Supplementary Material for "High Dimensional Inference in Partially Linear Models"

Ying Zhu[*] Zhuqing Yu[†] Guang Cheng[‡]

February 25, 2019

A Proofs, assumptions, and additional Results

As a general rule for this appendix, all the $c \in (0, \infty)$ constants denote positive universal constants. The specific values of these constants may change from place to place. For notational simplicity, we assume the regime of interest is $p \geq (n \vee 2)$; the modification to allow $p < (n \vee 2)$ is trivial.

When constructing an approximate inverse $\hat{\Theta}$ of $n^{-1} \sum_{i=1}^{n} \hat{X}_i^T \hat{X}_i$ in (5) and (6), we adopt the nodewise regression method proposed by van de Geer, et al. (2014). Since our analysis involves establishing $\left\| \hat{\Theta}_j - \Theta_j \right\|_1 = o_p(1)$, as in van de Geer et al. (2014), our illustration (Theorem 1 and Corollary 1) in the main paper requires a sparsity condition on the inverse $\Theta = \Sigma^{-1}$ of the population Hessian $\Sigma := \mathbb{E}\left(\tilde{X}_i^T \tilde{X}_i\right)$. The realism of most assumptions for Theorem 1 and Corollary 1 in Section A.1 hinges crucially on whether the exact sparsity of β_0 and Θ hold in the applications of interest. We relax the exact sparsity of Θ_j (assumed in most literature including van de Geer, et al., 2014) in Section A.2 to accommodate for approximate sparsity which permits all the entries in Θ_j to be non-zero as long as they decay sufficiently fast. This extension provides a more realistic interpretation of most practical problems.

A.1 Assumptions for Theorem 1 and Corollary 1

Assumption 1. *For $j = 1, ..., p$, $\left(X_{ij}, \tilde{X}_{ij}, \tilde{Y}_i, \varepsilon_i\right)_{i=1}^{n}$ are i.i.d. sub-Gaussian variables with parameters at most $O(1)$.*

Assumption 2. *(i) For $\Sigma = \mathbb{E}\left(\tilde{X}_i^T \tilde{X}_i\right)$, $\left[\Lambda_{\min}^2(\Sigma)\right]^{-1} = O(1)$ and $\Lambda_{\max}^2(\Sigma) = O(1)$; there exists a parameter ρ such that for any unit vector $a \in \mathbb{R}^p$, $\sup_{r \geq 1} r^{-\frac{1}{2}} \left(\mathbb{E}\left|a^T \tilde{X}_i^T\right|^r\right)^{\frac{1}{r}} \leq \rho$ for all $i = 1, ..., n$, and*

$$\left(\rho^4 \vee 1\right) \frac{(s_j \vee 1) \log p}{n} = O(1). \tag{1}$$

(ii) For $\Sigma_1 = \mathbb{E}\left[\mathbb{E}(X_i|Z_i)^T \mathbb{E}(X_i|Z_i)\right]$ (where $\mathbb{E}(X_i|Z_i) := (\mathbb{E}(X_{i1}|Z_i), ..., \mathbb{E}(X_{ip}|Z_i))$ is a row vector of dimension p), $\Lambda_{\max}^2(\Sigma_1) = O(1)$.

[*]Department of Statistics, Department of Computer Science. Purdue University. West Lafayette, IN
[†]AbbVie Inc., North Chicago, IL.
[‡]Department of Statistics. Purdue University. West Lafayette, IN

Remark. Part (ii) of Assumption 2 is only used in the analysis for the second debiased estimator $\tilde{b}$ in (6).

Assumption 3. *The conditional expectations* $\mathbb{E}(Y_i|Z_i)$ *and* $\mathbb{E}(X_{ij}|Z_i)$ $(j = 1, ..., p)$ *belong to* $\mathcal{F}$. *For any* $f \in \bar{\mathcal{F}} = \left\{\tilde{f} = f^{'} - f^{''} : f^{'}, f^{''} \in \mathcal{F}\right\}$ *and* $\alpha \in [0, 1]$, $\alpha f \in \bar{\mathcal{F}}$ *(that is,* $\bar{\mathcal{F}}$ *is star-shaped).*

Remark. This condition is needed to operationalize (5)-(6) which require estimators for the conditional expectations. In view of the following identity

$$g_0(Z_i) = \mathbb{E}(Y_i|Z_i) - \mathbb{E}(X_i|Z_i)\beta_0,$$

note that imposing conditions on the function class $\mathbb{E}(X_{ij}|Z_i)$s and $\mathbb{E}(Y_i|Z_i)$ belong to automatically restricts the function class g_0 belongs to. Assumption 3 is relatively mild and often seen in the literature of nonparametric statistics (see e.g., Wainwright, 2015); it is satisfied when the set $\bar{\mathcal{F}}$ is convex and includes the function $f = 0$. It is also satisfied by some non-convex sets of functions, which arise in sparse regression models.

Assumption 4. *(i) The initial estimator* $\hat{\beta}$ *satisfies that*

$$\begin{aligned} \left\|\hat{\beta} - \beta_0\right\|_2 &= O_p\left(\sqrt{\frac{s_0 \log p}{n}}\right), \\ \left\|\hat{\beta} - \beta_0\right\|_1 &= O_p\left(s_0\sqrt{\frac{\log p}{n}}\right). \end{aligned}$$

(ii) $\left\|\mathbb{E}(X|Z)\left(\hat{\beta} - \beta_0\right)\right\|_n = O_p\left(\sqrt{\frac{s_0 \log p}{n}}\right)$ *and the estimator* $\hat{g}$ *satisfies that*

$$\|\hat{g} - g\|_n^2 = O_p\left(\left(s_0 r_n^2\right) \vee \left(\frac{s_0 \log p}{n}\right)\right)$$

where r_n *is the critical radius.*

Remark. Part (ii) of Assumption 4 is only used in the analysis for the second debiased estimator $\tilde{b}$ in (6). Under mild conditions, the rates in Assumption 4 are satisfied by the initial estimators based on Zhu (2017). For special cases where Z_i has a low dimension and g_0 belongs to the mth order Sobolev ball $\mathcal{S}^m$, the initial estimators based on Müller and van de Geer (2015) or Yu, et al. (2017) also satisfy the rate requirements in Assumption 4.

Assumption 5. $\left(\|\pi_j\|_1 \vee 1\right) r_n^2 = O\left(\sqrt{\frac{\log p}{n}}\right)$ *where* r_n *is the critical radius. Moreover,*

$$\begin{aligned} \sqrt{n} s_j (s_0 \vee 1)\left(r_n^2 \vee \frac{\log p}{n}\right) &= o(1), \\ \left[s_j^2 (s_0 \vee 1)\frac{\log p}{\sqrt{n}}\right] \vee \left[s_j^3 (s_0 \vee 1)\left(r_n^2 \vee \frac{\log p}{n}\right)\sqrt{\log p}\right] &= o(1), \\ \left[\left(s_j^2 \vee 1\right)\sqrt{\frac{\log p}{n}}\right] \vee \left[\left(s_j^3 \vee 1\right)\left(r_n^2 \vee \frac{\log p}{n}\right)\right] &= o(1). \end{aligned}$$

Remark. With some algebraic manipulations, we show that $\sqrt{n}\left(\hat{b}_j - \beta_{0j}\right) = \frac{1}{\sqrt{n}}\Theta_j\tilde{X}^T\varepsilon + REM$ and $\sqrt{n}\left(\hat{b}_j - \beta_{0j}\right) = \frac{1}{\sqrt{n}}\Theta_j\tilde{X}^T\varepsilon + REM'$, where the leading term $\frac{1}{\sqrt{n}}\Theta_j\tilde{X}^T\varepsilon$ has an asymptotic normal distribution. Assumption 5 imposes requirements on the sparsity parameters s_0 and s_js as well as the rates (reflected by r_n) for the auxiliary estimators $\hat{f}_j$s so that $REM = o(1)$ and $REM' = o(1)$.

Assumption 6. *For all $l \neq j$, $j = 1, ..., p$, $\mathbb{E}\left[\frac{1}{n}\tilde{X}_l^T\left(\tilde{X}_j - \tilde{X}_{-j}\pi_j\right)\right] = O\left(\sqrt{\frac{\log p}{n}}\right)$ as $p \to \infty$ and $n \to \infty$.*

Remark. For a general sub-Gaussian matrix, Assumption 6 is needed in order to derive the scaling for λ_j, whose choice in (9) depends on an upper bound for $\frac{1}{n}\left\|\tilde{X}_{-j}^T\left(\tilde{X}_j - \pi_j\tilde{X}_{-j}\right)\right\|_\infty$. Intuitively, Assumption 6 says that as the number of terms ($\tilde{X}_{i,-j}$) used to approximate $\tilde{X}_{ij}$ increases (that is, as $p \to \infty$), $\mathbb{E}\left[\frac{1}{n}\tilde{X}_l^T\left(\tilde{X}_j - \pi_j\tilde{X}_{-j}\right)\right] = o(1)$ provided $\frac{\log p}{n} = o(1)$. If $\mathbb{E}\left(\tilde{X}_{ij}|\tilde{X}_{i,-j}\right) = \tilde{X}_{i,-j}\pi_j$ (e.g., when $\tilde{X}_i$ is a normal vector), then $\mathbb{E}\left[\frac{1}{n}\tilde{X}_l^T\left(\tilde{X}_j - \pi_j\tilde{X}_{-j}\right)\right] = 0$ for all $l \neq j$. This special case is considered in Theorem 2.2 in van de Geer, et al. (2014).

A.2 The case of approximate sparsity

With additional efforts, the exact sparsity assumptions of β_0 and π_j can be relaxed to accommodate for approximate sparsity, provided that the ordered coefficients decay sufficiently fast. To work with approximately sparse β_0 and π_j, we introduce two thresholded subsets:

$$S_{\underline{\tau}} := \{j \in \{1, 2, ..., p\} : |\beta_{0j}| > \underline{\tau}\}, \tag{2}$$
$$S_{\underline{\tau}_j} := \left\{l \in \{1, 2, ..., p\} \setminus j : |\pi_{jl}| > \underline{\tau}_j\right\}. \tag{3}$$

Let $|S_{\underline{\tau}}| := s_0$ and $\left|S_{\underline{\tau}_j}\right| := s_j$. Note that the newly defined s_0 and s_j generalize the previous exact sparsity parameters; in Theorem 1, we simply take $\underline{\tau} = 0$ and $\underline{\tau}_j = 0$. We also introduce four terms below. These terms are used in the proof for Theorem A.1 and appear in Assumptions 2A, 4A and 5A (to be stated). The roles these assumptions play in the case of approximately sparse β_0 and π_j are similar to those Assumptions 2, 4 and 5 play in the case of exact sparsity. Let

$$D_1 = s_0\sqrt{\frac{\log p}{n}} + \left(s_0\sqrt{\frac{\log p}{n}}\left\|\beta_{0,S^c_{\underline{\tau}}}\right\|_1\right)^{\frac{1}{2}} + \left\|\beta_{0,S^c_{\underline{\tau}}}\right\|_1,$$

$$D_2 = \sqrt{\frac{s_0\log p}{n}} + \left(\sqrt{\frac{\log p}{n}}\left\|\beta_{0,S^c_{\underline{\tau}}}\right\|_1\right)^{\frac{1}{2}},$$

$$B_{1j} = s_j\sqrt{\frac{\log p}{n}} + \left(s_j\sqrt{\frac{\log p}{n}}\left\|\pi_{j,S^c_{\underline{\tau}_j}}\right\|_1\right)^{\frac{1}{2}} + \left\|\pi_{j,S^c_{\underline{\tau}_j}}\right\|_1,$$

$$B_{2j} = \sqrt{\frac{s_j\log p}{n}} + \left(\sqrt{\frac{\log p}{n}}\left\|\pi_{j,S^c_{\underline{\tau}_j}}\right\|_1\right)^{\frac{1}{2}}.$$

Assumption 2A. *The conditions in Assumption 2 hold with the only change being that (1) is replaced by*

$$\left(\rho^4 \vee 1\right) \frac{\log p}{n} \left(\left|S_{\underline{\mathcal{T}}_j}\right| \vee 1\right) = O\left(1\right), \tag{4}$$

$$\left(\rho^4 \vee 1\right) \frac{\log p}{n} \left\|\pi_{j,S^c_{\underline{\mathcal{T}}_j}}\right\|_1^2 = O\left(B_{2j}^2\right). \tag{5}$$

Assumption 4A. *With* $\underline{\mathcal{T}} = \frac{c\sqrt{\frac{\log p}{n}}}{\Lambda^2_{\min}(\Sigma)}$ *in (2) for some universal constant* $c > 0$*, (i) the initial estimator* $\hat{\beta}$ *satisfies that* $\left\|\hat{\beta} - \beta_0\right\|_2 = O_p\left(D_2\right)$ *and* $\left\|\hat{\beta} - \beta_0\right\|_1 = O_p\left(D_1\right)$*; (ii)* $\left\|\mathbb{E}\left(X|Z\right)\left(\hat{\beta} - \beta_0\right)\right\|_n = O_p\left(D_2\right)$ *and the estimator* $\hat{g}$ *satisfies that* $\left\|\hat{g} - g\right\|_n^2 = O_p\left(\left(\left\|\beta_0\right\|_1 r_n^2\right) \vee D_2^2\right)$*.*

Remark. Under mild conditions, the rates in Assumption 4A are satisfied by the initial estimators based on Zhu (2017). As Assumption 4(ii), Assumption 4A(ii) is only used in the analysis for the second debiased estimator $\tilde{b}$ in (6).

Assumption 5A. $\left(\left\|\pi_j\right\|_1 \vee 1\right) r_n^2 = O\left(\sqrt{\frac{\log p}{n}}\right)$ *and*

$$\begin{aligned}
\sqrt{n}\max\left\{\left\|\pi_j\right\|_1 D_2^2,\ B_{1j}\sqrt{\frac{\log p}{n}},\ B_{1j}D_1,\ \left\|\pi_j\right\|_1\left(\left\|\beta_0\right\|_1 \vee 1\right)\left(r_n^2 \vee \frac{\log p}{n}\right)\right\} &= o(1),\\
\sqrt{n}\max\left\{\left(\left\|\pi_j\right\|_1^2 \vee \left\|\pi_j\right\|_1\right)\frac{\log p}{n},\ \left(\left\|\pi_j\right\|_1^3 \vee \left\|\pi_j\right\|_1\right)\left(r_n^2 \vee \frac{\log p}{n}\right)\sqrt{\frac{\log p}{n}}\right\} &= o(1),\\
\sqrt{n}\max\left\{\left(\left\|\pi_j\right\|_1^2 \vee \left\|\pi_j\right\|_1\right)\sqrt{\frac{\log p}{n}}D_1,\ \left(\left\|\pi_j\right\|_1^3 \vee \left\|\pi_j\right\|_1\right)\left(r_n^2 \vee \frac{\log p}{n}\right)D_1\right\} &= o(1),\\
\max\left\{B_{1j},\ \left(\left\|\pi_j\right\|_1^2 \vee \left\|\pi_j\right\|_1\right)\sqrt{\frac{\log p}{n}},\ \left(\left\|\pi_j\right\|_1^3 \vee \left\|\pi_j\right\|_1\right)\left(r_n^2 \vee \frac{\log p}{n}\right)\right\} &= o(1).
\end{aligned}$$

In comparison with the case of exact sparsity, approximate sparsity permits all the entries in β_0 (and π_j) to be non-zero at the expense of incurring an extra approximation error $\left(s_0\sqrt{\frac{\log p}{n}}\left\|\beta_{0,S^c_{\underline{\mathcal{T}}}}\right\|_1\right)^{\frac{1}{2}} + \left\|\beta_{0,S^c_{\underline{\mathcal{T}}}}\right\|_1$ (respectively, $\left(s_j\sqrt{\frac{\log p}{n}}\left\|\pi_{j,S^c_{\underline{\mathcal{T}}_j}}\right\|_1\right)^{\frac{1}{2}} + \left\|\pi_{j,S^c_{\underline{\mathcal{T}}_j}}\right\|_1$) in the upper bound for $\left\|\hat{\beta} - \beta_0\right\|_1$ (respectively, $\left\|\hat{\pi}_j - \pi_j\right\|_1$). Relative to Assumption 5, Assumption 5A imposes additional conditions on the "small coefficients" via $\left\|\beta_{0,S^c_{\underline{\mathcal{T}}}}\right\|_1$ and $\left\|\pi_{j,S^c_{\underline{\mathcal{T}}_j}}\right\|_1$ so that the remainder terms in the asymptotic expansions of $\sqrt{n}\left(\hat{b}_j - \beta_{0j}\right)$ and $\sqrt{n}\left(\tilde{b}_j - \beta_{0j}\right)$ are dominated by the leading term $\frac{1}{\sqrt{n}}\Theta_j\tilde{X}^T\varepsilon$ (which gives Theorem 2 below). Note that for the special case where $\left\|\pi_{j,S^c_{\underline{\mathcal{T}}_j}}\right\|_1 = O\left(\left(s_j \vee 1\right)\sqrt{\frac{\log p}{n}}\right)$ and $\left\|\beta_{0,S^c_{\underline{\mathcal{T}}}}\right\|_1 = O\left(\left(s_0 \vee 1\right)\sqrt{\frac{\log p}{n}}\right)$, we have $D_1 = \left(s_0 \vee 1\right)\sqrt{\frac{\log p}{n}}$, $D_2 = \sqrt{\frac{(s_0 \vee 1)\log p}{n}}$, $B_{1j} = \left(s_j \vee 1\right)\sqrt{\frac{\log p}{n}}$, and $B_{2j} = \sqrt{\frac{(s_j \vee 1)\log p}{n}}$ in Assumptions 2A, 4A and 5A; these terms take the same forms as those used in Assumptions 2, 4 and 5.

Theorem A.1. *Under Assumptions 1, 2A, 3, 4A, 5A and 6, if we choose* $\lambda_j \asymp \sqrt{\frac{\log p}{n}}$ *uniformly in* j *in (9) and* $\underline{\tau}_j = \frac{c^{'}\sqrt{\frac{\log p}{n}}}{\Lambda^2_{\min}(\Sigma)}$ *in (3) for some universal constant* $c^{'} > 0$*, then the claims in Theorem 1 still hold.*

A.3 Proof for Theorem 1 and Corollary 1

Recall the two versions of the debiased estimators:

$$\begin{aligned} \hat{b}_j &:= \hat{\beta}_j + \frac{1}{n}\hat{\Theta}_j \hat{X}^T \left(\hat{Y} - \hat{X}\hat{\beta}\right), && (6) \\ \tilde{b}_j &:= \hat{\beta}_j + \frac{1}{n}\hat{\Theta}_j \hat{X}^T \left(Y - X\hat{\beta} - \hat{g}(Z)\right); && (7) \end{aligned}$$

$\hat{g}(Z) := \{\hat{g}(Z_i)\}_{i=1}^n$, $\hat{Y} = Y - \hat{\mathbb{E}}\left(Y|Z\right) := \left\{Y_i - \hat{\mathbb{E}}\left(Y_i|Z_i\right)\right\}_{i=1}^n$ is an estimate for $\tilde{Y} = Y - \mathbb{E}\left(Y|Z\right) := \{Y_i - \mathbb{E}\left(Y_i|Z_i\right)\}_{i=1}^n$, and for $j = 1, ..., p$, $\hat{X}_j = X_j - \hat{\mathbb{E}}\left(X_j|Z\right) := \left\{X_{ij} - \hat{\mathbb{E}}\left(X_{ij}|Z_i\right)\right\}_{i=1}^n$ is an estimate for $\tilde{X}_j = X_j - \mathbb{E}\left(X_j|Z\right) = \{X_{ij} - \mathbb{E}\left(X_{ij}|Z_i\right)\}_{i=1}^n$. We write $\hat{Y} = \tilde{Y} + \hat{Y} - \tilde{Y} = \tilde{X}\beta_0 + \varepsilon + \hat{Y} - \tilde{Y}$ and $\hat{X} = \tilde{X} + \hat{X} - \tilde{X}$, which are used in the following derivations. We show in the following that $\hat{b}_j$ and $\tilde{b}_j$ have the same asymptotic distribution.

For (6), we obtain

$$\begin{aligned} & \hat{b}_j - \beta_{0j} \\ = \; & \hat{\beta}_j - \beta_{0j} + \frac{1}{n}\hat{\Theta}_j \hat{X}^T \left(\hat{Y} - \hat{X}\hat{\beta}\right) \\ = \; & \frac{1}{n}\hat{\Theta}_j \tilde{X}^T\varepsilon - \frac{1}{n}\hat{\Theta}_j \left(\tilde{X} - \hat{X}\right)^T \varepsilon + \hat{\beta}_j - \beta_{0j} + \frac{1}{n}\hat{\Theta}_j \hat{X}^T \left[\hat{X}\beta_0 - \left(\hat{X} - \tilde{X}\right)\beta_0 + \hat{Y} - \tilde{Y} - \hat{X}\hat{\beta}\right] \\ = \; & \frac{1}{n}\Theta_j \tilde{X}^T\varepsilon + \underbrace{\frac{1}{n}\left(\hat{\Theta}_j - \Theta_j\right)\tilde{X}^T\varepsilon}_{E_0} - \underbrace{\frac{1}{n}\hat{\Theta}_j\left(\tilde{X} - \hat{X}\right)^T\varepsilon}_{E_1} + \underbrace{\left(e_j - \frac{1}{n}\hat{\Theta}_j\hat{X}^T\hat{X}\right)\left(\hat{\beta} - \beta_0\right)}_{E_2} \\ & \underbrace{-\frac{1}{n}\hat{\Theta}_j\hat{X}^T\left(\hat{X} - \tilde{X}\right)\beta_0}_{E_3} + \underbrace{\frac{1}{n}\hat{\Theta}_j\hat{X}^T\left(\hat{Y} - \tilde{Y}\right)}_{E_4}. && (8) \end{aligned}$$

For (7), we obtain

$$
\begin{aligned}
& \tilde{b}_j - \beta_{0j} \\
= \;& \hat{\beta}_j - \beta_{0j} + \frac{1}{n}\hat{\Theta}_j \hat{X}^T \left(Y - X\hat{\beta} - \hat{g}(Z)\right) \\
= \;& \frac{1}{n}\hat{\Theta}_j \tilde{X}^T \varepsilon - \frac{1}{n}\hat{\Theta}_j \left(\tilde{X} - \hat{X}\right)^T \varepsilon + \hat{\beta}_j - \beta_{0j} \\
& + \frac{1}{n}\hat{\Theta}_j \hat{X}^T \left[\hat{X}\beta_0 - \left(\hat{X} - \tilde{X}\right)\beta_0 + (X - \tilde{X})\beta_0 - \hat{X}\hat{\beta} + \left(\hat{X} - \tilde{X}\right)\hat{\beta} - (X - \tilde{X})\hat{\beta} + g_0 - \hat{g}\right] \\
= \;& \frac{1}{n}\Theta_j \tilde{X}^T \varepsilon + \underbrace{\frac{1}{n}\left(\hat{\Theta}_j - \Theta_j\right)\tilde{X}^T \varepsilon}_{E_0} - \underbrace{\frac{1}{n}\hat{\Theta}_j \left(\tilde{X} - \hat{X}\right)^T \varepsilon}_{E_1} \\
& + \underbrace{\left(e_j - \frac{1}{n}\hat{\Theta}_j \hat{X}^T \hat{X}\right)\left(\hat{\beta} - \beta_0\right)}_{E_2} + \underbrace{\frac{1}{n}\hat{\Theta}_j \hat{X}^T \left(\hat{X} - \tilde{X}\right)\left(\hat{\beta} - \beta_0\right)}_{E_3'} \\
& - \underbrace{\frac{1}{n}\hat{\Theta}_j \tilde{X}^T \mathbb{E}\left(X|Z\right)\left(\hat{\beta} - \beta_0\right)}_{E_4'} - \underbrace{\frac{1}{n}\hat{\Theta}_j \left(\hat{X} - \tilde{X}\right)^T \mathbb{E}\left(X|Z\right)\left(\hat{\beta} - \beta_0\right)}_{E_5'} - \underbrace{\frac{1}{n}\hat{\Theta}_j \hat{X}^T \left(\hat{g} - g_0\right)}_{E_6'}. \qquad (9)
\end{aligned}
$$

By elementary inequalities, we have

$$
\begin{aligned}
E_0 &\leq \left\|\hat{\Theta}_j - \Theta_j\right\|_1 \left\|\frac{1}{n}\tilde{X}^T \varepsilon\right\|_\infty, \\
E_1 &\leq \left\|\hat{\Theta}_j\right\|_1 \left\|\frac{1}{n}\left(\tilde{X} - \hat{X}\right)^T \varepsilon\right\|_\infty, \\
E_2 &\leq \left\|e_j - \frac{1}{n}\hat{\Theta}_j \hat{X}^T \hat{X}\right\|_\infty \left\|\hat{\beta} - \beta_0\right\|_1, \\
E_3 &\leq \left\|\hat{\Theta}_j\right\|_1 \left\|\frac{1}{n}\hat{X}^T \left(\hat{X} - \tilde{X}\right)\right\|_\infty \left\|\beta_0\right\|_1, \\
E_4 &\leq \left\|\hat{\Theta}_j\right\|_1 \left\|\frac{1}{n}\hat{X}^T \left(\hat{Y} - \tilde{Y}\right)\right\|_\infty,
\end{aligned}
$$

and

$$
\begin{aligned}
E_3' &\leq \left\|\hat{\Theta}_j\right\|_1 \left\|\frac{1}{n}\hat{X}^T \left(\hat{X} - \tilde{X}\right)\right\|_\infty \left\|\hat{\beta} - \beta_0\right\|_1, \\
E_4' &\leq \left\|\hat{\Theta}_j\right\|_1 \left\|\frac{1}{n}\tilde{X}^T \mathbb{E}\left(X|Z\right)\right\|_\infty \left\|\hat{\beta} - \beta_0\right\|_1, \\
E_5' &\leq \left\|\hat{\Theta}_j\right\|_1 \max_{j=1,\ldots,p} \left\|\hat{X}_j - \tilde{X}_j\right\|_n \left\|\mathbb{E}\left(X|Z\right)\left(\hat{\beta} - \beta_0\right)\right\|_n, \\
E_6' &\leq \left\|\hat{\Theta}_j\right\|_1 \left\|\frac{1}{n}\hat{X}^T \left(\hat{g} - g_0\right)\right\|_\infty.
\end{aligned}
$$

We bound $\left\|e_j - \frac{1}{n}\hat{\Theta}_j \hat{X}^T \hat{X}\right\|_\infty$ with (28) and $\left\|\hat{\Theta}_j - \Theta_j\right\|_1$ with (25), which also implies that

$$
\left\|\hat{\Theta}_j\right\|_1 = O_p\left(\max_j s_j\right) + O_p\left(\max_j s_j^2 \sqrt{\frac{\log p}{n}} + \max_j s_j^3 \left(r_{nj}^2 \vee \frac{\log p}{n}\right)\right). \qquad (10)
$$

By Assumption 4, we have

$$\begin{aligned}
\left\|\hat{\beta}-\beta_0\right\|_1 &= O_p\left(s_0\sqrt{\frac{\log p}{n}}\right), \\
\left\|\mathbb{E}\left(X|Z\right)\left(\hat{\beta}-\beta_0\right)\right\|_n &= O_p\left(\sqrt{\frac{s_0\log p}{n}}\right). \quad (11)
\end{aligned}$$

By Assumption 1, standard tail bounds for sub-Exponential variables [e.g., Vershynin (2012)] yield

$$\begin{aligned}
\left\|\frac{1}{n}\tilde{X}^T\varepsilon\right\|_\infty &= O_p\left(\sqrt{\frac{\log p}{n}}\right), \\
\left\|\frac{1}{n}\tilde{X}^T\mathbb{E}\left(X|Z\right)\right\|_\infty &= O_p\left(\sqrt{\frac{\log p}{n}}\right), \quad (12)
\end{aligned}$$

where we have used the fact that $\mathbb{E}\left(X_{ij}|Z_i\right)=X_{ij}-\tilde{X}_{ij}$ is sub-Gaussian [implied by Assumption 1 and that sub-Gaussianity is preserved by linear operations]. Note that (11)-(12) only matter to the second debiased estimator $\tilde{b}_j$ in (7).

Note that we have

$$\begin{aligned}
\left\|\hat{X}_j-\tilde{X}_j\right\|_n &= \left\|\mathbb{E}\left(X_j|Z\right)-\hat{\mathbb{E}}\left(X_j|Z\right)\right\|_n, \\
\left|\frac{1}{n}\left(\tilde{X}_j-\hat{X}_j\right)^T\varepsilon\right| &= \left|\frac{1}{n}\left[\hat{\mathbb{E}}\left(X_j|Z\right)-\mathbb{E}\left(X_j|Z\right)\right]^T\varepsilon\right|, \\
\left|\frac{1}{n}\hat{X}_j^T\left(\hat{X}_{j'}-\tilde{X}_{j'}\right)\right| &\leq \left\|\mathbb{E}\left(X_j|Z\right)-\hat{\mathbb{E}}\left(X_j|Z\right)\right\|_n\left\|\mathbb{E}\left(X_{j'}|Z\right)-\hat{\mathbb{E}}\left(X_{j'}|Z\right)\right\|_n \\
&\quad+\left|\frac{1}{n}\tilde{X}_j^T\left[\hat{\mathbb{E}}\left(X_{j'}|Z\right)-\mathbb{E}\left(X_{j'}|Z\right)\right]\right|, \\
\left|\frac{1}{n}\hat{X}_j^T\left(\hat{Y}-\tilde{Y}\right)\right| &\leq \left\|\mathbb{E}\left(X_j|Z\right)-\hat{\mathbb{E}}\left(X_j|Z\right)\right\|_n\left\|\mathbb{E}\left(Y|Z\right)-\hat{\mathbb{E}}\left(Y|Z\right)\right\|_n+\left|\frac{1}{n}\tilde{X}_j^T\left[\hat{\mathbb{E}}\left(Y|Z\right)-\mathbb{E}\left(Y|Z\right)\right]\right|, \\
\left|\frac{1}{n}\hat{X}_j^T\left(\hat{g}-g_0\right)\right| &\leq \left\|\mathbb{E}\left(X_j|Z\right)-\hat{\mathbb{E}}\left(X_j|Z\right)\right\|_n\left\|\hat{g}-g_0\right\|_n+\left|\frac{1}{n}\tilde{X}_j^T\left(\hat{g}-g_0\right)\right|.
\end{aligned}$$

We write $\hat{f}_j:=\hat{\mathbb{E}}\left(X_j|Z\right)$ and $f_j:=\mathbb{E}\left(X_j|Z\right)$ for $j=1,...,p$, as well as $\hat{f}_0:=\hat{\mathbb{E}}\left(Y|Z\right)$ and $f_0:=\mathbb{E}\left(Y|Z\right)$. Under Assumptions 1 and 3, by standard argument for nonparametric least squares estimators, for any $t_n\geq r_n$,

$$\left\|\hat{f}_j-f_j\right\|_n\leq c_1t_n \quad (13)$$

with probability at least $1-c\exp\left(-c^{'}nt_n^2\right)$. Moreover, under Assumption 4,

$$\left\|\hat{g}-g_0\right\|_n=O_p\left(\sqrt{s_0}r_n\vee\sqrt{\frac{s_0\log p}{n}}\right). \quad (14)$$

For fixed elements $\tilde{f}_j\in\mathcal{F}$ and $\tilde{g}\in\mathcal{F}$, respectively, we can view $\tilde{f}_j-f_j$ and $\tilde{g}-g_0$ as functions of Z only. Additionally, note that $\mathbb{E}\left(\varepsilon_i|Z_i\right)=0$ and $\mathbb{E}\left(\tilde{Y}_i|Z_i\right)=\mathbb{E}\left(\tilde{X}_{ij}|Z_i\right)=0$ (by construction of $\tilde{Y}_i$ and $\tilde{X}_{ij}$, $i=1,...,n$, $j=1,...,p$). The remaining argument uses results from empirical processes theory and local function complexity. In particular, under the independent sampling assumption, Lemma A5 with (13)-(14) implies that

$$\begin{aligned}
\max_{j=1,\ldots,p}\left|\frac{1}{n}\left[\mathbb{E}\left(X_j|Z\right)-\hat{\mathbb{E}}\left(X_j|Z\right)\right]^T\varepsilon\right| &= O_p\left(t_n^2\right),\\
\max_{j,j'\in\{1,\ldots,p\}}\left|\frac{1}{n}\tilde{X}_j^T\left[\hat{\mathbb{E}}\left(X_{j'}|Z\right)-\mathbb{E}\left(X_{j'}|Z\right)\right]\right| &= O_p\left(t_n^2\right),\\
\max_{j=1,\ldots,p}\left|\frac{1}{n}\tilde{X}_j^T\left[\hat{\mathbb{E}}\left(Y|Z\right)-\mathbb{E}\left(Y|Z\right)\right]\right| &= O_p\left(t_n^2\right),\\
\max_{j=1,\ldots,p}\left|\frac{1}{n}\tilde{X}_j^T\left(\hat{g}-g_0\right)\right| &= O_p\left(s_0t_n^2\right)+O_p\left(\frac{s_0\log p}{n}\right),
\end{aligned}$$

provided that $t_n \geq r_n$ and $nt_n^2 \gtrsim \log p$. In our analysis, it the suffices to choose $t_n = \left(r_n \vee \sqrt{\frac{\log p}{n}}\right)$. Consequently,

$$\begin{aligned}
\left\|\frac{1}{n}\left(\tilde{X}-\hat{X}\right)^T\varepsilon\right\|_\infty &= O_p\left(r_n^2\vee\frac{\log p}{n}\right),\\
\left\|\frac{1}{n}\hat{X}^T\left(\hat{X}-\tilde{X}\right)\right\|_\infty &= O_p\left(r_n^2\vee\frac{\log p}{n}\right), &(15)\\
\left\|\frac{1}{n}\hat{X}^T\left(\hat{Y}-\tilde{Y}\right)\right\|_\infty &= O_p\left(r_n^2\vee\frac{\log p}{n}\right), &(16)\\
\left\|\frac{1}{n}\hat{X}^T\left(\hat{g}-g_0\right)\right\|_\infty &= O_p\left(s_0r_n^2\vee\frac{s_0\log p}{n}\right). &(17)
\end{aligned}$$

Note that (16) only matters to the first debiased estimator $\hat{b}_j$ in (6) while (17) only matters to the second debiased estimator $\tilde{b}_j$ in (7).

Putting all the pieces together, under Assumption 5, we apply the CLT and obtain $\frac{\sqrt{n}\left(\hat{b}_j-\beta_{0j}\right)}{\sigma_j}\xrightarrow{D}\mathcal{N}\left(0,\,1\right)$ and $\frac{\sqrt{n}\left(\tilde{b}_j-\beta_{0j}\right)}{\sigma_j}\xrightarrow{D}\mathcal{N}\left(0,\,1\right)$, where $\sigma_j^2=\Theta_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\Theta_j^T$. Now it remains to show that

$$\begin{aligned}
&\left|\hat{\Theta}_j\frac{\hat{X}^T\hat{X}}{n}\hat{\Theta}_j^T-\Theta_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\Theta_j^T\right|\\
\leq\ &\left|\hat{\Theta}_j\left[\frac{\hat{X}^T\hat{X}}{n}-\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\right]\hat{\Theta}_j^T\right|+\left|\hat{\Theta}_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\hat{\Theta}_j^T-\Theta_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\Theta_j^T\right|\\
\leq\ &\left\|\frac{\hat{X}^T\hat{X}}{n}-\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\right\|_\infty\left\|\hat{\Theta}\right\|_1^2+\left|\hat{\Theta}_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\hat{\Theta}_j^T-\Theta_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\Theta_j^T\right|\\
=\ &O_p\left(s_j^2\left(r_n^2\vee\frac{\log p}{n}\right)+s_j^2\sqrt{\frac{\log p}{n}}+s_j^3\left(r_{nj}^2\vee\frac{\log p}{n}\right)\right) &(18)\\
+\ &O_p\left(s_j^3\frac{\log p}{n}+s_j^5\left(r_n^2\vee\frac{\log p}{n}\right)^2+\left(s_j\vee 1\right)\sqrt{\frac{\log p}{n}}+\left(s_j^2\vee 1\right)\left(r_n^2\vee\frac{\log p}{n}\right)\right) &(19)
\end{aligned}$$

where (18) follows from (10) and (34)-(35); (19) follows from (27). Thus we have shown Theorem 1.

To show Corollary 1, we adopt the following result (Lemma A1) from Lemma 1.1 of Zhang and Cheng (2016). Combining Lemma A1 with the facts established in Theorem 1, the claim in

Corollary 1 follows.

Lemma A1. Let $\zeta_j = (\zeta_{1j}, ..., \zeta_{nj})$, $j = 1, ..., p$, be *i.i.d.* sub-Gaussian random variables with zero mean and variance $\Theta_j \mathbb{E}\left(\tilde{X}_i^T \tilde{X}_i\right) \Theta_j^T$. Assume that $\frac{(\log pn)^7}{n} \leq C_1 n^{-c_1}$ for some constants c_1, $C_1 > 0$. Moreover, there exists a sequence of positive numbers $\alpha_n \to \infty$ such that $\frac{\alpha_n}{p} = o(1)$ and

$$\alpha_n (\log p)^2 \max_{j=1,...,p} \lambda_j \sqrt{s_j} = o(1).$$

Then, for any $G \subseteq \{1, ..., p\}$,

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\frac{\max_{j \in G} \zeta_{ij}}{\sqrt{n}} \leq t \right) - \mathbb{P}\left(\frac{\max_{j \in G} \epsilon_{ij}}{\sqrt{n}} \leq t \right) \right| \lesssim n^{-c'}, \, c' > 0,$$

where $\epsilon_j = (\epsilon_{j1}, ..., \epsilon_{jn})$ are *i.i.d.* normal random variables with mean 0 and variance $\Theta_j \mathbb{E}\left(\tilde{X}_i^T \tilde{X}_i\right) \Theta_j^T$.

Lemma A2. Suppose Assumptions 1, 2(i) regarding $\Lambda^2_{\min}(\Sigma)$, 3, and 6 hold. If $\left(\|\pi_j\|_1 \vee 1\right) r_n^2 = O\left(\sqrt{\frac{\log p}{n}}\right)$, $\lambda_j \gtrsim \sqrt{\frac{\log p}{n}}$, and

$$\max_j s_j \left[r_n^2 \vee \frac{\log p}{n} \right] \leq c \Lambda^2_{\min}(\Sigma) \tag{20}$$

for some sufficiently small constant $c > 0$, then,

$$\max_j \|\hat{\pi}_j - \pi_j\|_2 = O_p\left(\max_j \sqrt{s_j} \lambda_j \right), \tag{21}$$

$$\max_j \|\hat{\pi}_j - \pi_j\|_1 = O_p\left(\max_j s_j \lambda_j \right). \tag{22}$$

Proof. First, write $\eta_j := \tilde{X}_j - \tilde{X}_{-j}\pi_j$ and

$$\begin{aligned} \tilde{X}_j &= \hat{X}_j + \hat{\mathbb{E}}(X_j|Z) - \mathbb{E}(X_j|Z) \\ &= \tilde{X}_{-j}\pi_j + \eta_j = \left[\hat{X}_{-j} + \hat{\mathbb{E}}(X_{-j}|Z) - \mathbb{E}(X_{-j}|Z) \right] \pi_j + \eta_j, \end{aligned}$$

thus we have

$$\begin{aligned} \hat{X}_j &= \hat{X}_{-j}\pi_j + \left[\hat{\mathbb{E}}(X_{-j}|Z) - \mathbb{E}(X_{-j}|Z) \right] \pi_j - \left[\hat{\mathbb{E}}(X_j|Z) - \mathbb{E}(X_j|Z) \right] + \eta_j \\ &= \hat{X}_{-j}\pi_j + u_j. \end{aligned}$$

where

$$u_j := \left[\hat{\mathbb{E}}(X_{-j}|Z) - \mathbb{E}(X_{-j}|Z) \right] \pi_j - \left[\hat{\mathbb{E}}(X_j|Z) - \mathbb{E}(X_j|Z) \right] + \eta_j. \tag{23}$$

By standard argument for the Lasso, applying Lemma A6 [which shows the $\frac{1}{n}\hat{X}_{-j}^T \hat{X}_{-j}$ satisfies a lower restricted eigenvalue (LRE) condition with probability at least $1 - o(1)$] and Lemma A7 along with Assumption 6 [which implies that $\max_j \left\| \frac{1}{n} \hat{X}_{-j}^T u_j \right\|_\infty = O_p\left(\sqrt{\frac{\log p}{n}}\right)$] yields (21) and (22).

Lemma A3. Suppose Assumptions in Lemma A2 hold. Let $\lambda_j \asymp \sqrt{\frac{\log p}{n}}$ uniformly in j. Then for every $j = 1, ..., p$, we have

$$\left| \hat{\tau}_j^2 - \tau_j^2 \right| = O_p\left(\max_j (s_j \vee 1) \sqrt{\frac{\log p}{n}} + \max_j \left(s_j^2 \vee 1 \right) \left(r_n^2 \vee \frac{\log p}{n} \right) \right), \tag{24}$$

where $\tau_j^2 := \mathbb{E}\left[\left(\tilde{X}_{ij} - \tilde{X}_{i,-j}\pi_j\right)^2\right]$; moreover,

$$\begin{aligned}
\left\|\hat{\Theta}_j - \Theta_j\right\|_1 &= O_p\left(\max_j s_j^2\sqrt{\frac{\log p}{n}} + \max_j s_j^3\left(r_n^2 \vee \frac{\log p}{n}\right)\right), && (25)\\
\left\|\hat{\Theta}_j - \Theta_j\right\|_2 &= O_p\left(\max_j s_j^{\frac{3}{2}}\sqrt{\frac{\log p}{n}} + \max_j s_j^{\frac{5}{2}}\left(r_n^2 \vee \frac{\log p}{n}\right)\right), && (26)\\
\left|\hat{\Theta}_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\hat{\Theta}_j^T - \Theta_{j,j}\right| &= O_p\left(\max_j s_j^3\frac{\log p}{n} + \max_j s_j^5\left(r_n^2 \vee \frac{\log p}{n}\right)^2\right)\\
&+ O_p\left(\max_j\left(s_j \vee 1\right)\sqrt{\frac{\log p}{n}} + \max_j\left(s_j^2 \vee 1\right)\left(r_n^2 \vee \frac{\log p}{n}\right)\right), && (27)\\
\left\|\hat{\Theta}_j\frac{\hat{X}^T\hat{X}}{n} - e_j\right\|_\infty &= O_p\left(\max_j s_j^2\sqrt{\frac{\log p}{n}} + \max_j s_j^3\left(r_n^2 \vee \frac{\log p}{n}\right)\right). && (28)
\end{aligned}$$

Proof. Note that we have $\hat{\tau}_j^2 := \frac{1}{n}\left\|\hat{X}_j - \hat{X}_{-j}\hat{\pi}_j\right\|_2^2 + \lambda_j\left\|\hat{\pi}_j\right\|_1$ and

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^n\left(\hat{X}_{ij} - \hat{X}_{i,-j}\hat{\pi}_j\right)^2 - \tau_j^2\right|\\
\leq\ &\frac{1}{n}\sum_{i=1}^n\left[\hat{X}_{i,-j}\left(\pi_j - \hat{\pi}_j\right)\right]^2 + \left|\frac{2}{n}\sum_{i=1}^n\left[u_{ij}\hat{X}_{i,-j}\left(\pi_j - \hat{\pi}_j\right)\right]\right| + \left|\frac{1}{n}\sum_{i=1}^n u_{ij}^2 - \tau_j^2\right|\\
\leq\ &\frac{2}{n}\sum_{i=1}^n\left[\tilde{X}_{i,-j}\left(\pi_j - \hat{\pi}_j\right)\right]^2 + \left\|\pi_j - \hat{\pi}_j\right\|_1^2\left\|\frac{2}{n}\sum_{i=1}^n\left(\hat{X}_{i,-j} - \tilde{X}_{i,-j}\right)^T\left(\hat{X}_{i,-j} - \tilde{X}_{i,-j}\right)\right\|_\infty\\
&+\left\|\frac{2}{n}\hat{X}_{-j}^T u_j\right\|_\infty\left\|\pi_j - \hat{\pi}_j\right\|_1 + \left|\frac{1}{n}\sum_{i=1}^n\left(u_{ij}^2 - \eta_{ij}^2\right)\right| + \left|\frac{1}{n}\sum_{i=1}^n\left(\eta_{ij}^2 - \tau_j^2\right)\right|.
\end{aligned}$$

Under Assumption 2(i), applying Lemma A2 and Lemma A4 yields

$$\frac{2}{n}\sum_{i=1}^n\left[\tilde{X}_{i,-j}\left(\pi_j - \hat{\pi}_j\right)\right]^2 = O_p\left(\frac{\max_j s_j\log p}{n}\right).$$

By choosing $t_n = \left(r_n \vee \sqrt{\frac{\log p}{n}}\right)$ as in the proof for Theorem 1, (13) and (22) imply that

$$\begin{aligned}
&\left\|\pi_j - \hat{\pi}_j\right\|_1^2\left\|\frac{2}{n}\sum_{i=1}^n\left(\hat{X}_{i,-j} - \tilde{X}_{i,-j}\right)^T\left(\hat{X}_{i,-j} - \tilde{X}_{i,-j}\right)\right\|_\infty\\
=\ &O_p\left(\max_j s_j^2\frac{\log p}{n}\right)O_p\left(r_n^2 \vee \frac{\log p}{n}\right).
\end{aligned}$$

By (22) and Lemma A7 along with Assumption 6, we have

$$\left\|\frac{2}{n}\hat{X}_{-j}^T u_j\right\|_\infty\left\|\pi_j - \hat{\pi}_j\right\|_1 = O_p\left(\sqrt{\frac{\log p}{n}}\right)O_p\left(\max_j s_j\sqrt{\frac{\log p}{n}}\right).$$

For the term $\left|\frac{1}{n}\sum_{i=1}^{n}\left(u_{ij}^2-\eta_{ij}^2\right)\right|$, it suffices to show

$$\begin{aligned}
\left\|\left[\hat{\mathbb{E}}\left(X_{-j}|Z\right)-\mathbb{E}\left(X_{-j}|Z\right)\right]\pi_j\right\|_n^2 &= O_p\left(\max_j s_j^2\left(r_n^2\vee\frac{\log p}{n}\right)\right) &(29)\\
\left\|\hat{\mathbb{E}}\left(X_j|Z\right)-\mathbb{E}\left(X_j|Z\right)\right\|_n^2 &= O_p\left(r_n^2\vee\frac{\log p}{n}\right) &(30)\\
\frac{1}{n}\eta_j^T\left[\hat{\mathbb{E}}\left(X_{-j}|Z\right)-\mathbb{E}\left(X_{-j}|Z\right)\right]\pi_j &= O_p\left(\max_j s_j\left(r_n^2\vee\frac{\log p}{n}\right)\right) &(31)\\
\frac{1}{n}\eta_j^T\left[\hat{\mathbb{E}}\left(X_j|Z\right)-\mathbb{E}\left(X_j|Z\right)\right] &= O_p\left(r_n^2\vee\frac{\log p}{n}\right). &(32)
\end{aligned}$$

In the above, (29) and (30) follow from (13) (where we choose $t_n=\left(r_n\vee\sqrt{\frac{\log p}{n}}\right)$). In terms of (31) and (32), for fixed elements $\tilde{f}_{j'}$, $f_{j'}\in\mathcal{F}$, we can view $\tilde{f}_{j'}-f_{j'}$ as functions of Z only. Meanwhile, η_j is a function of $\tilde{X}$ only, so $\mathbb{E}\left(\tilde{X}_{ij}|Z_i\right)=0$ $(i=1,...,n,\ j=1,...,p)$ implies that

$$\mathbb{E}\left[\eta_{ij}\tilde{f}_{j'}(Z_i)|Z_i\right]=\tilde{f}_{j'}(Z_i)\mathbb{E}\left[\tilde{X}_{ij}|Z_i\right]-\tilde{f}_{j'}(Z_i)\mathbb{E}\left[\tilde{X}_{i,-j}|Z_i\right]\pi_j=0. \tag{33}$$

Furthermore, $\mathbb{E}\left[\eta_{ij}\tilde{f}_{j'}(Z_i)\right]=0$. As for (15), we apply Lemma A5 with (13) and obtain the (31) and (32).

In terms of $\left|\frac{1}{n}\sum_{i=1}^{n}\left(\eta_{ij}^2-\tau_j^2\right)\right|$, we note that by Assumption 1 and the definition of η_{ij}, for $j=1,...,p$, an application of standard tail bounds for sub-Exponential variables yields

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left(\eta_{ij}^2-\tau_j^2\right)\right|=O_p\left(\sqrt{\frac{\log p}{n}}\right).$$

Moreover, by (22) and the choice $\lambda_j\asymp\sqrt{\frac{\log p}{n}}$, for $j=1,...,p$, we have

$$\lambda_j\left\|\hat{\pi}_j\right\|_1=O_p\left(\max_j s_j\sqrt{\frac{\log p}{n}}\right)+O_p\left(\sqrt{\frac{\log p}{n}}\right)O\left(\max_j s_j\sqrt{\frac{\log p}{n}}\right)$$

Putting the pieces together, we have (24).

Next we show (25) and (26). Note that

$$\left\|\hat{\Theta}_j-\Theta_j\right\|_1=\left\|\frac{\hat{M}_j}{\hat{\tau}_j^2}-\frac{M_j}{\tau_j^2}\right\|_1\leq\frac{\left\|\hat{\pi}_j-\pi_j\right\|_1}{\hat{\tau}_j^2}+\left\|\pi_j\right\|_1\left(\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2}\right)$$

where the first term is $O_p\left(s_j\sqrt{\frac{\log p}{n}}\right)$ by (22) and the fact that $\hat{\tau}_j^2=\tau_j^2+o_p(1)$ while $\frac{1}{\tau_j^2}\precsim 1$ [by Assumption 2]. For the second term, we have $\left\|\pi_j\right\|_1=O(s_j)$ and

$$\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2}=O_p\left\{\left[\max_j\left(s_j\vee 1\right)\sqrt{\frac{\log p}{n}}\right]\vee\left[\max_j\left(s_j^2\vee 1\right)\left(r_n^2\vee\frac{\log p}{n}\right)\right]\right\}.$$

Therefore, we have (25). Similarly, we can obtain (26) by exploiting

$$\left\|\hat{\Theta}_j-\Theta_j\right\|_2\leq\frac{\left\|\hat{\pi}_j-\pi_j\right\|_2}{\hat{\tau}_j^2}+\left\|\pi_j\right\|_2\left(\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2}\right).$$

We now show (28). Note that

$$\left\|\hat{\Theta}_j \frac{\hat{X}^T\hat{X}}{n} - e_j\right\|_\infty \leq \left\|\hat{\Theta}_j - \Theta_j\right\|_1 \left\|\frac{\hat{X}^T\hat{X}}{n}\right\|_\infty + \|\Theta_j\|_1 \left\|\frac{\hat{X}^T\hat{X}}{n} - \mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\right\|_\infty.$$

By (15), we have

$$\begin{aligned}\left\|\frac{\hat{X}^T\hat{X}}{n} - \frac{\tilde{X}^T\tilde{X}}{n}\right\|_\infty &\leq \left\|\frac{\tilde{X}^T(\hat{X}-\tilde{X})}{n}\right\|_\infty + \left\|\frac{(\hat{X}-\tilde{X})^T\tilde{X}}{n}\right\|_\infty + \left\|\frac{(\hat{X}-\tilde{X})^T(\hat{X}-\tilde{X})}{n}\right\|_\infty \\ &= O_p\left(r_n^2 \vee \frac{\log p}{n}\right). \end{aligned} \tag{34}$$

Moreover, by standard tail bounds for sub-Exponential variables, we have

$$\left\|\frac{\tilde{X}^T\tilde{X}}{n} - \mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\right\|_\infty = O_p\left(\sqrt{\frac{\log p}{n}}\right). \tag{35}$$

Consequently, we obtain (28).

Finally we show (27). Using the facts that $\Theta_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right) = e_j^T$, $\Theta_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\Theta_j^T = \Theta_{j,j}$, $\hat{\Theta}_{j,j} = \frac{1}{\hat{\tau}_j^2}$, and $\Theta_{j,j} = \frac{1}{\tau_j^2}$, we have

$$\begin{aligned}&\hat{\Theta}_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\hat{\Theta}_j^T - \Theta_{j,j} \\ =&(\hat{\Theta}_j - \Theta_j)\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)(\hat{\Theta}_j - \Theta_j)^T + 2\Theta_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)(\hat{\Theta}_j - \Theta_j)^T + \Theta_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\Theta_j^T - \Theta_{j,j} \\ =&(\hat{\Theta}_j - \Theta_j)\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)(\hat{\Theta}_j - \Theta_j)^T + \frac{2}{\hat{\tau}_j^2} - \frac{2}{\tau_j^2}.\end{aligned}$$

Note that

$$(\hat{\Theta}_j - \Theta_j)\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)(\hat{\Theta}_j - \Theta_j)^T \leq \Lambda_{\max}^2 \left\|\hat{\Theta}_j - \Theta_j\right\|_2^2.$$

Applying (26) yields the claim. □

Lemma A4. Let $U \in \mathbb{R}^{n\times p_1}$ be a sub-Gaussian matrix with parameter σ_U and each row be independently sampled. For some positive constants, $\underline{\alpha}$, $\bar{\alpha}$, α', and b, that only depend on σ_U and the eigenvalues of $\Sigma_U = \mathbb{E}(U_i^TU_i)$ for $i = 1, ..., n$, we have

$$\begin{aligned}\frac{\|U\Delta\|_2^2}{n} &\leq \frac{3\bar{\alpha}}{2}\|\Delta\|_2^2 + \alpha'\frac{\log p_1}{n}\|\Delta\|_1^2, \qquad \text{for all } \Delta \in \mathbb{R}^{p_1} \\ \frac{\|U\Delta\|_2^2}{n} &\geq \frac{\underline{\alpha}}{2}\|\Delta\|_2^2 - \alpha'\frac{\log p_1}{n}\|\Delta\|_1^2, \qquad \text{for all } \Delta \in \mathbb{R}^{p_1}\end{aligned}$$

with probability at least $1 - c_1\exp(-bn)$.

Remark. This result follows from Lemmas 12, 13 and 15 in Loh and Wainwright (2012).

Lemma A5. Suppose $\{v_i\}_{i=1}^n$ are *i.i.d.* sub-Gaussian variables with parameter at most σ_v and $\mathbb{E}(v_i|Z_i) = 0$. Let $\bar{\mathcal{F}}$ be a star-shaped function class. Then, there are positive universal constants (c, c_1, c_2) such that for any $t_n \geq r_n$ (where r_n is the *critical radius*),

$$\sup_{f\in\bar{\mathcal{F}},\, \|f\|_n \leq t_n}\left|\frac{1}{n}\sum_{i=1}^n v_i f(Z_i)\right| \leq ct_n^2 \tag{36}$$

with probability at least $1-c_1\exp\left(-c_2 n\frac{t_n^2}{\sigma_v^2}\right)$. If $\bar{\mathcal{F}}$ concerns a ball of a RKHS $\mathcal{H}$ equipped with a norm $\|\cdot\|_{\mathcal{H}}$, then for any $f\in\bar{\mathcal{F}}$ and $t_n\geq r_n$, we have

$$\left|\frac{1}{n}\sum_{i=1}^{n}v_i f(Z_i)\right|\leq c^{'}t_n^2\left\|f\right\|_{\mathcal{H}}+c^{''}t_n\left\|f\right\|_n \tag{37}$$

with probability at least $1-c_1^{'}\exp\left(-c_2^{'}n\frac{t_n^2}{\sigma_v^2}\right)$.

Remark. The proof for bound (36) follows the proof for Lemma 13.2 in Wainwright (2015). The second bound (37) follows from Lemma 1 in Raskutti et al. (2012). If $\|f\|_{\mathcal{H}}\precsim 1$ and $\|f\|_n\precsim t_n$, then $\left|\frac{1}{n}\sum_{i=1}^{n}v_i f(Z_i)\right|\leq c_3 t_n^2$ with probability at least $1-c_1^{'}\exp\left(-c_2^{'}n\frac{t_n^2}{\sigma_v^2}\right)$. □

Lemma A6 (LRE condition). Suppose Assumptions 1, 3, and (20) hold. Then, for any

$$\Delta_j\in\mathbb{C}(J(\pi_j)):=\left\{\Delta\in\mathbb{R}^{p-1}\,:\,\left|\Delta_{J(\pi_j)^c}\right|_1\leq 3\left|\Delta_{J(\pi_j)}\right|_1\right\} \tag{38}$$

[$J(\pi_j)$ is the support of π_j] and every $j=1,...,p$, we have

$$\Delta_j^T\frac{\hat{X}_{-j}^T\hat{X}_{-j}}{n}\Delta_j\geq\kappa_1\left\|\Delta_j\right\|_2^2 \tag{39}$$

with probability at least $1-c\exp\left(-c^{'}nt_n^2+c^{''}\log p\right)$ for any $t_n\geq r_n$, where $\kappa_1=c_0^{'}\Lambda_{\min}^2(\Sigma)$ for some universal constant $c_0^{'}>0$.

Proof. By elementary inequalities, we have

$$\begin{aligned}\left|\Delta_j^T\frac{\hat{X}_{-j}^T\hat{X}_{-j}}{n}\Delta_j\right| &\geq \left|\Delta_j^T\frac{\tilde{X}_{-j}^T\tilde{X}_{-j}}{n}\Delta_j\right|-\left|\Delta_j^T\left(\frac{\tilde{X}_{-j}^T\tilde{X}_{-j}-\hat{X}_{-j}^T\hat{X}_{-j}}{n}\right)\Delta_j\right|\\ &\geq \left|\Delta_j^T\frac{\tilde{X}_{-j}^T\tilde{X}_{-j}}{n}\Delta_j\right|-\left\|\frac{\tilde{X}_{-j}^T\tilde{X}_{-j}-\hat{X}_{-j}^T\hat{X}_{-j}}{n}\right\|_\infty\left\|\Delta_j\right\|_1^2\\ &\geq \left|\Delta_j^T\frac{\tilde{X}_{-j}^T\tilde{X}_{-j}}{n}\Delta_j\right|-\left(\left\|\frac{\tilde{X}_{-j}^T(\hat{X}_{-j}-\tilde{X}_{-j})}{n}\right\|_\infty+\left\|\frac{(\hat{X}_{-j}-\tilde{X}_{-j})^T\hat{X}_{-j}}{n}\right\|_\infty\right)\left\|\Delta_j\right\|_1^2\\ &\geq \left|\Delta_j^T\frac{\tilde{X}_{-j}^T\tilde{X}_{-j}}{n}\Delta_j\right|-\left\|\frac{\tilde{X}_{-j}^T(\hat{X}_{-j}-\tilde{X}_{-j})}{n}\right\|_\infty\left\|\Delta_j\right\|_1^2\\ &\quad -\left\|\frac{(\hat{X}_{-j}-\tilde{X}_{-j})^T\tilde{X}_{-j}}{n}\right\|_\infty\left\|\Delta_j\right\|_1^2-\left\|\frac{(\hat{X}_{-j}-\tilde{X}_{-j})^T(\hat{X}_{-j}-\tilde{X}_{-j})}{n}\right\|_\infty\left\|\Delta_j\right\|_1^2.\end{aligned}$$

We first bound $\left\|\frac{\tilde{X}_{-j}^T(\hat{X}_{-j}-\tilde{X}_{-j})}{n}\right\|_\infty$. Note that in terms of the $(l,\,l^{'})$th element of the matrix $\frac{\tilde{X}_{-j}^T(\hat{X}_{-j}-\tilde{X}_{-j})}{n}$, for any $t_n\geq r_n$, we have

$$\left|\frac{1}{n}\tilde{X}_l^T(\hat{X}_{l^{'}}-\tilde{X}_{l^{'}})\right|=\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{X}_{il}\left[\hat{f}_{l^{'}}(Z_i)-f_{l^{'}}(Z_i)\right]\right|.$$

Lemma A5 and (13) imply that, for any $t_n\geq r_n$,

$$\max_{l,l'}\left|\frac{1}{n}\tilde{X}_l^T(\hat{X}_{l'}-\tilde{X}_{l'})\right|\leq c_0t_n^2 \tag{40}$$

with probability at least $1-c_1\exp\left(-c_2nt_n^2+c_3\log p\right)$. To bound the term $\left\|\frac{(\hat{X}_{-j}-\tilde{X}_{-j})^T(\hat{X}_{-j}-\tilde{X}_{-j})}{n}\right\|_\infty$, we have

$$\left\|\frac{(\hat{X}_{-j}-\tilde{X}_{-j})^T(\hat{X}_{-j}-\tilde{X}_{-j})}{n}\right\|_\infty\leq\max_{l'}\left\|\hat{f}_{l'}(Z)-f_{l'}(Z)\right\|_n^2\leq c_0^{'}t_n^2 \tag{41}$$

with probability at least $1-c_1^{'}\exp\left(-c_2^{'}nt_n^2+c_3^{'}\log p\right)$.

To bound $\left|\Delta_j^T\frac{\tilde{X}_{-j}^T\tilde{X}_{-j}}{n}\Delta_j\right|$ from below, we apply the second result in Lemma A4; since this result holds for all $\Delta_j\in\mathbb{R}^{p-1}$, we can specialize it to any Δ_j in the restricted sets specified in (38). Putting everything above together and choosing $t_n=r_n\vee\sqrt{\frac{\log p}{n}}$ yields

$$\Delta_j^T\frac{\hat{X}_{-j}^T\hat{X}_{-j}}{n}\Delta_j\geq c_1\Lambda_{\min}^2(\Sigma)\left\|\Delta_j\right\|_2^2-c_2\left(\frac{\log p}{n}\vee r_n^2\right)\left\|\Delta_j\right\|_1^2 \tag{42}$$

with probability at least $1-c\exp\left(-c^{'}nt_n^2+c^{''}\log p\right)$. As a result, the cone condition $\left\|\hat{\pi}_j-\pi_j\right\|_1\leq 4\sqrt{s_j}\left\|\hat{\pi}_j-\pi_j\right\|_2$ implied by (38) along with condition (20) yields (39).□

Lemma A7 (Upper bound on $\max_j\left\|\frac{1}{n}\hat{X}_{-j}^Tu_j\right\|_\infty$). Suppose $\left(\left\|\pi_j\right\|_1\vee 1\right)r_n^2=O\left(\sqrt{\frac{\log p}{n}}\right)$, Assumptions 1 and 3 hold. Then, we have $\max_{j=1,...,p}\left\|\frac{1}{n}\hat{X}_{-j}^Tu_j\right\|_\infty\precsim\max_{l,j}\mathbb{E}\left[\frac{1}{n}\tilde{X}_l^T\left(\tilde{X}_j-\tilde{X}_{-j}\pi_j\right)\right]+\sqrt{\frac{\log p}{n}}$ with probability at least $1-c_1^{'}\exp\left(-c_2^{'}\log p\right)-c_3^{'}\exp\left(-c_4^{'}nt_n^2+c_5^{'}\log p\right)$.

Proof. Recall the definition of u_j in (23) and

$$\hat{X}_{-j}=\tilde{X}_{-j}-\hat{\mathbb{E}}\left(X_{-j}|Z\right)+\mathbb{E}\left(X_{-j}|Z\right)$$

where $\tilde{X}_{-j}=X_{-j}-\mathbb{E}\left(X_{-j}|Z\right)$. For any $l\neq j$, after expanding $\left|\frac{1}{n}\hat{X}_l^Tu_j\right|$, we note that in order to bound $\max_j\left\|\frac{1}{n}\hat{X}_{-j}^Tu_j\right\|_\infty$, we need to bound $\max_{l'}\left\|\hat{f}_{l'}(Z)-f_{l'}(Z)\right\|_n^2$, $\max_{l,l'}\left|\frac{1}{n}\tilde{X}_l^T(\hat{X}_{l'}-\tilde{X}_{l'})\right|$ $(l\,l'=1,...,p,\ l\neq j)$, $\max_{l=1,...,p,\,l\neq j}\left|\frac{1}{n}\eta_j^T(\hat{X}_l-\tilde{X}_l)\right|$, and $\max_{l=1,...,p,\,l\neq j}\left|\frac{1}{n}\tilde{X}_l^T\eta_j\right|$. In particular, the first two terms are bounded in (41) and (40); the fourth term can be bounded by a standard sub-Exponential concentration inequality; for the third term, we evoke (33) and the argument that is used to bound (40), which yields

$$\max_{l\neq j}\left|\frac{1}{n}\eta_j^T(\hat{X}_l-\tilde{X}_l)\right|\leq c_0^{''}t_n^2$$

with probability at least $1-c_1^{''}\exp\left(-c_2^{''}nt_n^2+c_3^{''}\log p\right)$. Putting the pieces together and choosing $t_n=r_n\vee\sqrt{\frac{\log p}{n}}$, the claim in Lemma A7 follows. □

Lemma A8. Let the shifted function class $\bar{\mathcal{F}}$ be star-shaped. Then the function $t\mapsto\frac{\mathcal{G}_n(t;\mathcal{F})}{t}$ is non-increasing on $(0,\infty)$; as a result, the *critical inequality* has a smallest positive solution (the *critical radius*).

Remark. This is Lemma 13.1 in Wainwright (2015).

Lemma A9: Let $N_n(t;\,\Omega(\tilde{r}_n;\,\mathcal{F}))$ denote the $t-$covering number of the set

$$\Omega(\tilde{r}_n;\,\mathcal{F})=\left\{f\in\bar{\mathcal{F}}\,:\,|f|_n\leq\tilde{r}_n\right\}$$

in the $\mathcal{L}^2(\mathbb{P}_n)$ norm. Then the smallest positive solution (the *critical radius*) to the *critical inequality* is bounded above by any $\tilde{r}_n\in(0,\,\sigma^{\dagger}]$ such that

$$\frac{c^{'}}{\sqrt{n}}\int_{\frac{\tilde{r}_n^2}{4\sqrt{2}\sigma^{\dagger}}}^{\tilde{r}_n}\sqrt{\log N_n(t;\,\Omega(\tilde{r}_n;\,\mathcal{F}))}dt\leq\frac{\tilde{r}_n^2}{4\sigma^{\dagger}}.$$

Remark. This result has been established in van der Vaart and Wellner (1996), van de Geer (2000), Barlett and Mendelson (2002), Koltchinski (2006), Wainwright (2015), etc.

Lemma A10 (Approximately sparse π_j). Suppose Assumptions 1, 2(i) regarding $\Lambda^2_{\min}(\Sigma)$, 3, and 6 hold. If $\left(\left\|\pi_j\right\|_1\vee 1\right)r_n^2=O\left(\sqrt{\frac{\log p}{n}}\right)$, $\lambda_j\gtrsim\sqrt{\frac{\log p}{n}}$, and

$$\left\|\pi_j\right\|_1\left(\frac{\log p}{n}\vee r_n^2\right)\leq c\sqrt{\frac{\log p}{n}}\Lambda^2_{\min}(\Sigma) \tag{43}$$

for some sufficiently small constant $c>0$, then,

$$\max_j\left\|\hat{\pi}_j-\pi_j\right\|_2 \;=\; O_p\left(\lambda_j\sqrt{s_j}+\sqrt{\lambda_j\left\|\pi_{j,S^c_{\underline{\tau}_j}}\right\|_1}\right), \tag{44}$$

$$\max_j\left\|\hat{\pi}_j-\pi_j\right\|_1 \;=\; O_p\left(\lambda_j s_j+\sqrt{\lambda_j s_j\left\|\pi_{j,S^c_{\underline{\tau}_j}}\right\|_1}+\left\|\pi_{j,S^c_{\underline{\tau}_j}}\right\|_1\right). \tag{45}$$

Proof. Let $\hat{\Delta}_j=\hat{\pi}_j-\pi_j$. The basic inequality and the choice of λ_j yield that $\left\|\hat{\Delta}_{j,S^c_{\underline{\tau}_j}}\right\|_1\leq 3\left\|\hat{\Delta}_{j,S_{\underline{\tau}_j}}\right\|_1+4\left\|\pi_{j,S^c_{\underline{\tau}_j}}\right\|_1$. Consequently,

$$\left\|\hat{\Delta}_j\right\|_1\leq 4\left\|\hat{\Delta}_{j,S_{\underline{\tau}_j}}\right\|_1+4\left\|\pi_{j,S^c_{\underline{\tau}_j}}\right\|_1\leq 4\sqrt{s_j}\left\|\hat{\Delta}_j\right\|_2+4\left\|\pi_{j,S^c_{\underline{\tau}_j}}\right\|_1. \tag{46}$$

Moreover, we have

$$\sum_{l=1,\ldots,p,\,l\neq j}|\pi_{jl}|\geq\sum_{l\in S_{\underline{\tau}_j}}|\pi_{jl}|\geq\underline{\tau}_j s_j$$

and therefore $s_j\leq\underline{\tau}_j^{-1}\left\|\pi_j\right\|_1$. Putting the pieces together yields

$$\left\|\hat{\Delta}_j\right\|_1\leq 4\sqrt{\underline{\tau}_j^{-1}\left\|\pi_j\right\|_1}\left\|\hat{\Delta}_j\right\|_2+4\left\|\pi_{j,S^c_{\underline{\tau}_j}}\right\|_1. \tag{47}$$

Therefore, for any vector $\hat{\Delta}_j$ satisfying (47), applying (42) yields

$$\hat{\Delta}_j^T\frac{\hat{X}_{-j}^T\hat{X}_{-j}}{n}\hat{\Delta}_j\geq\left\|\hat{\Delta}_j\right\|_2^2\left\{c_1\Lambda^2_{\min}(\Sigma)-c_2\underline{\tau}_j^{-1}\left\|\pi_j\right\|_1\left(\frac{\log p}{n}\vee r_n^2\right)\right\}-c_3\left\|\pi_{j,S^c_{\underline{\tau}_j}}\right\|_1^2\left(\frac{\log p}{n}\vee r_n^2\right). \tag{48}$$

With the choice of

$$\left(\Lambda_{\min}^2(\Sigma)\right)^{-\frac{1}{2}}\left\|\pi_{j,S^c_{\underline{T}_j}}\right\|_1\sqrt{\frac{\log p}{n}\vee r_n^2}\asymp\delta^*,$$

under condition (43), we have

$$\left|\hat{\Delta}_j^T\frac{\hat{X}_{-j}^T\hat{X}_{-j}}{n}\hat{\Delta}_j\right|\geq c_4\Lambda_{\min}^2(\Sigma)\left\|\hat{\Delta}_j\right\|_2^2$$

for any $\hat{\Delta}_j$ such that $\left\|\hat{\Delta}_j\right\|_2\geq\delta^*$. We can then apply Theorem 1 in Negahban, et. al (2010) to obtain (44). By (46), we also obtain (45). □

A.4 Proof for Theorem A.1

For Theorem A.1, we apply (44) and (45) when proving Lemma A3 and obtain

$$\begin{aligned}
\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2} &= O_p\left(\max_j B_{2j}^2\right)+O_p\left(\max_j\left(\|\pi_j\|_1\vee 1\right)\sqrt{\frac{\log p}{n}}\right)\\
&+ O_p\left(\max_j\left(\|\pi_j\|_1^2\vee 1\right)\left(r_n^2\vee\frac{\log p}{n}\right)\right),\\
\left\|\hat{\Theta}_j-\Theta_j\right\|_1 &= O_p\left(\max_j B_{1j}\right)+O_p\left(\max_j\left(\|\pi_j\|_1^2\vee\|\pi_j\|_1\right)\sqrt{\frac{\log p}{n}}\right)\\
&+ O_p\left(\max_j\left(\|\pi_j\|_1^3\vee\|\pi_j\|_1\right)\left(r_n^2\vee\frac{\log p}{n}\right)\right),\\
\left\|\hat{\Theta}_j-\Theta_j\right\|_2 &= O_p\left(\max_j B_{2j}\right)+O_p\left(\max_j\left(\|\pi_j\|_1\|\pi_j\|_2\vee\|\pi_j\|_2\right)\sqrt{\frac{\log p}{n}}\right)\\
&+ O_p\left(\max_j\left(\|\pi_j\|_1^2\|\pi_j\|_2\vee\|\pi_j\|_2\right)\left(r_n^2\vee\frac{\log p}{n}\right)\right),\\
\left|\hat{\Theta}_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\hat{\Theta}_j^T-\Theta_{j,j}\right| &= O_p\left(\max_j B_{2j}^2\right)+O_p\left(\max_j\left(\|\pi_j\|_1^2\|\pi_j\|_2^2\vee\|\pi_j\|_2^2\right)\frac{\log p}{n}\right)\\
&+ O_p\left(\max_j\left(\|\pi_j\|_1^4\|\pi_j\|_2^2\vee\|\pi_j\|_2^2\right)\left(r_n^2\vee\frac{\log p}{n}\right)^2\right)\\
&+ O_p\left\{\max_j\left(\|\pi_j\|_1\vee 1\right)\sqrt{\frac{\log p}{n}}+\max_j\left(\|\pi_j\|_1^2\vee 1\right)\left(r_n^2\vee\frac{\log p}{n}\right)\right\},\\
\left\|\hat{\Theta}_j\frac{\hat{X}^T\hat{X}}{n}-e_j\right\|_\infty &= O_p\left(\max_j B_{1j}\right)+O_p\left(\max_j\left(\|\pi_j\|_1^2\vee\|\pi_j\|_1\right)\sqrt{\frac{\log p}{n}}\right)\\
&+ O_p\left(\max_j\left(\|\pi_j\|_1^3\vee\|\pi_j\|_1\right)\left(r_n^2\vee\frac{\log p}{n}\right)\right).
\end{aligned}$$

Now, we adopt the same argument as in the proof for Theorem 1 with the following minor differences: in showing the remainder terms E_0-E_4 and $E_3^{'}-E_6^{'}$ are $o_p\left(\frac{1}{\sqrt{n}}\right)$ as well as that $\left|\hat{\Theta}_j\frac{\hat{X}^T\hat{X}}{n}\hat{\Theta}_j^T-\Theta_j\mathbb{E}\left(\tilde{X}_i^T\tilde{X}_i\right)\Theta_j^T\right|=o_p(1)$, we apply the above rates and Assumptions 4A-5A (which replace Assumptions 4-5). Putting these pieces together gives the claims in Theorem A.1. □