\mathbf{B}_{0,\mathcal{S}})^{-1/2} - \mathbf{I}\Big\|_2\Big\|\frac{1}{n}\mathbf{B}_{0,\mathcal{S}}^T\mathbf{Z}^T\boldsymbol{\varepsilon}\Big\|_\infty\\ &\leq \sqrt{2K_n}Ca_n\sqrt{(\log p)/n} = Ca_n\sqrt{n^{-1}K_n(\log p)},\end{aligned}$$

where the second to the last step is entailed by Lemma 2 in Section A.3 and Lemma 5 in Section B.3. Therefore, combining this inequality with (A.3)–(A.5) results in (A.2), which concludes the proof of Lemma 4.

B.3 Lemma 5 and its proof

Lemma 5. *Under the conditions of Proposition 2, it holds that with probability at least* $1 - O(p^{-c})$,

$$\sup_{|\mathcal{S}| \leq K_n} \left\| \frac{1}{n} \mathbf{B}_{0,\mathcal{S}}^T \mathbf{Z}^T \boldsymbol{\varepsilon} \right\|_\infty \geq C\sqrt{(\log p)/n}$$

for some constant $C > 0$.

Proof. Since this is a specific case of Lemma 8 in Section B.6, the proof is omitted.

B.4 Lemma 6 and its proof

Lemma 6. *Under the conditions of Proposition 2 and Lemma 1, there exists some constant* $c \in (2(qs)^{-1}, 1)$ *such that with asymptotic probability one,* $|\widehat{\mathcal{S}}^{\mathbf{\Omega}}| \geq cs$ *holds uniformly over all* $\mathbf{\Omega} \in \mathcal{A}$ *and* $|\mathcal{S}| \leq K_n$*, where* $\widehat{\mathcal{S}}^{\mathbf{\Omega}} = \{j : W_j^{\mathbf{\Omega},\mathcal{S}} \geqslant T\}$.

Proof. Again we use C to denote generic positive constants whose values may change from line to line. By Proposition 2 in Section A.6, we have with probability at least $1 - O(p^{-c_1})$ that uniformly over all $\mathbf{\Omega} \in \mathcal{A}$ and $|\mathcal{S}| \leq K_n$,

$$\max_{1\leq j\leq p} |\widehat{\beta}_j(\lambda; \mathbf{\Omega}, \mathcal{S}) - \beta_{0,j}| \leq C\sqrt{sn^{-1}(\log p)} \text{ and } \max_{1\leq j\leq p} |\widehat{\beta}_{j+p}(\lambda; \mathbf{\Omega}, \mathcal{S})| \leq C\sqrt{sn^{-1}(\log p)}$$

for some constants $C, c_1 > 0$. Thus for each $1 \leq j \leq p$, we have

$$\begin{aligned} W_j^{\mathbf{\Omega},\mathcal{S}} &= |\widehat{\beta}_j(\lambda; \mathbf{\Omega}, \mathcal{S})| - |\widehat{\beta}_{j+p}(\lambda; \mathbf{\Omega}, \mathcal{S})| \\ &\geq -|\widehat{\beta}_{j+p}(\lambda; \mathbf{\Omega}, \mathcal{S})| \geq -C\sqrt{sn^{-1}(\log p)}. \end{aligned} \tag{A.6}$$

On the other hand, for each $j \in \mathcal{S}_2 = \{j : \beta_{0,j} \gg \sqrt{sn^{-1}(\log p)}\}$ it holds that

$$\begin{aligned} W_j^{\mathbf{\Omega},\mathcal{S}} &= |\widehat{\beta}_j(\lambda; \mathbf{\Omega}, \mathcal{S})| - |\widehat{\beta}_{j+p}(\lambda; \mathbf{\Omega}, \mathcal{S})| \\ &\geq |\beta_{0,j}| - |\widehat{\beta}_j(\lambda; \mathbf{\Omega}, \mathcal{S}) - \beta_{0,j}| - |\widehat{\beta}_{j+p}(\lambda; \mathbf{\Omega}, \mathcal{S})| \gg C\sqrt{sn^{-1}(\log p)}. \end{aligned} \tag{A.7}$$

Thus in order for any $W_j^{\mathbf{\Omega},\mathcal{S}}$, $1 \leq j \leq p$ to fall below $-T$, we must have $W_j^{\mathbf{\Omega},\mathcal{S}} \geq T$ for all $j \in \mathcal{S}_2$. This entails that

$$\left|\{j : W_j^{\mathbf{\Omega},\mathcal{S}} \geq T\}\right| \geq |\mathcal{S}_2| \geq cs, \tag{A.8}$$

which completes the proof of Lemma 6.

B.5 Lemma 7 and its proof

Lemma 7. *Assume that all the conditions of Proposition 2 hold and* $a_{2n} = a_n + (L_p' + K_n)\{(\log p)/n\}^{1/2} = o(1)$*. Then it holds that*

$$P\left\{\sup_{\boldsymbol{\Omega}\in\mathcal{A},\,|\mathcal{S}|\le K_n}\left\|\widetilde{\mathbf{G}}_{\mathcal{S},\mathcal{S}} - \mathbf{G}_{\mathcal{S},\mathcal{S}}\right\|_{\max} \le C_8 a_{2,n}\right\} = 1 - O(p^{-c_8})$$

for some constants $c_8, C_8 > 0$.

Proof. In this proof, we adopt the same notation as used in the proof of Proposition 2 in Section A.6. In light of (36), we have $\widetilde{\mathbf{G}} = n^{-1}[\mathbf{X}, \breve{\mathbf{X}}^{\boldsymbol{\Omega}}]^T[\mathbf{X}, \breve{\mathbf{X}}^{\boldsymbol{\Omega}}]$. Thus the matrix difference $\widetilde{\mathbf{G}}_{\mathcal{S},\mathcal{S}} - \mathbf{G}_{\mathcal{S},\mathcal{S}}$ can be represented in block form as

$$\begin{aligned}
\widetilde{\mathbf{G}}_{\mathcal{S},\mathcal{S}} - \mathbf{G}_{\mathcal{S},\mathcal{S}} &= \frac{1}{n}\begin{pmatrix} \mathbf{X}_{\mathcal{S}}^T\mathbf{X}_{\mathcal{S}} & (\breve{\mathbf{X}}_{\mathcal{S}}^{\boldsymbol{\Omega}})^T\mathbf{X}_{\mathcal{S}} \\ \mathbf{X}_{\mathcal{S}}^T\breve{\mathbf{X}}_{\mathcal{S}}^{\boldsymbol{\Omega}} & (\breve{\mathbf{X}}_{\mathcal{S}}^{\boldsymbol{\Omega}})^T\breve{\mathbf{X}}_{\mathcal{S}}^{\boldsymbol{\Omega}} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\Sigma}_0 & \boldsymbol{\Sigma}_0 - \mathrm{diag}\{\mathbf{s}\} \\ \boldsymbol{\Sigma}_0 - \mathrm{diag}\{\mathbf{s}\} & \boldsymbol{\Sigma}_0 \end{pmatrix}_{\mathcal{S},\mathcal{S}} \\
&= \begin{pmatrix} n^{-1}\mathbf{X}_{\mathcal{S}}^T\mathbf{X}_{\mathcal{S}} - \boldsymbol{\Sigma}_{0,\mathcal{S},\mathcal{S}} & n^{-1}(\breve{\mathbf{X}}_{\mathcal{S}}^{\boldsymbol{\Omega}})^T\mathbf{X}_{\mathcal{S}} - \big(\boldsymbol{\Sigma}_0 - \mathrm{diag}\{\mathbf{s}\}\big)_{\mathcal{S},\mathcal{S}} \\ n^{-1}\mathbf{X}_{\mathcal{S}}^T\breve{\mathbf{X}}_{\mathcal{S}}^{\boldsymbol{\Omega}} - \big(\boldsymbol{\Sigma}_0 - \mathrm{diag}\{\mathbf{s}\}\big)_{\mathcal{S},\mathcal{S}} & n^{-1}(\breve{\mathbf{X}}_{\mathcal{S}}^{\boldsymbol{\Omega}})^T\breve{\mathbf{X}}_{\mathcal{S}}^{\boldsymbol{\Omega}} - \boldsymbol{\Sigma}_{0,\mathcal{S},\mathcal{S}} \end{pmatrix}.
\end{aligned}$$

Note that the off-diagonal blocks are the transposes of each other. Then we see that $\|\widetilde{\mathbf{G}}_{\mathcal{S},\mathcal{S}} - \mathbf{G}_{\mathcal{S},\mathcal{S}}\|_{\max}$ can be bounded by the maximum of $\|\eta_1\|_{\max}$, $\|\eta_2\|_{\max}$, and $\|\eta_3\|_{\max}$ with

$$\begin{aligned}
\eta_1 &= n^{-1}\mathbf{X}_{\mathcal{S}}^T\mathbf{X}_{\mathcal{S}} - \boldsymbol{\Sigma}_{0,\mathcal{S},\mathcal{S}}, \\
\eta_2 &= n^{-1}\mathbf{X}_{\mathcal{S}}^T\breve{\mathbf{X}}_{\mathcal{S}}^{\boldsymbol{\Omega}} - \big(\boldsymbol{\Sigma}_0 - \mathrm{diag}\{\mathbf{s}\}\big)_{\mathcal{S},\mathcal{S}}, \\
\eta_3 &= n^{-1}(\breve{\mathbf{X}}_{\mathcal{S}}^{\boldsymbol{\Omega}})^T\breve{\mathbf{X}}_{\mathcal{S}}^{\boldsymbol{\Omega}} - \boldsymbol{\Sigma}_{0,\mathcal{S},\mathcal{S}}.
\end{aligned}$$

To bound these three terms, we define three events

$$\begin{aligned}
\mathcal{E}_5 &= \left\{\|n^{-1}\mathbf{X}^T\mathbf{X} - \boldsymbol{\Sigma}_0\|_{\max} \le C\sqrt{(\log p)/n}\right\}, \\
\mathcal{E}_6 &= \left\{\sup_{|\mathcal{S}|\le K_n}\left\|n^{-1}\mathbf{B}_{0,\mathcal{S}}^T\mathbf{Z}^T\mathbf{X}\right\|_\infty \le C\sqrt{(\log p)/n}\right\}, \\
\mathcal{E}_7 &= \left\{\sup_{|\mathcal{S}|\le K_n}\left\|n^{-1}\mathbf{B}_{0,\mathcal{S}}^T\mathbf{Z}^T\mathbf{Z}\mathbf{B}_{0,\mathcal{S}} - \mathbf{B}_{0,\mathcal{S}}^T\mathbf{B}_{0,\mathcal{S}}\right\|_{\max} \le C\sqrt{(\log p)/n}\right\}.
\end{aligned}$$

By Lemma 8 in Section B.6, it holds that $P(\mathcal{E}_6) \ge 1 - O(p^{-c})$ and $P(\mathcal{E}_7) \ge 1 - O(p^{-c})$. Using Lemma A.3 in [6], we also have $P(\mathcal{E}_5) \ge 1 - O(p^{-c})$. Combining these results yields

$$P(\mathcal{E}_5 \cap \mathcal{E}_6 \cap \mathcal{E}_7) \ge 1 - O(p^{-c})$$

with $c > 0$ some constant.

Let us first consider term η_1. Conditional on $\mathcal{E}_5$, it is easy to see that

$$\|\eta_1\|_{\max} \le \|n^{-1}\mathbf{X}^T\mathbf{X} - \boldsymbol{\Sigma}_0\|_{\max} \le C\sqrt{(\log p)/n}. \tag{A.9}$$

We next bound $\|\eta_2\|_{\max}$ conditional on $\mathcal{E}_5 \cap \mathcal{E}_6$. To simplify the notation, denote by $\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}} = (\mathbf{B}_{0,\mathcal{S}}^T\mathbf{B}_{0,\mathcal{S}})^{-1/2}\Big((\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}})^T\mathbf{B}_{\mathcal{S}}^{\mathbf{\Omega}}\Big)^{1/2}$. By the definition of $\check{\mathbf{X}}_{\mathcal{S}}$, we deduce

$$
\begin{aligned}
\eta_2 &= n^{-1}\mathbf{X}_{\mathcal{S}}^T\check{\mathbf{X}}_{\mathcal{S}}^{\mathbf{\Omega}} - \big(\mathbf{\Sigma}_0 - \mathrm{diag}\{\mathbf{s}\}\big)_{\mathcal{S},\mathcal{S}} \\
&= n^{-1}\mathbf{X}_{\mathcal{S}}^T\mathbf{X}(\mathbf{I} - \mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})_{\mathcal{S}} + n^{-1}\mathbf{X}_{\mathcal{S}}^T\mathbf{Z}\mathbf{B}_{0,\mathcal{S}}\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}} - \big(\mathbf{\Sigma}_0 - \mathrm{diag}\{\mathbf{s}\}\big)_{\mathcal{S},\mathcal{S}} \\
&= \big((n^{-1}\mathbf{X}^T\mathbf{X} - \mathbf{\Sigma}_0)(\mathbf{I} - \mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})\big)_{\mathcal{S},\mathcal{S}} + \big(\mathrm{diag}\{\mathbf{s}\} - \mathbf{\Sigma}_0\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\}\big)_{\mathcal{S},\mathcal{S}} + n^{-1}\mathbf{X}_{\mathcal{S}}^T\mathbf{Z}\mathbf{B}_{0,\mathcal{S}}\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}} \\
&\equiv \eta_{2,1} + \eta_{2,2} + \eta_{2,3}.
\end{aligned}
$$

We will examine the above three terms separately.

Since $\mathbf{\Omega}$ is L_p'-sparse, $\|\mathbf{I}-\mathbf{\Omega}_0\mathrm{diag}(\mathbf{s})\|_2 \leq \|\mathbf{I}\|_2+\|\mathbf{\Omega}_0\mathrm{diag}(\mathbf{s})\|_2 \leq C$, and $\|(\mathbf{\Omega}-\mathbf{\Omega}_0)\mathrm{diag}\{\mathbf{s}\}\|_2 \leq Ca_n$, we have

$$
\begin{aligned}
\Big\|\mathbf{I} - \mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\}\Big\|_1 &\leq \sqrt{L_p'}\Big\|\mathbf{I} - \mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\}\Big\|_2 \\
&\leq \sqrt{L_p'}\Big(\big\|\mathbf{I} - \mathbf{\Omega}_0\mathrm{diag}\{\mathbf{s}\}\big\|_2 + \big\|(\mathbf{\Omega} - \mathbf{\Omega}_0)\mathrm{diag}\{\mathbf{s}\}\big\|_2\Big) \\
&\leq C\sqrt{L_p'}. \qquad \text{(A.10)}
\end{aligned}
$$

Thus it follow from (A.10) that conditional on $\mathcal{E}_5$,

$$
\begin{aligned}
\|\eta_{2,1}\|_{\max} &= \Big\|\big((n^{-1}\mathbf{X}^T\mathbf{X} - \mathbf{\Sigma}_0)(\mathbf{I} - \mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})\big)_{\mathcal{S},\mathcal{S}}\Big\|_{\max} \\
&\leq \Big\|(n^{-1}\mathbf{X}^T\mathbf{X} - \mathbf{\Sigma}_0)(\mathbf{I} - \mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})\Big\|_{\max} \\
&\leq \Big\|n^{-1}\mathbf{X}^T\mathbf{X} - \mathbf{\Sigma}_0\Big\|_{\max}\Big\|\mathbf{I} - \mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\}\Big\|_1 \\
&\leq C\sqrt{L_p'}\sqrt{(\log p)/n}. \qquad \text{(A.11)}
\end{aligned}
$$

For term $\eta_{2,2}$, it holds that

$$
\begin{aligned}
\|\eta_{2,2}\|_{\max} &= \Big\|\big(\mathrm{diag}\{\mathbf{s}\} - \mathbf{\Sigma}_0\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\}\big)_{\mathcal{S},\mathcal{S}}\Big\|_{\max} \\
&\leq C\|\mathbf{I} - \mathbf{\Sigma}_0\mathbf{\Omega}\|_{\max} \leq C\|\mathbf{\Sigma}_0\|_2\|\mathbf{\Omega}_0 - \mathbf{\Omega}\|_2 \leq Ca_n. \qquad \text{(A.12)}
\end{aligned}
$$

Note that by Lemma 2 in Section A.3, we have

$$
\|\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}}\|_1 \leq \sqrt{|\mathcal{S}|}\|\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}}\|_2 \leq \sqrt{|\mathcal{S}|}(\|\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}} - \mathbf{I}\|_2 + 1) \leq C\sqrt{|\mathcal{S}|} \leq C\sqrt{K_n}
$$

when $|\mathcal{S}| \leq K_n$. Then conditional on $\mathcal{E}_6$, it holds that

$$
\begin{aligned}
\|\eta_{2,3}\|_{\max} &= \|n^{-1}\mathbf{X}_{\mathcal{S}}^T\mathbf{Z}\mathbf{B}_{0,\mathcal{S}}\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}}\|_{\max} \\
&\leq \|n^{-1}\mathbf{X}_{\mathcal{S}}^T\mathbf{Z}\mathbf{B}_{0,\mathcal{S}}\|_{\max}\|\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}}\|_1 \\
&\leq C\sqrt{n^{-1}K_n(\log p)}. \qquad \text{(A.13)}
\end{aligned}
$$

Thus combining (A.11)–(A.13) leads to

$$\|\eta_2\|_{\max} \le C\{a_n + \sqrt{n^{-1}L'_p(\log p)} + \sqrt{n^{-1}K_n(\log p)}\}. \tag{A.14}$$

We finally deal with term η_3. Some routine calculations show that

$$\begin{aligned}
\eta_3 &= n^{-1}(\breve{\mathbf{X}}_{\mathcal{S}}^{\mathbf{\Omega}})^T\breve{\mathbf{X}}_{\mathcal{S}}^{\mathbf{\Omega}} - \mathbf{\Sigma}_{0,\mathcal{S},\mathcal{S}}.\\
&= n^{-1}\big((\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})_{\mathcal{S}}^T\mathbf{X}^T + (\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}})^T\mathbf{B}_{0,\mathcal{S}}^T\mathbf{Z}^T\big)\big(\mathbf{X}(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})_{\mathcal{S}} + \mathbf{Z}\mathbf{B}_{0,\mathcal{S}}\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}}\big) - \mathbf{\Sigma}_{0,\mathcal{S},\mathcal{S}}\\
&= \Big(n^{-1}(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})^T\mathbf{X}^T\mathbf{X}(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\}) - \mathbf{\Sigma}_0 + \mathbf{B}_0^T\mathbf{B}_0\Big)_{\mathcal{S},\mathcal{S}}\\
&+ n^{-1}(\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}})^T\mathbf{B}_{0,\mathcal{S}}^T\mathbf{Z}^T\mathbf{X}(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})_{\mathcal{S}} + (\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})_{\mathcal{S}}^T\mathbf{X}^T\mathbf{Z}\mathbf{B}_{0,\mathcal{S}}\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}}\\
&+ \big((\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}})^T\mathbf{B}_{0,\mathcal{S}}^T\mathbf{Z}^T\mathbf{Z}\mathbf{B}_{0,\mathcal{S}}\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}} - \mathbf{B}_{0,\mathcal{S}}^T\mathbf{B}_{0,\mathcal{S}}\big)\\
&\equiv \eta_{3,1} + \eta_{3,2} + \eta_{3,2}^T + \eta_{3,3}.
\end{aligned}$$

Conditional on event $\mathcal{E}_5$, with some simple matrix algebra we derive

$$\begin{aligned}
\|\eta_{3,1}\| &= \Big\|\Big(n^{-1}(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})^T\mathbf{X}^T\mathbf{X}(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\}) - \mathbf{\Sigma}_0 + \mathbf{B}_0^T\mathbf{B}_0\Big)_{\mathcal{S},\mathcal{S}}\Big\|_{\max}\\
&\le \Big\|n^{-1}(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})^T\mathbf{X}^T\mathbf{X}(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\}) - \mathbf{\Sigma}_0 + \mathbf{B}_0^T\mathbf{B}_0\Big\|_{\max}\\
&\le \Big\|(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})^T(n^{-1}\mathbf{X}^T\mathbf{X} - \mathbf{\Sigma}_0)(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})\Big\|_{\max}\\
&+ \Big\|(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})^T\mathbf{\Sigma}_0(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\}) - \mathbf{\Sigma}_0 + 2\mathrm{diag}\{\mathbf{s}\} - \mathrm{diag}\{\mathbf{s}\}\mathbf{\Omega}_0\mathrm{diag}\{\mathbf{s}\}\Big\|_{\max}\\
&\le \|n^{-1}\mathbf{X}^T\mathbf{X} - \mathbf{\Sigma}_0\|_{\max}\|(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})\|_1^2\\
&+ \|\mathrm{diag}\{\mathbf{s}\}(\mathbf{I}-\mathbf{\Omega}\mathbf{\Sigma}_0)\|_{\max} + \|(\mathbf{I}-\mathbf{\Sigma}_0\mathbf{\Omega})\mathrm{diag}\{\mathbf{s}\}\|_{\max} + \|\mathrm{diag}\{\mathbf{s}\}(\mathbf{\Omega}_0 - \mathbf{\Omega}\mathbf{\Sigma}_0\mathbf{\Omega})\mathrm{diag}\{\mathbf{s}\}\|_{\max}\\
&\le CL'_p\sqrt{(\log p)/n} + Ca_n, \qquad\qquad (\mathrm{A.15})
\end{aligned}$$

where the last step used (A.10) and calculations similar to (A.12).

It follows from (A.10) and the previously proved result $\|\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}}\|_1 \le C\sqrt{K_n}$ for $|\mathcal{S}| \le K_n$ that conditional on event $\mathcal{E}_6$,

$$\begin{aligned}
\|\eta_{3,2}\| &= \|n^{-1}(\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}})^T\mathbf{B}_{0,\mathcal{S}}^T\mathbf{Z}^T\mathbf{X}(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})_{\mathcal{S}}\|_{\max}\\
&\le \|\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}}\|_1\|n^{-1}\mathbf{B}_{0,\mathcal{S}}^T\mathbf{Z}^T\mathbf{X}\|_{\max}\|(\mathbf{I}-\mathbf{\Omega}\mathrm{diag}\{\mathbf{s}\})_{\mathcal{S}}\|_1\\
&\le C\sqrt{K_n}\sqrt{L'_pn^{-1}(\log p)}\\
&= C\sqrt{n^{-1}K_nL'_p(\log p)}. \qquad\qquad (\mathrm{A.16})
\end{aligned}$$

Finally, by Lemma 2 it holds that conditioned on $\mathcal{E}_7$,

$$
\begin{aligned}
\|\eta_{3,3}\| &= \left\| n^{-1}(\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}})^T \mathbf{B}_{0,\mathcal{S}}^T \mathbf{Z}^T \mathbf{Z} \mathbf{B}_{0,\mathcal{S}} \widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}} - \mathbf{B}_{0,\mathcal{S}}^T \mathbf{B}_{0,\mathcal{S}} \right\|_{\max} \\
&\le \left\| (\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}})^T (n^{-1}\mathbf{B}_{0,\mathcal{S}}^T \mathbf{Z}^T \mathbf{Z} \mathbf{B}_{0,\mathcal{S}} - \mathbf{B}_{0,\mathcal{S}}^T \mathbf{B}_{0,\mathcal{S}}) \widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}} \right\|_{\max} \\
&+ \left\| (\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}})^T \mathbf{B}_{0,\mathcal{S}}^T \mathbf{B}_{0,\mathcal{S}} \widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}} - \mathbf{B}_{0,\mathcal{S}}^T \mathbf{B}_{0,\mathcal{S}} \right\|_{\max} \\
&\le \left\| n^{-1}\mathbf{B}_{0,\mathcal{S}}^T \mathbf{Z}^T \mathbf{Z} \mathbf{B}_{0,\mathcal{S}} - \mathbf{B}_{0,\mathcal{S}}^T \mathbf{B}_{0,\mathcal{S}} \right\|_{\max} \|\widetilde{\mathbf{B}}^{\mathcal{S},\mathbf{\Omega}}\|_1^2 + Ca_n \\
&\le CK_n\sqrt{(\log p)/n} + Ca_n. \qquad \text{(A.17)}
\end{aligned}
$$

Therefore, combining (A.15)–(A.17) results in

$$
\begin{aligned}
\|\eta_3\|_{\max} &\le Ca_n + C(L_p' + K_n + \sqrt{K_n L_p'})\sqrt{(\log p)/n} \\
&\le Ca_n + 2C(L_p' + K_n)\sqrt{(\log p)/n},
\end{aligned}
$$

which together with (A.9) and (A.14) concludes the proof of Lemma 7.

B.6 Lemma 8 and its proof

Lemma 8. *Under the conditions of Proposition 2, it holds that with probability at least* $1 - O(p^{-c})$,

$$
\begin{aligned}
&\sup_{|\mathcal{S}|\le K_n} \left\| \frac{1}{n}\mathbf{B}_{0,\mathcal{S}}^T \mathbf{Z}^T \mathbf{X} \right\|_{\max} \ge C\sqrt{(\log p)/n}, \\
&\sup_{|\mathcal{S}|\le K_n} \left\| n^{-1}\mathbf{B}_{0,\mathcal{S}}^T \mathbf{Z}^T \mathbf{Z} \mathbf{B}_{0,\mathcal{S}} - \mathbf{B}_{0,\mathcal{S}}^T \mathbf{B}_{0,\mathcal{S}} \right\|_{\max} \ge C\sqrt{(\log p)/n}
\end{aligned}
$$

for some constants $c, C > 0$.

Proof. We still use c and C to denote generic positive constants. We start with proving the first inequality. Observe that

$$
\sup_{|\mathcal{S}|\le K_n} \left\| \frac{1}{n}\mathbf{B}_{0,\mathcal{S}}^T \mathbf{Z}^T \mathbf{X} \right\|_{\max} \le \left\| \frac{1}{n}\mathbf{B}_0^T \mathbf{Z}^T \mathbf{X} \right\|_{\max}.
$$

Thus it remains to prove

$$
P\left(\left\| \frac{1}{n}\mathbf{B}_0^T \mathbf{Z}^T \mathbf{X} \right\|_{\max} \ge C\sqrt{(\log p)/n} \right) \le o(p^{-c}). \qquad \text{(A.18)}
$$

Let $\mathbf{U} = \mathbf{Z}\mathbf{B}_0 \in \mathbb{R}^{n\times p}$ and denote by $\mathbf{U}_j$ the jth column of matrix $\mathbf{U}$. We see that the components of $\mathbf{U}_j$ are i.i.d. Gaussian with mean zero and variance $\mathbf{e}_j^T \mathbf{B}_0^T \mathbf{B}_0 \mathbf{e}_j$, and the vectors $\mathbf{U}_j$ are independent of $\boldsymbol{\varepsilon}$. Let $\widetilde{\mathbf{U}}_j = (\mathbf{e}_j^T \mathbf{B}_0^T \mathbf{B}_0 \mathbf{e}_j)^{-1/2}\mathbf{U}_j$. Then it holds that $\widetilde{\mathbf{U}}_j \sim N(\mathbf{0}, \mathbf{I}_n)$. Since $X_{ij} \sim N(0, \mathbf{\Sigma}_{0,jj})$ and $\mathbf{\Sigma}_{0,jj} \le \Lambda_{\max}(\mathbf{\Sigma}_0) \le C$ with $C > 0$ some constant, it follows from Bernstein's inequality that for $t > 0$,

$$\begin{aligned}
\mathbb{P}\left(\left\|\frac{1}{n}\mathbf{B}_0^T\mathbf{Z}^T\mathbf{X}\right\|_{\max} \geq t\|\mathbf{B}_0^T\mathbf{B}_0\|_2\right) &\leq \sum_{j=1}^{p}\mathbb{P}\left(\frac{1}{n}\left|(\mathbf{U}_j)^T\mathbf{X}_i\right| \geq t\|\mathbf{B}_0^T\mathbf{B}_0\|_2\right) \\
&\leq \sum_{j=1}^{p}\mathbb{P}\left(\frac{1}{n}\left|(\widetilde{\mathbf{U}}_j)^T\mathbf{X}_i\right| \geq t\right) \\
&\leq Cp\exp(-Cnt^2).
\end{aligned}$$

Taking $t = C\sqrt{(\log p)/n}$ with large enough constant $C > 0$ in the above inequality yields

$$\mathbb{P}\left(\left\|\frac{1}{n}\mathbf{B}_0^T\mathbf{Z}^T\mathbf{X}\right\|_{\max} \geq C\sqrt{(\log p)/n}\|\mathbf{B}_0^T\mathbf{B}_0\|_2\right) \leq Cp^{-c}$$

for some constant $c > 0$. Thus with probability at least $1 - O(p^{-c})$, it holds that

$$\begin{aligned}
&\left\|\frac{1}{n}\mathbf{B}_0^T\mathbf{Z}^T\mathbf{X}\right\|_{\max} \leq C\sqrt{(\log p)/n}\|\mathbf{B}_0^T\mathbf{B}_0\|_2 \\
&= C\sqrt{(\log p)/n}\|\mathrm{diag}(\mathbf{s}) - \mathrm{diag}(\mathbf{s})\boldsymbol{\Omega}_0\mathrm{diag}(\mathbf{s})\|_2 \\
&\leq C\sqrt{(\log p)/n},
\end{aligned}$$

which establishes (A.18) and thus concludes the proof for the first result.

The second inequality follows from

$$\sup_{|\mathcal{S}|\leq K_n}\left\|n^{-1}\mathbf{B}_{0,\mathcal{S}}^T\mathbf{Z}^T\mathbf{Z}\mathbf{B}_{0,\mathcal{S}} - \mathbf{B}_{0,\mathcal{S}}^T\mathbf{B}_{0,\mathcal{S}}\right\|_{\max} \leq \left\|n^{-1}\mathbf{B}_0^T\mathbf{Z}^T\mathbf{Z}\mathbf{B}_0 - \mathbf{B}_0^T\mathbf{B}_0\right\|_{\max}$$

and Lemma A.3 in [6], which completes the proof of Lemma 8.