Hypothesis Testing in Large-scale Functional Linear Regression

Kaijie Xue and Fang Yao

Nankai University and Peking University

Supplementary Material

S1 Notations

For clarity, we first recall the notations used throughout the supplementary material. For any vector $u = (u_1, \ldots, u_s)'$, let $||u||_q = (\sum_{l=1}^{s} |u_l|^q)^{1/q}$ for $q \geq 1$, $||u||_\infty = \max_{l \leq s} |u_l|$, and $||u||_0$ denotes the cardinality of supp(u), where supp$(u) = \{l : u_l \neq 0\}$. If $S =$ supp(u), let u_S be the vector containing only nonzero elements of u, while u_{S^c} contains only zero elements. For any matrix $C = [c_{ij}]$, denote $||C||_\infty = \max_{i,j} |c_{ij}|$, and if C is symmetric, $\lambda_{\min}(C)$ and $\lambda_{\max}(C)$ are the minimum and maximum eigenvalues. For real sequences a_n and b_n, $a_n \sim b_n$ if and only if $c_1 \leq \lim_{n\to\infty} |a_n/b_n| \leq c_2$ for some positive constants c_1, c_2. Recall that $||X||_{\phi_1} = \sup_{q\geq 1} q^{-1/2}\{E(|X|^q)\}^{1/q}$ is the sub-Gaussian norm, and $||X||_{\phi_2} = \sup_{q\geq 1} q^{-1}\{E(|X|^q)\}^{1/q}$ the sub-exponential norm, similar to those adopted in Ning and Liu (2017).

Recall that $\mathcal{P}_n = \{1, \ldots, p_n\}$ is the index set representing all functional predictors, while $\mathcal{H}_n \subseteq \mathcal{P}_n$ is any nonempty subset of $\mathcal{P}_n$, with cardinality $|\mathcal{H}_n| = h_n \leq p_n$, and denote $\mathcal{H}_n^c = \mathcal{P}_n \setminus \mathcal{H}_n$. For each $j \leq p_n$, let $\Lambda_j = \text{diag}\{\omega_{j1}^{1/2}, \ldots, \omega_{js_n}^{1/2}\}$, $\Lambda = \Lambda_{\mathcal{P}_n}$ be the block diagonal matrix with $\{\Lambda_j : j \leq p_n\}$ as diagonal submatrices, and $\Lambda_{\mathcal{H}_n}$ be the block diagonal matrix with $\{\Lambda_j : j \in \mathcal{H}_n\}$ as diagonal submatrices. Similarly, define $\hat{\Lambda}_j = \text{diag}\{\hat{\omega}_{j1}^{1/2}, \ldots, \hat{\omega}_{js_n}^{1/2}\}$, $\hat{\Lambda} = \hat{\Lambda}_{\mathcal{P}_n}$ and $\hat{\Lambda}_{\mathcal{H}_n}$. Furthermore, we denote $\breve{\eta} = \Lambda\eta = (\breve{\eta}_1^{'}, \ldots, \breve{\eta}_{p_n}^{'})'$ with $\breve{\eta}_j = \Lambda_j \eta_j = (\eta_{j1}\omega_{j1}^{1/2}, \ldots, \eta_{js_n}\omega_{js_n}^{1/2})'$. Similarly, denote $\nu = \eta - \eta^*$ and $\breve{\nu} = \breve{\eta} - \breve{\eta}^* = \Lambda\nu$, where η^* and $\breve{\eta}^*$ are true values of η and $\breve{\eta}$, respectively. For any $\mathcal{H}_n \subseteq \mathcal{P}_n$, $\breve{\nu}_{\mathcal{H}_n}$ is constructed by stacking the vectors $\{\breve{\nu}_j = \Lambda_j(\eta_j - \eta_j^*) : j \in \mathcal{H}_n\}$ in a column.

Without loss of generality, let $\Theta = (G_1, \ldots, G_n)' = [\Theta_{\mathcal{H}_n}, \Theta_{\mathcal{H}_n^c}]$, $\Theta_{\mathcal{H}_n} = (E_1, \ldots, E_n)'$, $\Theta_{\mathcal{H}_n^c} = (F_1, \ldots, F_n)'$, where $\Theta_{\mathcal{H}_n}$ is constructed by stacking $\{\Theta_j : j \in \mathcal{H}_n\}$ in a row. Similarly, denote $\breve{\Theta} = (\breve{G}_1, \ldots, \breve{G}_n)' = [\breve{\Theta}_{\mathcal{H}_n}, \breve{\Theta}_{\mathcal{H}_n^c}] = \Theta\Lambda^{-1} = [\Theta_{\mathcal{H}_n}\Lambda_{\mathcal{H}_n}^{-1}, \Theta_{\mathcal{H}_n^c}\Lambda_{\mathcal{H}_n^c}^{-1}]$, $\breve{\Theta}_{\mathcal{H}_n} = (\breve{E}_1, \ldots, \breve{E}_n)'$, $\breve{\Theta}_{\mathcal{H}_n^c} = (\breve{F}_1, \ldots, \breve{F}_n)'$, where $\breve{\Theta}_{\mathcal{H}_n}$ is constructed by stacking $\{\breve{\Theta}_j = \Theta_j\Lambda_j^{-1} : j \in \mathcal{H}_n\}$ in a row. Further, we denote a series of quantities derived from the information matrix I.

$$I = E(G_iG_i{}'), \quad I_{\mathcal{H}_n\mathcal{H}_n} = E(E_iE_i{}'), \quad I_{\mathcal{H}_n\mathcal{H}_n^c} = E(E_iF_i{}'), \quad I_{\mathcal{H}_n^c\mathcal{H}_n} = E(F_iE_i{}'),$$

$$I_{\mathcal{H}_n^c\mathcal{H}_n^c} = E(F_iF_i{}'), \quad I_{\mathcal{H}_n|\mathcal{H}_n^c} = I_{\mathcal{H}_n\mathcal{H}_n} - w'I_{\mathcal{H}_n^c\mathcal{H}_n} \quad w = I^{-1}_{\mathcal{H}_n^c\mathcal{H}_n^c}I_{\mathcal{H}_n^c\mathcal{H}_n} = (w_1, \ldots, w_{h_ns_n}),$$

$$\breve{I} = E(\breve{G}_i\breve{G}_i') = \Lambda^{-1}I\Lambda^{-1}, \quad \breve{I}_{\mathcal{H}_n\mathcal{H}_n} = \Lambda^{-1}_{\mathcal{H}_n}I_{\mathcal{H}_n\mathcal{H}_n}\Lambda^{-1}_{\mathcal{H}_n}, \quad \breve{I}_{\mathcal{H}_n\mathcal{H}_n^c} = \Lambda^{-1}_{\mathcal{H}_n}I_{\mathcal{H}_n\mathcal{H}_n^c}\Lambda^{-1}_{\mathcal{H}_n^c},$$

$$\breve{I}_{\mathcal{H}_n^c\mathcal{H}_n} = \Lambda^{-1}_{\mathcal{H}_n^c}I_{\mathcal{H}_n^c\mathcal{H}_n}\Lambda^{-1}_{\mathcal{H}_n}, \quad \breve{I}_{\mathcal{H}_n^c\mathcal{H}_n^c} = \Lambda^{-1}_{\mathcal{H}_n^c}I_{\mathcal{H}_n^c\mathcal{H}_n^c}\Lambda^{-1}_{\mathcal{H}_n^c}, \quad \breve{I}_{\mathcal{H}_n|\mathcal{H}_n^c} = \Lambda^{-1}_{\mathcal{H}_n}I_{\mathcal{H}_n|\mathcal{H}_n^c}\Lambda^{-1}_{\mathcal{H}_n},$$

$$\rho_n = \max_{l \le h_ns_n} ||w_l||_0 = \max_{l \le h_ns_n} \rho_{nl}, \quad \rho_{nl} = ||w_l||_0.$$

Using the $\hat{\eta}$ obtained from (2.5) in the main paper under the conditions of Theorem 1, we further denote

$$\hat{S}_i = (\hat{S}_{i1}, \ldots, \hat{S}_{i,h_ns_n}) = \hat{\Lambda}^{-1}_{\mathcal{H}_n}(\hat{w}'F_i - E_i)(Y_i - F_i'\hat{\eta}_{\mathcal{H}_n^c}), \quad i = 1, \ldots, n,$$

$$S_i = (S_{i1}, \ldots, S_{i,h_ns_n}) = \Lambda^{-1}_{\mathcal{H}_n}(w'F_i - E_i)\epsilon_i, \qquad i = 1, \ldots, n,$$

$$\hat{T}^* = n^{-1/2}\sum_{i=1}^n \hat{S}_i, \quad T^* = n^{-1/2}\sum_{i=1}^n S_i, \quad \hat{T}^*_e = n^{-1/2}\sum_{i=1}^n e_i\hat{S}_i, \quad T^*_e = n^{-1/2}\sum_{i=1}^n e_iS_i,$$

$$c_B(\alpha) = \inf\{t \in \mathbb{R} : P_e(||\hat{T}^*_e||_\infty \le t) \ge 1 - \alpha\}, \quad \alpha \in (0, 1),$$

$$\phi^*(\alpha) = \inf\{t \in \mathbb{R} : P_e(||T^*_e||_\infty \le t) \ge 1 - \alpha\}, \quad \alpha \in (0, 1),$$

where $e = (e_1, \ldots, e_n)'$ consists of i.i.d. standard normal random variables independent of the data, $P_e(\cdot)$ refers to the probability measure with respect to e, $c_B(\alpha)$ denotes the $100(1-\alpha)$th percentile of $||\hat{T}^*_e||_\infty$, and $\phi^*(\alpha)$ denotes the $100(1-\alpha)$th percentile of $||T^*_e||_\infty$. It is also worth to notice that $\{S_i : i \le n\}$ are i.i.d. centered random vectors such that $E(S_iS_i{}') = \sigma^2\breve{I}_{\mathcal{H}_n|\mathcal{H}_n^c}$.

We present below a series of auxiliary lemmas, quantifying various error bounds for a number of quantities that are useful for showing the main theorems.

S2 Auxiliary Lemmas and Proofs

Lemma 1. *(a) Under conditions (P1)–(P4),* $|\rho_\lambda(t_1) - \rho_\lambda(t_2)| \leq \lambda L|t_1 - t_2|$, *for all* $t_1, t_2 \in \mathbb{R}$.

(b) Under conditions (P1)–(P4), $|\rho'_\lambda(t)| \leq \lambda L$, *for all* $t \neq 0$.

(c) Under conditions (P1)–(P5), $\lambda L|t| \leq \rho_\lambda(t) + 2^{-1}\mu t^2$, *for all* $t \in \mathbb{R}$.

Proof. The proof of this lemma can be referred to Lemma 4 in Loh and Wainwright (2015). □

Lemma 2. *Under conditions (P1)–(P4), if* $P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta) \geq 0$, *where* η^* *is the true version of* η *and* $P_{\lambda_n}(\eta) = \sum_{j=1}^{p_n} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j \eta_j||_2)$, *then*

$$0 \leq P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta) \leq \lambda_n s_n^{1/2} L\{\sum_{j\in\mathcal{A}_n} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2 - \sum_{j\in\mathcal{A}_n^c} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2\},$$

where $\mathcal{A}_n \subseteq \mathcal{P}_n$ *is the index set corresponding to the largest* q_n *elements of* $\{n^{-1/2}||\Theta_j\eta_j||_2 : j \leq p_n\}$, *and* $\mathcal{A}_n^c = \mathcal{P}_n \setminus \mathcal{A}_n$.

Proof. First, we define $f_n(t) = t/\rho_{\lambda_n s_n^{1/2}}(t)$ for $t > 0$, and $f_n(t) = (\lambda_n s_n^{1/2} L)^{-1}$ for $t = 0$. Under conditions (P1)–(P4), we have that $f_n(t)$ is nondecreasing in t

for $t \geq 0$. Then, it follows that

$$\sum_{j\in\mathcal{A}_n^c} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2) = \sum_{j\in\mathcal{A}_n^c} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2 \big\{f_n(n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2)\big\}^{-1}$$

$$\geq \big\{f_n(\max_{j\in\mathcal{A}_n^c} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2)\big\}^{-1} \sum_{j\in\mathcal{A}_n^c} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2. \qquad \text{(S2.1)}$$

Moreover, we also have

$$\sum_{j\in\mathcal{A}_n} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2) = \sum_{j\in\mathcal{A}_n} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2 \big\{f_n(n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2)\big\}^{-1}$$

$$\leq \big\{f_n(\max_{j\in\mathcal{A}_n^c} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2)\big\}^{-1} \sum_{j\in\mathcal{A}_n} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2. \qquad \text{(S2.2)}$$

By combining (S2.1), (S2.2) with $P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta) \geq 0$, we have

$$0 \leq P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta) = \sum_{j=1}^{p_n} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j\eta_j^*||_2) - \sum_{j=1}^{p_n} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j\eta_j||_2)$$

$$= \sum_{j=1}^{q_n} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j\eta_j^*||_2) - \sum_{j=1}^{p_n} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j\eta_j||_2)$$

$$\leq \sum_{j=1}^{q_n} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2) + \sum_{j=1}^{q_n} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j\eta_j||_2) - \sum_{j=1}^{p_n} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j\eta_j||_2)$$

$$= \sum_{j=1}^{q_n} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2) - \sum_{j=q_n+1}^{p_n} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j\eta_j||_2)$$

$$\leq \sum_{j\in\mathcal{A}_n} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2) - \sum_{j\in\mathcal{A}_n^c} \rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2)$$

$$\leq \big\{f_n(\max_{j\in\mathcal{A}_n^c} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2)\big\}^{-1}\big\{\sum_{j\in\mathcal{A}_n} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2 - \sum_{j\in\mathcal{A}_n^c} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2\big\}$$

$$\leq \lambda_n s_n^{1/2} L\big\{\sum_{j\in\mathcal{A}_n} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2 - \sum_{j\in\mathcal{A}_n^c} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2\big\},$$

which completes the proof. □

Lemma 3. *Under conditions (A1), (A2), (A5), (B1), we have with probability tending to one, for some constants c_1 and c_2, where I here denotes the $p_n s_n \times p_n s_n$ identity matrix,*

$$\textit{det}(\hat{\Lambda}) \neq 0, \qquad ||\hat{\Lambda}\Lambda^{-1} - I||_\infty \leq c_1 n^{\beta/2-1/2}, \qquad ||\Lambda\hat{\Lambda}^{-1} - I||_\infty \leq c_2 n^{\beta/2-1/2}.$$

Proof. First, under condition (A2) and by Bernstein inequality, we have that for any $t > 0$,

$$P\Big\{|n^{-1}\sum_{i=1}^{n}(\omega_{jk}^{-1}\theta_{ijk}^2 - 1)| \geq t\Big\} \leq 2\exp\big\{-c_0\min(c_1^{-2}t^2, c_1^{-1}t)n\big\},$$

for some universal constants $c_0, c_1 > 0$, uniformly in $j = 1, \ldots, p_n$, $k = 1, \ldots, s_n$. It then follows from union bound inequality that

$$P\Big\{\max_{j\leq p_n}\max_{k\leq s_n}|n^{-1}\sum_{i=1}^{n}(\omega_{jk}^{-1}\theta_{ijk}^2 - 1)| \geq t\Big\} \leq 2p_n s_n\exp\big\{-c_0\min(c_1^{-2}t^2, c_1^{-1}t)n\big\},$$

by choosing $t = c_2\{\log(p_n s_n)/n\}^{1/2}$, for some sufficiently large $c_2 > 0$, we have that

$$\max_{j\leq p_n}\max_{k\leq s_n}|n^{-1}\sum_{i=1}^{n}(\omega_{jk}^{-1}\theta_{ijk}^2 - 1)| \leq c_2\{\log(p_n s_n)/n\}^{1/2},$$

with probability tending to one. Accordingly, we have that

$$\begin{aligned}||\hat{\Lambda}\Lambda^{-1} - I||_\infty &= \max_{j\leq p_n}\max_{k\leq s_n}|\big(n^{-1}\sum_{i=1}^{n}\omega_{jk}^{-1}\theta_{ijk}^2\big)^{1/2} - 1| \\ &\leq \max_{j\leq p_n}\max_{k\leq s_n}|n^{-1}\sum_{i=1}^{n}(\omega_{jk}^{-1}\theta_{ijk}^2 - 1)| \leq c_2\{\log(p_n s_n)/n\}^{1/2},\end{aligned}$$

with probability tending to one. Under (A5) and (B1), we have that $||\hat{\Lambda}\Lambda^{-1} - I||_\infty \leq c_3 n^{\beta/2-1/2}$, with probability tending to one, for some constant $c_3 > 0$. Together with (A1) and (B1), it can be deduced that

$$\begin{aligned}
&\lambda_{\min}(\hat{\Lambda}) \geq \lambda_{\min}(\Lambda) - |\lambda_{\min}(\hat{\Lambda}) - \lambda_{\min}(\Lambda)| \geq \lambda_{\min}(\Lambda) - ||\hat{\Lambda} - \Lambda||_\infty \\
\geq &\lambda_{\min}(\Lambda) - \lambda_{\max}(\Lambda)||\hat{\Lambda}\Lambda^{-1} - I||_\infty \geq c_4 s_n^{-a/2} - c_5 n^{\beta/2-1/2} \\
\geq &c_6 s_n^{-a/2},
\end{aligned}$$

with probability tending to one, for some constants $c_4, c_5, c_6 > 0$. It follows that $\det(\hat{\Lambda}) \neq 0$, with probability tending to one. Moreover, we have

$$\begin{aligned}
||\Lambda\hat{\Lambda}^{-1} - I||_\infty &= \max_{j \leq p_n} \max_{k \leq s_n} |\omega_{jk}^{1/2} \big(n^{-1} \sum_{i=1}^{n} \theta_{ijk}^2\big)^{-1/2} - 1| \\
&\leq \big\{ \max_{j \leq p_n} \max_{k \leq s_n} \omega_{jk}^{1/2} \big(n^{-1} \sum_{i=1}^{n} \theta_{ijk}^2\big)^{-1/2} \big\} ||\hat{\Lambda}\Lambda^{-1} - I||_\infty \\
&\leq (||\Lambda\hat{\Lambda}^{-1} - I||_\infty + 1)||\hat{\Lambda}\Lambda^{-1} - I||_\infty,
\end{aligned}$$

which further implies that

$$\begin{aligned}
||\Lambda\hat{\Lambda}^{-1} - I||_\infty &\leq ||\hat{\Lambda}\Lambda^{-1} - I||_\infty / (1 - ||\hat{\Lambda}\Lambda^{-1} - I||_\infty) \\
&\leq c_3 n^{\beta/2-1/2} / (1 - c_3 n^{\beta/2-1/2}) \\
&\leq 2c_3 n^{\beta/2-1/2},
\end{aligned}$$

with probability tending to one, which completes the proof. □

Lemma 4. *Under conditions (A1), (A2), (A5), (B1),we have the following with* *probability tending to one, for some positive constants* c_0, c_1 *and* c_3*, where* ϵ *denotes the* $n \times 1$ *random error vector.*

1) $||n^{-1}\Theta'\Theta - E(G_iG_i{}')||_\infty \leq c_0 n^{\beta/2-1/2}$, $||n^{-1}\breve{\Theta}'\breve{\Theta} - E(\breve{G}_i\breve{G}_i')||_\infty \leq c_1 n^{\beta/2-1/2}$,

2) $||n^{-1}\Theta'\epsilon||_\infty \leq c_2 n^{\beta/2-1/2}$, $||n^{-1}\breve{\Theta}'\epsilon||_\infty \leq c_3 n^{\beta/2-1/2}$,

3) $\max_{l \leq h_n s_n} ||n^{-1}\sum_{i=1}^n (w_l{}'F_i - E_{il})F_i{}'||_\infty \leq c_4 n^{\beta/2-1/2}$,

4) $\max_{l \leq h_n s_n} ||n^{-1}\sum_{i=1}^n (w_l{}'F_i - E_{il})\{E(E_{il}^2)\}^{-1/2}F_i{}'||_\infty \leq c_5 n^{\beta/2-1/2}$,

5) $\max_{l \leq h_n s_n} |n^{-1}\sum_{i=1}^n (w_l{}'F_i - E_{il})\epsilon_i| \leq c_6 n^{\beta/2-1/2}$,

6) $\max_{l \leq h_n s_n} |n^{-1}\sum_{i=1}^n (w_l{}'F_i - E_{il})\{E(E_{il}^2)\}^{-1/2}\epsilon_i| \leq c_7 n^{\beta/2-1/2}$.

Proof. By using Bernstein inequality and union bound inequality repeatedly, the Lemma holds trivially. □

Lemma 5. *Under conditions (A1), (A2), (A5)-(A6), (B1) and* $H_0 : \|\beta_j\|_{L^2} = 0$ *for all* $j \in \mathcal{H}_n$*, we have the following with probability tending to one, for some positive constants* c_0 *and* c_1*,*

1) $||n^{-1}\breve{\Theta}'_{\mathcal{H}_n^c}(Y - \Theta_{\mathcal{H}_n^c}\eta_{\mathcal{H}_n^c})||_\infty \leq c_0(n^{\beta/2-1/2} + q_n s_n^{-\delta})$,

2) $||n^{-1}\Theta_{\mathcal{H}_n^c}{}'(Y - \Theta_{\mathcal{H}_n^c}\eta_{\mathcal{H}_n^c})||_\infty \leq c_1(n^{\beta/2-1/2} + q_n s_n^{-\delta})$.

Proof. First, under $H_0 : \|\beta_j\|_{L^2} = 0$ for all $j \in \mathcal{H}_n$, we have that

$$\begin{aligned}
&||n^{-1}\breve{\Theta}'_{\mathcal{H}_n^c}(Y - \Theta_{\mathcal{H}_n^c}\eta_{\mathcal{H}_n^c})||_\infty = ||n^{-1}\sum_{i=1}^n \breve{F}_i(Y_i - F_i'\eta_{\mathcal{H}_n^c})||_\infty \\
=&||n^{-1}\sum_{i=1}^n \breve{F}_i\epsilon_i + n^{-1}\sum_{i=1}^n\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^\infty \big\{\breve{F}_i\theta_{ijk} - E(\breve{F}_i\theta_{ijk})\big\}\eta_{jk} + n^{-1}\sum_{i=1}^n\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^\infty E(\breve{F}_i\theta_{ijk})\eta_{jk}||_\infty \\
\le&||n^{-1}\sum_{i=1}^n \breve{F}_i\epsilon_i + n^{-1}\sum_{i=1}^n\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^\infty \big\{\breve{F}_i\theta_{ijk} - E(\breve{F}_i\theta_{ijk})\big\}\eta_{jk}||_\infty + \\
&||n^{-1}\sum_{i=1}^n\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^\infty E(\breve{F}_i\theta_{ijk})\eta_{jk}||_\infty. \qquad \text{(S2.3)}
\end{aligned}$$

In addition, for each $l = 1, \ldots, (p_n - h_n)s_n$, we have

$$E\{(n^{-1}\sum_{i=1}^n \breve{F}_{il}\epsilon_i)^2\} = n^{-1}\sigma^2 \sim n^{-1}. \qquad \text{(S2.4)}$$

Moreover, for each $l = 1, \ldots, (p_n - h_n)s_n$, we have

$$\begin{aligned}
&E\Big[n^{-1}\sum_{i=1}^n\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^\infty \big\{\breve{F}_{il}\theta_{ijk} - E(\breve{F}_{il}\theta_{ijk})\big\}\eta_{jk}\Big]^2 \\
\le&n^{-2}\sum_{i=1}^n E\Big[\sum_{j\in\mathcal{H}_n^c}|\sum_{k=s_n+1}^\infty \big\{\breve{F}_{il}\theta_{ijk} - E(\breve{F}_{il}\theta_{ijk})\big\}\eta_{jk}|\Big]^2 \\
\le&n^{-2}\sum_{i=1}^n E\Big[\sum_{j=1}^{q_n}|\sum_{k=s_n+1}^\infty \big\{\breve{F}_{il}\theta_{ijk} - E(\breve{F}_{il}\theta_{ijk})\big\}\eta_{jk}|\Big]^2 \\
\le&n^{-2}q_n\sum_{i=1}^n E\Big[\sum_{j=1}^{q_n}|\sum_{k=s_n+1}^\infty \big\{\breve{F}_{il}\theta_{ijk} - E(\breve{F}_{il}\theta_{ijk})\big\}\eta_{jk}|^2\Big] \\
\le&n^{-2}q_n\sum_{i=1}^n\sum_{j=1}^{q_n}\Big(\sum_{k=s_n+1}^\infty \eta_{jk}^2k^{2\delta}\Big)\Big\{\sum_{k_1=s_n+1}^\infty E(\breve{F}_{il}^2\theta_{ijk_1}^2)k_1^{-2\delta}\Big\} \\
=&n^{-2}q_n\sum_{i=1}^n\sum_{j=1}^{q_n}\Big(\sum_{k=s_n+1}^\infty \eta_{jk}^2k^{2\delta}\Big)\Big\{\sum_{k_1=s_n+1}^\infty E(\breve{F}_{il}^2\breve{\theta}_{ijk_1}^2)\omega_{jk_1}k_1^{-2\delta}\Big\} \\
\le&cn^{-1}q_n^2s_n^{-2\delta} = o(n^{-1}), \qquad \text{(S2.5)}
\end{aligned}$$

where the last inequality is by (A1), (A2), (A6) and (B1). Therefore, by combining (S2.4) with (S2.5), we have that for any $l = 1, \ldots, (p_n - h_n)s_n$, there exists a universal constant $c_0 > 0$ such that

$$\lim_{n\to\infty} P\big[|n^{-1}\sum_{i=1}^{n}\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\big\{\breve{F}_{il}\theta_{ijk} - E(\breve{F}_{il}\theta_{ijk})\big\}\eta_{jk}| \le c_0|n^{-1}\sum_{i=1}^{n}\breve{F}_{il}\epsilon_i|\big] = 1,$$

which implies that for any $t > 0$,

$$\begin{aligned}
&\lim_{n\to\infty} P\big[|n^{-1}\sum_{i=1}^{n}\breve{F}_{il}\epsilon_i + n^{-1}\sum_{i=1}^{n}\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\big\{\breve{F}_{il}\theta_{ijk} - E(\breve{F}_{il}\theta_{ijk})\big\}\eta_{jk}| \ge t\big] \\
\le &\lim_{n\to\infty} P\big(|n^{-1}\sum_{i=1}^{n}\breve{F}_{il}\epsilon_i| \ge t\big) + \lim_{n\to\infty} P\big[|n^{-1}\sum_{i=1}^{n}\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\big\{\breve{F}_{il}\theta_{ijk} - E(\breve{F}_{il}\theta_{ijk})\big\}\eta_{jk}| \ge t\big] \\
\le &\lim_{n\to\infty} P\big(|n^{-1}\sum_{i=1}^{n}\breve{F}_{il}\epsilon_i| \ge t\big) + \lim_{n\to\infty} P\big(|n^{-1}\sum_{i=1}^{n}\breve{F}_{il}\epsilon_i| \ge c_0^{-1}t\big) \\
\le &2\lim_{n\to\infty} P\big(|n^{-1}\sum_{i=1}^{n}\breve{F}_{il}\epsilon_i| \ge c_1 t\big),
\end{aligned} \tag{S2.6}$$

where $c_1 = \min(1, c_0^{-1}) > 0$. Moreover, under (A2) and by Bernstein inequality, we have that for any $l = 1, \ldots, (p_n - h_n)s_n$, and any $t > 0$,

$$\lim_{n\to\infty} P\big(|n^{-1}\sum_{i=1}^{n}\breve{F}_{il}\epsilon_i| \ge c_1 t\big) \le \lim_{n\to\infty} 2\exp\big\{-c_2\min(c_3^{-2}t^2, c_3^{-1}t)n\big\}, \tag{S2.7}$$

where c_2 and c_3 are some universal positive constants. By combining (S2.6) with (S2.7), we have that for any $l = 1, \ldots, (p_n - h_n)s_n$, and any $t > 0$,

$$\begin{aligned}
&\lim_{n\to\infty} P\big[|n^{-1}\sum_{i=1}^{n}\breve{F}_{il}\epsilon_i + n^{-1}\sum_{i=1}^{n}\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\big\{\breve{F}_{il}\theta_{ijk} - E(\breve{F}_{il}\theta_{ijk})\big\}\eta_{jk}| \ge t\big] \\
\le &2\lim_{n\to\infty} P\big(|n^{-1}\sum_{i=1}^{n}\breve{F}_{il}\epsilon_i| \ge c_1 t\big) \le \lim_{n\to\infty} 4\exp\big\{-c_2\min(c_3^{-2}t^2, c_3^{-1}t)n\big\},
\end{aligned}$$

invoking union bound inequality, we have that for any $t > 0$,

$$\lim_{n\to\infty} P\big[||n^{-1}\sum_{i=1}^{n}\breve{F}_i\epsilon_i + n^{-1}\sum_{i=1}^{n}\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\big\{\breve{F}_i\theta_{ijk} - E(\breve{F}_i\theta_{ijk})\big\}\eta_{jk}||_\infty \geq t\big]$$

$$\leq \lim_{n\to\infty} 4p_n s_n \exp\big\{-c_2\min(c_3^{-2}t^2, c_3^{-1}t)n\big\},$$

by choosing $t = c_4\{\log(p_n s_n)/n\}^{1/2}$, for some sufficiently large $c_4 > 0$, we have that

$$||n^{-1}\sum_{i=1}^{n}\breve{F}_i\epsilon_i + n^{-1}\sum_{i=1}^{n}\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\big\{\breve{F}_i\theta_{ijk} - E(\breve{F}_i\theta_{ijk})\big\}\eta_{jk}||_\infty$$

$$\leq c_4\{\log(p_n s_n)/n\}^{1/2} \leq c_5 n^{\beta/2-1/2} \tag{S2.8}$$

with probability tending to one, for some $c_5 > 0$. Furthermore, we have

$$||n^{-1}\sum_{i=1}^{n}\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}E(\breve{F}_i\theta_{ijk})\eta_{jk}||_\infty = \max_{l\leq(p_n-h_n)s_n}|n^{-1}\sum_{i=1}^{n}\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}E(\breve{F}_{il}\theta_{ijk})\eta_{jk}|$$

$$\leq \max_{l\leq(p_n-h_n)s_n} n^{-1}\sum_{i=1}^{n}\sum_{j\in\mathcal{H}_n^c}|\sum_{k=s_n+1}^{\infty}E(\breve{F}_{il}\theta_{ijk})\eta_{jk}|$$

$$\leq \max_{l\leq(p_n-h_n)s_n} n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{q_n}|\sum_{k=s_n+1}^{\infty}E(\breve{F}_{il}\theta_{ijk})\eta_{jk}|$$

$$\leq \max_{l\leq(p_n-h_n)s_n} n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{q_n}\Big[\sum_{k=s_n+1}^{\infty}\big\{E(\breve{F}_{il}\theta_{ijk})\big\}^2 k^{-2\delta}\Big]^{1/2}\Big(\sum_{k_1=s_n+1}^{\infty}\eta_{jk_1}^2 k_1^{2\delta}\Big)^{1/2}$$

$$\leq n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{q_n}\Big(\sum_{k=s_n+1}^{\infty}\omega_{jk}k^{-2\delta}\Big)^{1/2}\Big(\sum_{k_1=s_n+1}^{\infty}\eta_{jk_1}^2 k_1^{2\delta}\Big)^{1/2} \leq c_6 q_n s_n^{-\delta}, \tag{S2.9}$$

for some constant $c_6 > 0$, where the last inequality is by (A1) and (A6). By combining (S2.1), (S2.8) with (S2.9), we conclude that $||n^{-1}\breve{\Theta}'_{\mathcal{H}_n^c}(Y - \Theta_{\mathcal{H}_n^c}\eta_{\mathcal{H}_n^c})||_\infty \leq c_7(n^{\beta/2-1/2} + q_n s_n^{-\delta})$, with probability tending to one, for some constant $c_7 > 0$, which completes the proof of part 1). For part 2), we have that

$$
\begin{aligned}
&||n^{-1}\Theta_{\mathcal{H}_n^c}{}'(Y-\Theta_{\mathcal{H}_n^c}\eta_{\mathcal{H}_n^c})||_\infty = ||n^{-1}\Lambda_{\mathcal{H}_n^c}\breve{\Theta}'_{\mathcal{H}_n^c}(Y-\Theta_{\mathcal{H}_n^c}\eta_{\mathcal{H}_n^c})||_\infty \\
\le&\lambda_{\max}(\Lambda_{\mathcal{H}_n^c})||n^{-1}\breve{\Theta}'_{\mathcal{H}_n^c}(Y-\Theta_{\mathcal{H}_n^c}\eta_{\mathcal{H}_n^c})||_\infty \\
\le&c_8(n^{\beta/2-1/2}+q_n s_n^{-\delta}),
\end{aligned}
$$

with probability tending to one, for some constant $c_8 > 0$, which completes the proof. □

Lemma 6. *Under conditions (A1)–(A3), (A5), (B1) and (P1)–(P5), we have the following with probability tending to one, for some positive constants c_0 and c_1, where $\breve{\nu} = \Lambda\nu = \breve{\eta} - \breve{\eta}^* = \Lambda(\eta - \eta^*)$,*

1) $n^{-1}\sum_{j=1}^{p_n}||\Theta_j(\eta_j-\eta_j^*)||_2^2 \le m_1\{1+o(1)\}||\breve{\nu}||_2^2$, where the constant m_1 is defined in (A3).

2) $||\breve{\nu}||_1 \le c_0 s_n^{1/2}\sum_{j=1}^{p_n} n^{-1/2}||\Theta_j(\eta_j-\eta_j^*)||_2, \quad \lambda_n||\breve{\nu}||_1 \le c_1\big\{P_{\lambda_n}(\eta^*) + P_{\lambda_n}(\eta) + ||\breve{\nu}||_2^2\big\}$.

Proof. First, we have that

$$n^{-1}\sum_{j=1}^{p_n}||\Theta_j(\eta_j-\eta_j^*)||_2^2 = n^{-1}\sum_{j=1}^{p_n}||\breve{\Theta}_j(\breve{\eta}_j-\breve{\eta}_j^*)||_2^2 = \sum_{j=1}^{p_n}(\breve{\eta}_j-\breve{\eta}_j^*)'(n^{-1}\breve{\Theta}_j'\breve{\Theta}_j)(\breve{\eta}_j-\breve{\eta}_j^*)$$

$$\leq \lambda_{\max}(\breve{I})||\breve{\nu}||_2^2 + \sum_{j=1}^{p_n}(\breve{\eta}_j-\breve{\eta}_j^*)'\big\{n^{-1}\breve{\Theta}_j'\breve{\Theta}_j - E(n^{-1}\breve{\Theta}_j'\breve{\Theta}_j)\big\}(\breve{\eta}_j-\breve{\eta}_j^*)$$

$$\leq \lambda_{\max}(\breve{I})||\breve{\nu}||_2^2 + ||n^{-1}\breve{\Theta}'\breve{\Theta} - E(n^{-1}\breve{\Theta}'\breve{\Theta})||_\infty \sum_{j=1}^{p_n}||\breve{\eta}_j-\breve{\eta}_j^*||_1^2$$

$$\leq \big\{\lambda_{\max}(\breve{I}) + s_n||n^{-1}\breve{\Theta}'\breve{\Theta} - E(n^{-1}\breve{\Theta}'\breve{\Theta})||_\infty\big\}||\breve{\nu}||_2^2$$

$$\leq (m_1 + c_2 s_n n^{\beta/2-1/2})||\breve{\nu}||_2^2,$$

for some constant $c_2 > 0$, where the last inequality is by (A3) and Lemma 4. Since $s_n n^{\beta/2-1/2} = o(1)$, it follows that $n^{-1}\sum_{j=1}^{p_n}||\Theta_j(\eta_j-\eta_j^*)||_2^2 \leq m_1\{1 + o(1)\}||\breve{\nu}||_2^2$, with probability tending to one, which completes the proof of part 1). Moreover, we have

$$||\breve{\nu}||_1 = \sum_{j=1}^{p_n}\sum_{k=1}^{s_n}|\breve{\eta}_{jk}-\breve{\eta}_{jk}^*| \leq s_n^{1/2}\sum_{j=1}^{p_n}||\breve{\eta}_j-\breve{\eta}_j^*||_2$$

$$\leq m_0^{-1/2}s_n^{1/2}\sum_{j=1}^{p_n}\big\{\lambda_{\min}(\breve{I})\big\}^{1/2}||\breve{\eta}_j-\breve{\eta}_j^*||_2 \leq m_0^{-1/2}s_n^{1/2}\sum_{j=1}^{p_n}\big\{(\breve{\eta}_j-\breve{\eta}_j^*)'E(n^{-1}\breve{\Theta}_j'\breve{\Theta}_j)(\breve{\eta}_j-\breve{\eta}_j^*)\big\}^{1/2}$$

$$\leq m_0^{-1/2}s_n^{1/2}\sum_{j=1}^{p_n}\Big\{n^{-1}||\Theta_j(\eta_j-\eta_j^*)||_2^2 + ||E(n^{-1}\breve{\Theta}'\breve{\Theta}) - n^{-1}\breve{\Theta}'\breve{\Theta}||_\infty||\breve{\eta}_j-\breve{\eta}_j^*||_1^2\Big\}^{1/2}$$

$$\leq m_0^{-1/2}s_n^{1/2}\sum_{j=1}^{p_n}\Big\{n^{-1/2}||\Theta_j(\eta_j-\eta_j^*)||_2 + ||E(n^{-1}\breve{\Theta}'\breve{\Theta}) - n^{-1}\breve{\Theta}'\breve{\Theta}||_\infty^{1/2}||\breve{\eta}_j-\breve{\eta}_j^*||_1\Big\}$$

$$\leq \Big\{m_0^{-1/2}s_n^{1/2}\sum_{j=1}^{p_n}n^{-1/2}||\Theta_j(\eta_j-\eta_j^*)||_2\Big\} + c_3 s_n^{1/2}n^{\beta/4-1/4}||\breve{\nu}||_1,$$

for some constant $c_3 > 0$, where the last inequality is by Lemma 4. Invoking

(B1), we have that $||\breve{\nu}||_1 \leq c_4 s_n^{1/2} \sum_{j=1}^{p_n} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2$, with probability tending to one, for some constant $c_4 > 0$, which completes the proof of the first inequality of $2)$. For the second inequality of $2)$, by combining the first inequality of $2)$ with Lemma 1(c), we have

$$\begin{aligned}
&\lambda_n||\breve{\nu}||_1 \leq c_4 \lambda_n s_n^{1/2} \sum_{j=1}^{p_n} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2 \\
=& c_4 L^{-1} \sum_{j=1}^{p_n} \lambda_n s_n^{1/2} L n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2 \\
\leq& c_4 L^{-1} \sum_{j=1}^{p_n} \left\{\rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2) + 2^{-1}\mu n^{-1}||\Theta_j(\eta_j - \eta_j^*)||_2^2\right\} \\
\leq& c_4 L^{-1}\left\{P_{\lambda_n}(\eta^*) + P_{\lambda_n}(\eta) + 2^{-1}\mu n^{-1} \sum_{j=1}^{p_n} ||\Theta_j(\eta_j - \eta_j^*)||_2^2\right\}. \qquad \text{(S2.10)}
\end{aligned}$$

Finally, by combining (S2.10) with part $1)$, we have that $\lambda_n||\breve{\nu}||_1 \leq c_5\{P_{\lambda_n}(\eta^*) + P_{\lambda_n}(\eta) + ||\breve{\nu}||_2^2\}$, with probability tending to one, for some constant $c_5 > 0$, which completes the proof of part $2)$. □

Lemma 7. *Under conditions (A1)–(A3), (A5), (B1)–(B3), we have the following with probability tending to one, for some positive constant* c_0, c_1 *and* c_2*, where* $\rho_{nl} = ||w_l||_0$ *and* $\rho_n = \max_{l \leq h_n s_n} ||w_l||_0 = \max_{l \leq h_n s_n} \rho_{nl}$*,*

1) $\max_{l \leq h_n s_n} ||n^{-1} \sum_{i=1}^n F_i F_i'(\hat{w}_l - w_l)||_\infty \leq c_0 n^{\beta/2 - 1/2}$,

2) $\max_{l \leq h_n s_n} ||\hat{w}_l - w_l||_2 \leq c_1 \rho_n^{1/2} s_n^a n^{\beta/2-1/2}$, $\max_{l \leq h_n s_n} ||\hat{w}_l - w_l||_1 \leq c_2 \rho_n s_n^a n^{\beta/2-1/2}$.

Proof. First, by Lemma 4, we have

$$\max_{l\leq h_n s_n} ||n^{-1}\sum_{i=1}^{n}(w_l{}'F_i - E_{il})F_i{}'||_\infty \leq c_0 n^{\beta/2-1/2},$$

with probability tending to one, for some constant $c_0 > 0$. Moreover, under (A5), (B1) and (B3), we have $\tau_n \sim n^{\beta/2-1/2}$. By choosing $\tau_n = c_0 n^{\beta/2-1/2}$, we have

$$\max_{l\leq h_n s_n} ||n^{-1}\sum_{i=1}^{n}(w_l{}'F_i - E_{il})F_i{}'||_\infty \leq \tau_n = c_0 n^{\beta/2-1/2}. \tag{S2.11}$$

Under (7) of the main paper, we have

$$\max_{l\leq h_n s_n} ||n^{-1}\sum_{i=1}^{n}(\hat{w}_l' F_i - E_{il})F_i{}'||_\infty \leq \tau_n = c_0 n^{\beta/2-1/2}. \tag{S2.12}$$

By combining (S2.11) with (S2.12), we have

$$\begin{aligned}
&\max_{l\leq h_n s_n} ||n^{-1}\sum_{i=1}^{n} F_i F_i{}'(\hat{w}_l - w_l)||_\infty \\
\leq &\max_{l\leq h_n s_n} ||n^{-1}\sum_{i=1}^{n}(w_l{}'F_i - E_{il})F_i{}'||_\infty + \max_{l\leq h_n s_n} ||n^{-1}\sum_{i=1}^{n}(\hat{w}_l' F_i - E_{il})F_i{}'||_\infty \\
\leq &2\tau_n = 2c_0 n^{\beta/2-1/2},
\end{aligned} \tag{S2.13}$$

which completes the proof of part 1). In addition, by the definition of the Dantizig selector method, we have

$$P\Big(\bigcap_{l=1}^{h_n s_n}\{||\hat{w}_l||_1 \leq ||w_l||_1\}\Big) \to 1, \qquad \text{as} \quad n\to\infty. \tag{S2.14}$$

Denote $S_l = \{j : w_{lj} \neq 0\}$ as the support set for each vector w_l, then (S2.14) implies that

$$\begin{aligned}
&\max_{l\leq h_n s_n} ||\hat{w}_l - w_l||_1 \leq \max_{l\leq h_n s_n} ||(\hat{w}_l - w_l)_{S_l}||_1 + \max_{l\leq h_n s_n} ||(\hat{w}_l)_{S_l^c}||_1 \\
\leq &\max_{l\leq h_n s_n} ||(\hat{w}_l - w_l)_{S_l}||_1 + \max_{l\leq h_n s_n} \left\{||(w_l)_{S_l}||_1 - ||(\hat{w}_l)_{S_l}||_1\right\} \\
\leq &2\max_{l\leq h_n s_n} ||(\hat{w}_l - w_l)_{S_l}||_1 \leq 2\max_{l\leq h_n s_n} {\rho_{nl}}^{1/2}||(\hat{w}_l - w_l)_{S_l}||_2 \\
\leq &2{\rho_n}^{1/2}\max_{l\leq h_n s_n} ||\hat{w}_l - w_l||_2, \qquad \text{(S2.15)}
\end{aligned}$$

with probability tending to one. Furthermore, we have

$$\begin{aligned}
&\max_{l\leq h_n s_n} (\hat{w}_l - w_l)'(n^{-1}\sum_{i=1}^n F_i {F_i}')(\hat{w}_l - w_l) \\
\leq &\max_{l\leq h_n s_n} ||\hat{w}_l - w_l||_1 ||n^{-1}\sum_{i=1}^n F_i {F_i}'(\hat{w}_l - w_l)||_\infty \\
\leq &2c_0 n^{\beta/2-1/2}\max_{l\leq h_n s_n} ||\hat{w}_l - w_l||_1, \qquad \text{(S2.16)}
\end{aligned}$$

with probability tending to one, where the last inequality is by (S2.13). In addition, we also have

$$\begin{aligned}
&\max_{l\leq h_n s_n} (\hat{w}_l - w_l)'(n^{-1}\sum_{i=1}^n F_i {F_i}')(\hat{w}_l - w_l) \\
\geq &\max_{l\leq h_n s_n} \left\{\lambda_{\min}(I)||\hat{w}_l - w_l||_2^2 - ||\hat{w}_l - w_l||_1^2 ||E(G_i {G_i}') - n^{-1}\Theta'\Theta||_\infty\right\} \\
\geq &\lambda_{\min}(I)\Big(\max_{l\leq h_n s_n} ||\hat{w}_l - w_l||_2^2\Big) - ||E(G_i {G_i}') - n^{-1}\Theta'\Theta||_\infty\Big(\max_{l\leq h_n s_n} ||\hat{w}_l - w_l||_1^2\Big) \\
\geq &c_1 s_n^{-a}\Big(\max_{l\leq h_n s_n} ||\hat{w}_l - w_l||_2^2\Big) - c_2 n^{\beta/2-1/2}\Big(\max_{l\leq h_n s_n} ||\hat{w}_l - w_l||_1^2\Big) \\
\geq &(c_1 s_n^{-a} - c_3\rho_n n^{\beta/2-1/2})\Big(\max_{l\leq h_n s_n} ||\hat{w}_l - w_l||_2^2\Big), \qquad \text{(S2.17)}
\end{aligned}$$

with probability tending to one, where the second last inequality is by (A1), (A3) and Lemma 4, while the last inequality is by (S2.15). Combining (S2.15), (S2.16), (S2.17) with (B2) entails that

$$\max_{l \le h_n s_n} ||\hat{w}_l - w_l||_2 \le c_4 \rho_n^{1/2} s_n^a n^{\beta/2-1/2}, \tag{S2.18}$$

with probability tending to one, for some constant $c_4 > 0$. Finally, combining (S2.15) with (S2.18) entails that $\max_{l \le h_n s_n} ||\hat{w}_l - w_l||_1 \le c_5 \rho_n s_n^a n^{\beta/2-1/2}$, with probability tending to one, for some constant $c_5 > 0$, which completes the proof of part 2). □

Lemma 8. *Under conditions (A1)–(A3), (A5)–(A6), (B1)–(B3) and (P1)–(P5), we have the following with probability tending to one, for some positive constant* c_0 *and* c_1, *where* $\hat{S}_i = (\hat{S}_{i1}, \ldots, \hat{S}_{i,h_n s_n}) = \hat{\Lambda}_{\mathcal{H}_n}^{-1}(\hat{w}' F_i - E_i)(Y_i - F_i' \hat{\eta}_{\mathcal{H}_n^c})$ *and* $S_i = (S_{i1}, \ldots, S_{i,h_n s_n}) = \Lambda_{\mathcal{H}_n}^{-1}(w' F_i - E_i)\epsilon_i$,

$$\begin{aligned}\max_{l \le h_n s_n} |n^{-1/2} \sum_{i=1}^{n} (\hat{S}_{il} - S_{il})| \le & c_0 (n^{\beta-1/2} \rho_n s_n^{3a/2} + \lambda_n n^{\beta/2} q_n s_n^{a+1} + n^{\beta/2+1/2} q_n s_n^{-\delta} \log s_n \\ & + n^{\beta} \rho_n q_n s_n^{3a/2-\delta} \log s_n),\end{aligned}$$

$$\begin{aligned}\max_{l \le h_n s_n} \big\{n^{-1} \sum_{i=1}^{n} (\hat{S}_{il} - S_{il})^2\big\}^{1/2} \le & c_1 \big\{\lambda_n n^{\beta} q_n s_n^{a/2+1} + n^{\beta/2} q_n s_n^{-\delta} \log s_n + \rho_n s_n^{3a/2} n^{\beta-1/2} (\log n)^{1/2} \\ & + \lambda_n \rho_n q_n s_n^{2a+1} n^{3\beta/2-1/2} + n^{\beta-1/2} \rho_n q_n s_n^{3a/2-\delta} \log s_n\big\}.\end{aligned}$$

Proof. First, for each $l = 1, \ldots, h_n s_n$, we have

$$\hat{S}_{il} - S_{il} = \Delta_{1il} + \Delta_{2il} + \Delta_{3il} + \Delta_{4il} + \Delta_{5il} + \Delta_{6il},$$

where we denote

$$\Delta_{1il} = \Big[\big\{E(E_{il}^2)/(n^{-1}\sum_{i_1=1}^{n} E_{i_1 l}^2)\big\}^{1/2} - 1\Big]\Big[\big\{E(E_{il}^2)\big\}^{-1/2}(w_l{}'F_i - E_{il})\epsilon_i\Big],$$

$$\Delta_{2il} = \Big[\big\{E(E_{il}^2)/(n^{-1}\sum_{i_1=1}^{n} E_{i_1 l}^2)\big\}^{1/2}\Big]\Big[\big\{E(E_{il}^2)\big\}^{-1/2}(w_l{}'F_i - E_{il})F_i{}'(\eta_{\mathcal{H}_n^c} - \hat{\eta}_{\mathcal{H}_n^c})\Big],$$

$$\Delta_{3il} = \Big[\big\{E(E_{il}^2)/(n^{-1}\sum_{i_1=1}^{n} E_{i_1 l}^2)\big\}^{1/2}\Big]\Big[\big\{E(E_{il}^2)\big\}^{-1/2}(w_l{}'F_i - E_{il})(\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})\Big],$$

$$\Delta_{4il} = \Big[\big\{E(E_{il}^2)/(n^{-1}\sum_{i_1=1}^{n} E_{i_1 l}^2)\big\}^{1/2}\Big]\Big[\big\{E(E_{il}^2)\big\}^{-1/2}(\hat{w}_l - w_l)'F_i\epsilon_i\Big],$$

$$\Delta_{5il} = \Big[\big\{E(E_{il}^2)/(n^{-1}\sum_{i_1=1}^{n} E_{i_1 l}^2)\big\}^{1/2}\Big]\Big[\big\{E(E_{il}^2)\big\}^{-1/2}(\hat{w}_l - w_l)'F_iF_i{}'(\eta_{\mathcal{H}_n^c} - \hat{\eta}_{\mathcal{H}_n^c})\Big],$$

$$\Delta_{6il} = \Big[\big\{E(E_{il}^2)/(n^{-1}\sum_{i_1=1}^{n} E_{i_1 l}^2)\big\}^{1/2}\Big]\Big[\big\{E(E_{il}^2)\big\}^{-1/2}(\hat{w}_l - w_l)'F_i(\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})\Big].$$

Accordingly, one can show that

$$\begin{aligned}\max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}(\hat{S}_{il} - S_{il})| &\le \max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}\Delta_{1il}| + \max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}\Delta_{2il}|\\ &+ \max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}\Delta_{3il}| + \max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}\Delta_{4il}|\\ &+ \max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}\Delta_{5il}| + \max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}\Delta_{6il}|. \qquad \text{(S2.19)}\end{aligned}$$

For $\max_{l \leq h_n s_n} |n^{-1/2} \sum_{i=1}^n \Delta_{1il}|$, we have

$$
\begin{aligned}
&\max_{l \leq h_n s_n} |n^{-1/2} \sum_{i=1}^n \Delta_{1il}| \\
=&\max_{l \leq h_n s_n} \Big| n^{-1/2} \sum_{i=1}^n \Big[\big\{ E(E_{il}^2)/(n^{-1} \sum_{i_1=1}^n E_{i_1 l}^2) \big\}^{1/2} - 1 \Big] \Big[\big\{ E(E_{il}^2) \big\}^{-1/2} (w_l{}' F_i - E_{il}) \epsilon_i \Big] \Big| \\
\leq& ||\Lambda \hat{\Lambda}^{-1} - I||_\infty \max_{l \leq h_n s_n} |n^{-1/2} \sum_{i=1}^n (w_l{}' F_i - E_{il}) \{ E(E_{il}^2) \}^{-1/2} \epsilon_i| \\
=& n^{1/2} ||\Lambda \hat{\Lambda}^{-1} - I||_\infty \max_{l \leq h_n s_n} |n^{-1} \sum_{i=1}^n (w_l{}' F_i - E_{il}) \{ E(E_{il}^2) \}^{-1/2} \epsilon_i| \\
\leq& c_0 n^{\beta - 1/2},
\end{aligned} \tag{S2.20}
$$

with probability tending to one, for some constant $c_0 > 0$, where the last inequality is by Lemma 3 and Lemma 4.

For $\max_{l \leq h_n s_n} |n^{-1/2} \sum_{i=1}^n \Delta_{2il}|$, we have

$$
\begin{aligned}
&\max_{l \leq h_n s_n} |n^{-1/2} \sum_{i=1}^n \Delta_{2il}| \\
\leq& \big(1 + ||\Lambda \hat{\Lambda}^{-1} - I||_\infty\big) \max_{l \leq h_n s_n} \Big| n^{-1/2} \sum_{i=1}^n \Big[\big\{ E(E_{il}^2) \big\}^{-1/2} (w_l{}' F_i - E_{il}) F_i{}' (\eta_{\mathcal{H}_n^c} - \hat{\eta}_{\mathcal{H}_n^c}) \Big] \Big| \\
\leq& n^{1/2} \big(1 + ||\Lambda \hat{\Lambda}^{-1} - I||_\infty\big) ||\eta_{\mathcal{H}_n^c} - \hat{\eta}_{\mathcal{H}_n^c}||_1 \max_{l \leq h_n s_n} ||n^{-1} \sum_{i=1}^n (w_l{}' F_i - E_{il}) \{ E(E_{il}^2) \}^{-1/2} F_i{}'||_\infty \\
\leq& n^{1/2} \big(1 + ||\Lambda \hat{\Lambda}^{-1} - I||_\infty\big) ||\hat{\eta} - \eta||_1 \max_{l \leq h_n s_n} ||n^{-1} \sum_{i=1}^n (w_l{}' F_i - E_{il}) \{ E(E_{il}^2) \}^{-1/2} F_i{}'||_\infty \\
\leq& c_1 \lambda_n n^{\beta/2} q_n s_n^{a/2+1},
\end{aligned} \tag{S2.21}
$$

with probability tending to one, for some constant $c_1 > 0$, where the last inequality is by Lemma 3, Lemma 4 and Theorem 1.

For $\max_{l\le h_n s_n} |n^{-1/2}\sum_{i=1}^n \Delta_{3il}|$, we have

$$\begin{aligned}
&\max_{l\le h_n s_n} |n^{-1/2}\sum_{i=1}^n \Delta_{3il}| \\
\le&\big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\big)\max_{l\le h_n s_n}\Big|n^{-1/2}\sum_{i=1}^n\Big[\big\{E(E_{il}^2)\big\}^{-1/2}(w_l{}'F_i-E_{il})(\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^\infty\theta_{ijk}\eta_{jk})\Big]\Big| \\
\le&n^{1/2}\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)\Big[\max_{l\le h_n s_n}\max_{i\le n}\big|\big\{E(E_{il}^2)\big\}^{-1/2}(w_l{}'F_i-E_{il})\big|\Big] \\
&\cdot\Big(n^{-1}\sum_{i=1}^n\big|\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^\infty\theta_{ijk}\eta_{jk}\big|\Big). \qquad \text{(S2.22)}
\end{aligned}$$

Regarding $\max_{l\le h_n s_n}\max_{i\le n}\big|\big\{E(E_{il}^2)\big\}^{-1/2}(w_l{}'F_i-E_{il})\big|$, based on (A2), (A5) and (B1), we have

$$\begin{aligned}
&\max_{l\le h_n s_n}\max_{i\le n}\big|\big\{E(E_{il}^2)\big\}^{-1/2}(w_l{}'F_i-E_{il})\big| \\
\le&c_1\{\log(np_n s_n)\}^{1/2}\le c_2 n^{\beta/2}, \qquad \text{(S2.23)}
\end{aligned}$$

with probability tending to one, for some constants $c_1, c_2>0$.

Regarding $n^{-1}\sum_{i=1}^n\big|\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^\infty\theta_{ijk}\eta_{jk}\big|$, we have

$$\begin{aligned}
&E\Big(n^{-1}\sum_{i=1}^n\big|\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^\infty\theta_{ijk}\eta_{jk}\big|\Big)\le E\Big(n^{-1}\sum_{i=1}^n\sum_{j=1}^{q_n}\big|\sum_{k=s_n+1}^\infty\theta_{ijk}\eta_{jk}\big|\Big) \\
\le&E\Big\{n^{-1}\sum_{i=1}^n\sum_{j=1}^{q_n}\big(\sum_{k=s_n+1}^\infty\theta_{ijk}^2k^{-2\delta}\big)^{1/2}\big(\sum_{k_1=s_n+1}^\infty\eta_{jk_1}^2k_1^{2\delta}\big)^{1/2}\Big\} \\
\le&n^{-1}\sum_{i=1}^n\sum_{j=1}^{q_n}\big(\sum_{k_1=s_n+1}^\infty\eta_{jk_1}^2k_1^{2\delta}\big)^{1/2}E\big\{\big(\sum_{k=s_n+1}^\infty\theta_{ijk}^2k^{-2\delta}\big)^{1/2}\big\} \\
\le&n^{-1}\sum_{i=1}^n\sum_{j=1}^{q_n}\big(\sum_{k_1=s_n+1}^\infty\eta_{jk_1}^2k_1^{2\delta}\big)^{1/2}\big(\sum_{k=s_n+1}^\infty\omega_{jk}k^{-2\delta}\big)^{1/2} \\
\le&O(q_n s_n^{-\delta}), \qquad \text{(S2.24)}
\end{aligned}$$

where the last inequality is by (A1) and (A6). Hence, by combining (S2.22), (S2.23) with (S2.24), we have

$$\max_{l\le h_n s_n} |n^{-1/2}\sum_{i=1}^{n}\Delta_{3il}| \le O_p(n^{\beta/2+1/2}q_n s_n^{-\delta}). \tag{S2.25}$$

For $\max_{l\le h_n s_n} |n^{-1/2}\sum_{i=1}^{n}\Delta_{4il}|$, we have

$$\begin{aligned}
&\max_{l\le h_n s_n} |n^{-1/2}\sum_{i=1}^{n}\Delta_{4il}|\\
\le&\big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\big)\max_{l\le h_n s_n}\Big|n^{-1/2}\sum_{i=1}^{n}\Big[\big\{E(E_{il}^2)\big\}^{-1/2}(\hat{w}_l-w_l)'F_i\epsilon_i\Big]\Big|\\
\le&n^{1/2}\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)\Big[\max_{l\le h_n s_n}\big\{E(E_{il}^2)\big\}^{-1/2}\Big]\\
&\cdot\Big(\max_{l\le h_n s_n}||\hat{w}_l-w_l||_1\Big)\Big(||n^{-1}\Theta'\epsilon||_\infty\Big)\\
\le&n^{1/2}\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)\Big[\big\{\lambda_{\min}(\Lambda)\big\}^{-1}\Big]\\
&\cdot\Big(\max_{l\le h_n s_n}||\hat{w}_l-w_l||_1\Big)\Big(||n^{-1}\Theta'\epsilon||_\infty\Big)\\
\le&c_3 n^{\beta-1/2}\rho_n s_n^{3a/2},
\end{aligned} \tag{S2.26}$$

with probability tending to one, for some constant $c_3>0$, where the last inequality is by Lemma 3, Lemma 4, Lemma 7 and (A1).

For $\max_{l\le h_n s_n} |n^{-1/2}\sum_{i=1}^n \Delta_{5il}|$, we have

$$
\begin{aligned}
&\max_{l\le h_n s_n} |n^{-1/2}\sum_{i=1}^n \Delta_{5il}| \\
\le&\big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\big)\max_{l\le h_n s_n}\Big|n^{-1/2}\sum_{i=1}^n\Big[\big\{E(E_{il}^2)\big\}^{-1/2}(\hat{w}_l-w_l)'F_iF_i'(\eta_{\mathcal{H}_n^c}-\hat{\eta}_{\mathcal{H}_n^c})\Big]\Big| \\
\le& n^{1/2}\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)\Big[\big\{\lambda_{\min}(\Lambda)\big\}^{-1}\Big]\Big(||\hat{\eta}-\eta||_1\Big)\Big\{\max_{l\le h_n s_n}||n^{-1}\sum_{i=1}^n F_iF_i'(\hat{w}_l-w_l)||_\infty\Big\} \\
\le& c_4\lambda_n n^{\beta/2}q_n s_n^{a+1}, \qquad \text{(S2.27)}
\end{aligned}
$$

with probability tending to one, for some constant $c_4>0$, where the last inequality is by Lemma 3, Lemma 7, Theorem 1 and (A1).

For $\max_{l\le h_n s_n} |n^{-1/2}\sum_{i=1}^n \Delta_{6il}|$, we have

$$
\begin{aligned}
&\max_{l\le h_n s_n} |n^{-1/2}\sum_{i=1}^n \Delta_{6il}| \\
\le&\big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\big)\max_{l\le h_n s_n}\Big|n^{-1/2}\sum_{i=1}^n\Big[\big\{E(E_{il}^2)\big\}^{-1/2}(\hat{w}_l-w_l)'F_i(\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})\Big]\Big| \\
\le& n^{1/2}\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)\Big[\max_{l\le h_n s_n}\big\{E(E_{il}^2)\big\}^{-1/2}\Big]\Big(\max_{l\le h_n s_n}||\hat{w}_l-w_l||_1\Big) \\
&\cdot\Big\{||n^{-1}\sum_{i=1}^n F_i(\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})||_\infty\Big\} \\
\le& n^{1/2}\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)\Big[\big\{\lambda_{\min}(\Lambda)\big\}^{-1}\Big]\Big(\max_{l\le h_n s_n}||\hat{w}_l-w_l||_1\Big) \\
&\cdot\Big(\max_{l\le(p_n-h_n)s_n}\max_{i\le n}\big|F_{il}\big|\Big)\Big(n^{-1}\sum_{i=1}^n\big|\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}\big|\Big) \\
\le& O_p(n^{\beta}\rho_n q_n s_n^{3a/2-\delta}), \qquad \text{(S2.28)}
\end{aligned}
$$

where the last inequality is by Lemma 3, Lemma 7, (A1), (A2) and (S2.24).

In summary, by combining (S2.19), (S2.20), (S2.21), (S2.25), (S2.26), (S2.27) with (S2.28), we have

$$\begin{aligned}
&\max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}(\hat{S}_{il}-S_{il})|\le \max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}\Delta_{1il}|+\max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}\Delta_{2il}|\\
&+\max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}\Delta_{3il}|+\max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}\Delta_{4il}|\\
&+\max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}\Delta_{5il}|+\max_{l\le h_n s_n}|n^{-1/2}\sum_{i=1}^{n}\Delta_{6il}|\\
&\le c_0 n^{\beta-1/2}+c_1\lambda_n n^{\beta/2}q_n s_n^{a/2+1}+O_p(n^{\beta/2+1/2}q_n s_n^{-\delta})\\
&+c_3 n^{\beta-1/2}\rho_n s_n^{3a/2}+c_4\lambda_n n^{\beta/2}q_n s_n^{a+1}+O_p(n^{\beta}\rho_n q_n s_n^{3a/2-\delta})\\
&\le c_5(n^{\beta-1/2}\rho_n s_n^{3a/2}+\lambda_n n^{\beta/2}q_n s_n^{a+1}+n^{\beta/2+1/2}q_n s_n^{-\delta}\log s_n\\
&+n^{\beta}\rho_n q_n s_n^{3a/2-\delta}\log s_n),
\end{aligned}$$

with probability tending to one, for some constant $c_5>0$, which completes the proof of part 1).

Next, we start to prove part 2). First, we have

$$\begin{aligned}
&\max_{l\le h_n s_n}\{n^{-1}\sum_{i=1}^{n}(\hat{S}_{il}-S_{il})^2\}\\
\le&100\big(\max_{l\le h_n s_n}n^{-1}\sum_{i=1}^{n}\Delta_{1il}^2+\max_{l\le h_n s_n}n^{-1}\sum_{i=1}^{n}\Delta_{2il}^2+\max_{l\le h_n s_n}n^{-1}\sum_{i=1}^{n}\Delta_{3il}^2\\
&+\max_{l\le h_n s_n}n^{-1}\sum_{i=1}^{n}\Delta_{4il}^2+\max_{l\le h_n s_n}n^{-1}\sum_{i=1}^{n}\Delta_{5il}^2+\max_{l\le h_n s_n}n^{-1}\sum_{i=1}^{n}\Delta_{6il}^2\big).
\end{aligned}\tag{S2.29}$$

For $\max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}\Delta_{1il}^2$, we have

$$
\begin{aligned}
&\max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}\Delta_{1il}^2\\
\le&\Big(||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)^2 \max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}\Big[\big\{E(E_{il}^2)\big\}^{-1/2}(w_l{}'F_i-E_{il})\epsilon_i\Big]^2\\
\le&\Big(||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)^2 \max_{l\le h_n s_n}\max_{i\le n}\Big[\big\{E(E_{il}^2)\big\}^{-1/2}(w_l{}'F_i-E_{il})\epsilon_i\Big]^2\\
\le&c_6 n^{2\beta-1}\log n, \qquad\qquad \text{(S2.30)}
\end{aligned}
$$

with probability tending to one, for some constant $c_6>0$, where the last inequality is by Lemma 3 and (A2).

For $\max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}\Delta_{2il}^2$, we have

$$
\begin{aligned}
&\max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}\Delta_{2il}^2\\
\le&\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)^2 \max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}\Big[\big\{E(E_{il}^2)\big\}^{-1/2}(w_l{}'F_i-E_{il})F_i{}'(\eta_{\mathcal{H}_n^c}-\hat{\eta}_{\mathcal{H}_n^c})\Big]^2\\
\le&\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)^2\Big(||\hat{\eta}-\eta||_1\Big)^2\Big[\max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}\big|\big|\big\{E(E_{il}^2)\big\}^{-1/2}(w_l{}'F_i-E_{il})F_i{}'\big|\big|_\infty^2\Big]\\
\le&\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)^2\Big(||\hat{\eta}-\eta||_1\Big)^2\Big[\max_{l\le h_n s_n}\max_{i\le n}\max_{l_1\le(p_n-h_n)s_n}\big|\big\{E(E_{il}^2)\big\}^{-1/2}(w_l{}'F_i-E_{il})F_{il_1}\big|^2\Big]\\
\le&c_7\lambda_n^2 n^{2\beta}q_n^2 s_n^{a+2}, \qquad\qquad \text{(S2.31)}
\end{aligned}
$$

with probability tending to one, for some constant $c_7>0$, where the last inequality is by Lemma 3, Theorem 1 and (A2).

For $\max_{l\le h_n s_n} n^{-1}\sum_{i=1}^n \Delta_{3il}^2$, we have

$$
\begin{aligned}
&\max_{l\le h_n s_n} n^{-1}\sum_{i=1}^n \Delta_{3il}^2\\
\le&\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)^2 \max_{l\le h_n s_n} n^{-1}\sum_{i=1}^n \Big[\big\{E(E_{il}^2)\big\}^{-1/2}({w_l}'F_i-E_{il})(\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})\Big]^2\\
\le&\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)^2\Big[\max_{l\le h_n s_n}\max_{i\le n}\big|\big\{E(E_{il}^2)\big\}^{-1/2}({w_l}'F_i-E_{il})\big|^2\Big]\\
&\cdot\Big\{n^{-1}\sum_{i=1}^n(\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\Big\}. \qquad \text{(S2.32)}
\end{aligned}
$$

Regarding $\max_{l\le h_n s_n}\max_{i\le n}\big|\big\{E(E_{il}^2)\big\}^{-1/2}({w_l}'F_i-E_{il})\big|^2$, we have

$$
\begin{aligned}
&\max_{l\le h_n s_n}\max_{i\le n}\big|\big\{E(E_{il}^2)\big\}^{-1/2}({w_l}'F_i-E_{il})\big|^2\\
\le& c_8\log(nh_ns_n)\le c_8\log(np_ns_n)\\
\le& c_9 n^{\beta}, \qquad \text{(S2.33)}
\end{aligned}
$$

with probability tending to one, for some constants $c_8, c_9>0$, where the last inequality is by (A2), (A5) and (B1).

Regarding $n^{-1}\sum_{i=1}^{n}(\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2$, we have

$$\begin{aligned}
&E\big\{n^{-1}\sum_{i=1}^{n}(\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk})^2\big\}\le E\big\{n^{-1}\sum_{i=1}^{n}(\sum_{j\in\mathcal{H}_n^c}|\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}|)^2\big\}\\
\le& E\big\{n^{-1}\sum_{i=1}^{n}(\sum_{j=1}^{q_n}|\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}|)^2\big\}\le E\big(n^{-1}q_n\sum_{i=1}^{n}\sum_{j=1}^{q_n}|\sum_{k=s_n+1}^{\infty}\theta_{ijk}\eta_{jk}|^2\big)\\
\le& E\big\{n^{-1}q_n\sum_{i=1}^{n}\sum_{j=1}^{q_n}\big(\sum_{k=s_n+1}^{\infty}\theta_{ijk}^2k^{-2\delta}\big)\big(\sum_{k_1=s_n+1}^{\infty}\eta_{jk_1}^2k_1^{2\delta}\big)\big\}\\
=& n^{-1}q_n\sum_{i=1}^{n}\sum_{j=1}^{q_n}\big(\sum_{k=s_n+1}^{\infty}\omega_{jk}k^{-2\delta}\big)\big(\sum_{k_1=s_n+1}^{\infty}\eta_{jk_1}^2k_1^{2\delta}\big)\\
=& O(q_n^2s_n^{-2\delta}). \qquad\qquad (S2.34)
\end{aligned}$$

Hence, by combining (S2.32), (S2.33), (S2.34) with Lemma 3, one can show that

$$\max_{l\le h_ns_n} n^{-1}\sum_{i=1}^{n}\Delta_{3il}^2\le O_p(n^\beta q_n^2 s_n^{-2\delta}). \qquad (S2.35)$$

For $\max_{l\le h_ns_n} n^{-1}\sum_{i=1}^{n}\Delta_{4il}^2$, we have

$$\begin{aligned}
&\max_{l\le h_ns_n} n^{-1}\sum_{i=1}^{n}\Delta_{4il}^2\\
\le&\Big(1+||\Lambda\hat\Lambda^{-1}-I||_\infty\Big)^2\Big[\big\{\lambda_{\min}(\Lambda)\big\}^{-1}\Big]^2\Big(\max_{l\le h_ns_n}||\hat w_l-w_l||_1\Big)^2\Big(n^{-1}\sum_{i=1}^{n}\big|\big|F_i\epsilon_i\big|\big|_\infty^2\Big)\\
\le&\Big(1+||\Lambda\hat\Lambda^{-1}-I||_\infty\Big)^2\Big[\big\{\lambda_{\min}(\Lambda)\big\}^{-1}\Big]^2\Big(\max_{l\le h_ns_n}||\hat w_l-w_l||_1\Big)^2\Big(\max_{i\le n}\max_{l\le(p_n-h_n)s_n}\big|F_{il}\epsilon_i\big|\Big)^2\\
\le& c_{10}\rho_n^2s_n^{3a}n^{2\beta-1}\log n, \qquad\qquad (S2.36)
\end{aligned}$$

with probability tending to one, for some constant $c_{10}>0$, where the last inequality is by Lemma 3, Lemma 7, (A1) and (A2).

For $\max_{l\le h_n s_n} n^{-1}\sum_{i=1}^n \Delta_{5il}^2$, we have

$$\begin{aligned}
&\max_{l\le h_n s_n} n^{-1}\sum_{i=1}^n \Delta_{5il}^2\\
\le&\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)^2\Big[\big\{\lambda_{\min}(\Lambda)\big\}^{-1}\Big]^2\Big(\max_{l\le h_n s_n}||\hat{w}_l-w_l||_1\Big)^2\Big(||\hat{\eta}-\eta||_1\Big)^2\Big(\max_{i\le n}||F_i||_\infty^4\Big)\\
\le&c_{11}\lambda_n^2\rho_n^2 q_n^2 s_n^{4a+2}n^{3\beta-1}, \qquad\qquad \text{(S2.37)}
\end{aligned}$$

with probability tending to one, for some constant $c_{11}>0$, where the last inequality is by Lemma 3, Lemma 7, Theorem 1, (A1) and (A2).

For $\max_{l\le h_n s_n} n^{-1}\sum_{i=1}^n \Delta_{6il}^2$, we have

$$\begin{aligned}
&\max_{l\le h_n s_n} n^{-1}\sum_{i=1}^n \Delta_{6il}^2\\
\le&\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)^2\Big[\big\{\lambda_{\min}(\Lambda)\big\}^{-1}\Big]^2\Big(\max_{l\le h_n s_n}||\hat{w}_l-w_l||_1\Big)^2\\
&\cdot\Big\{n^{-1}\sum_{i=1}^n\big|\big|F_i(\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^\infty\theta_{ijk}\eta_{jk})\big|\big|_\infty^2\Big\}\\
=&\Big(1+||\Lambda\hat{\Lambda}^{-1}-I||_\infty\Big)^2\Big[\big\{\lambda_{\min}(\Lambda)\big\}^{-1}\Big]^2\Big(\max_{l\le h_n s_n}||\hat{w}_l-w_l||_1\Big)^2\Big(\max_{i\le n}\big|\big|F_i\big|\big|_\infty^2\Big)\\
&\cdot\Big\{n^{-1}\sum_{i=1}^n(\sum_{j\in\mathcal{H}_n^c}\sum_{k=s_n+1}^\infty\theta_{ijk}\eta_{jk})^2\Big\}\\
\le&O_p(\rho_n^2 q_n^2 s_n^{3a-2\delta}n^{2\beta-1}), \qquad\qquad \text{(S2.38)}
\end{aligned}$$

where the last inequality is by Lemma 3, Lemma 7, (S2.34), (A1) and (A2).

In summary, by combining (S2.30), (S2.31), (S2.35), (S2.36), (S2.37), (S2.38) with (S2.29), we observe that

$$\begin{aligned}&\max_{l\le h_ns_n}\big\{n^{-1}\sum_{i=1}^{n}(\hat{S}_{il}-S_{il})^2\big\}\\ \le& c_{12}\big\{\lambda_n^2n^{2\beta}q_n^2s_n^{a+2}+n^{\beta}q_n^2s_n^{-2\delta}(\log s_n)^2+\rho_n^2s_n^{3a}n^{2\beta-1}\log n\\ &+\lambda_n^2\rho_n^2q_n^2s_n^{4a+2}n^{3\beta-1}+n^{2\beta-1}\rho_n^2q_n^2s_n^{3a-2\delta}(\log s_n)^2\big\},\end{aligned} \tag{S2.39}$$

with probability tending to one, for some constant $c_{12}>0$, which further implies that

$$\begin{aligned}&\max_{l\le h_ns_n}\big\{n^{-1}\sum_{i=1}^{n}(\hat{S}_{il}-S_{il})^2\big\}^{1/2}\\ \le& c_{13}\big\{\lambda_nn^{\beta}q_ns_n^{a/2+1}+n^{\beta/2}q_ns_n^{-\delta}\log s_n+\rho_ns_n^{3a/2}n^{\beta-1/2}(\log n)^{1/2}\\ &+\lambda_n\rho_nq_ns_n^{2a+1}n^{3\beta/2-1/2}+n^{\beta-1/2}\rho_nq_ns_n^{3a/2-\delta}\log s_n\big\},\end{aligned} \tag{S2.40}$$

with probability tending to one, for some constant $c_{13}>0$, which completes the proof of part 2). □

S3 Proofs of Main Theorems

Proof of Theorem 1. Recall that we denote ν as $\nu=\eta-\eta^*$, and $\breve{\nu}=\breve{\eta}-\breve{\eta}^*=\Lambda\nu$. By first order necessary condition of the optimization theory, any local minimizer $\hat{\eta}$ of $Q_n(\eta)$ from (5) of the main paper must satisfy $\hat{\eta}\in\{\eta:\langle\nabla L_n(\eta)+\nabla P_{\lambda_n}(\eta),-\nu\rangle\ge 0,\quad ||\nu||_1\le R_n\}$, where $\langle a,b\rangle=a'b$ for any vectors a, b. Hence, in order to prove Theorem 1, it is sufficient to show that for any $\eta \in \{\eta : \langle \nabla L_n(\eta) + \nabla P_{\lambda_n}(\eta), -\nu\rangle \geq 0, \quad ||\nu||_1 \leq R_n\}$, parts 1) and 2) of Theorem 1 hold. First, we assume

$$\eta \in \{\eta : \langle \nabla L_n(\eta) + \nabla P_{\lambda_n}(\eta), -\nu\rangle \geq 0, \quad ||\nu||_1 \leq R_n\}. \tag{S3.41}$$

Then, it can be deduced that

$$\langle \nabla L_n(\eta) - \nabla L_n(\eta^*), \nu\rangle = \breve{\nu}'(n^{-1}\breve{\Theta}'\breve{\Theta})\breve{\nu} = \breve{\nu}'E(n^{-1}\breve{\Theta}'\breve{\Theta})\breve{\nu} - \breve{\nu}'\{E(n^{-1}\breve{\Theta}'\breve{\Theta}) - n^{-1}\breve{\Theta}'\breve{\Theta}\}\breve{\nu}$$
$$\geq \lambda_{\min}(\breve{I})||\breve{\nu}||_2^2 - ||E(n^{-1}\breve{\Theta}'\breve{\Theta}) - n^{-1}\breve{\Theta}'\breve{\Theta}||_\infty ||\breve{\nu}||_1^2 \geq m_0||\breve{\nu}||_2^2 - c_0 n^{\beta/2-1/2}||\breve{\nu}||_1^2, \tag{S3.42}$$

with probability tending to one, for some constant $c_0 > 0$, where the last inequality is by (A3) and Lemma 4. In addition, with a little abuse of notation, denote

$$P_{\lambda_n,\mu}(\eta) = P_{\lambda_n}(\eta) + 2^{-1}\mu n^{-1}\sum_{j=1}^{p_n}||\Theta_j\eta_j||_2^2 = \sum_{j=1}^{p_n}\{\rho_{\lambda_n s_n^{1/2}}(n^{-1/2}||\Theta_j\eta_j||_2) + 2^{-1}\mu n^{-1}||\Theta_j\eta_j||_2^2\}.$$

Under (P5), it is not hard to verify that $P_{\lambda_n,\mu}(\eta)$ is convex in η, i.e.,

$$P_{\lambda_n,\mu}(\eta^*) - P_{\lambda_n,\mu}(\eta) \geq \langle \nabla P_{\lambda_n,\mu}(\eta), -\nu\rangle,$$

which further implies that

$$\langle \nabla P_{\lambda_n}(\eta), -\nu\rangle \leq P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta) + 2^{-1}\mu n^{-1}\sum_{j=1}^{p_n}||\Theta_j(\eta_j - \eta_j^*)||_2^2. \tag{S3.43}$$

By combining (S3.41), (S3.42) with (S3.43), we have

$$
\begin{aligned}
&m_0||\breve{\nu}||_2^2 - c_0 n^{\beta/2-1/2}||\breve{\nu}||_1^2 \leq \langle \nabla L_n(\eta) - \nabla L_n(\eta^*), \nu\rangle \\
&= \langle \nabla L_n(\eta) + \nabla P_{\lambda_n}(\eta), \nu\rangle + \langle \nabla P_{\lambda_n}(\eta), -\nu\rangle + \langle -\nabla L_n(\eta^*), \nu\rangle \\
&\leq P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta) + 2^{-1}\mu n^{-1}\sum_{j=1}^{p_n}||\Theta_j(\eta_j - \eta_j^*)||_2^2 + ||n^{-1}\breve{\Theta}'(Y - \Theta\eta^*)||_\infty ||\breve{\nu}||_1.
\end{aligned}
\tag{S3.44}
$$

By combining (S3.41), (S3.44) with Lemma 5, we have an upper bound for $m_0||\breve{\nu}||_2^2$, with probability tending to one for some $c_1 > 0$,

$$
P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta) + 2^{-1}\mu n^{-1}\sum_{j=1}^{p_n}||\Theta_j(\eta_j - \eta_j^*)||_2^2 + c_1(n^{\beta/2-1/2}R_n + q_n s_n^{-\delta})||\breve{\nu}||_1.
\tag{S3.45}
$$

By combining (S3.45) with Lemma 6 and (A9), we have $\big[m_0 - 2^{-1}m_1\mu\{1 + o(1)\}\big]||\breve{\nu}||_2^2 \leq \{1 + o(1)\}\big\{P_{\lambda_n}(\eta^*) - P_{\lambda_n}(\eta)\big\}$. Together with Lemma 2 and (A3), we have

$$
\begin{aligned}
0 \leq& \big[m_0 - 2^{-1}m_1\mu\{1 + o(1)\}\big]||\breve{\nu}||_2^2 \\
\leq& c_2\lambda_n s_n^{1/2}\big\{\sum_{j\in\mathcal{A}_n} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2 - \sum_{j\in\mathcal{A}_n^c} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2\big\},
\end{aligned}
\tag{S3.46}
$$

with probability tending to one, for some constant $c_2 > 0$. On one hand, (S3.46) implies that

$$||\breve{\nu}||_2^2 \leq c_3\lambda_n s_n^{1/2} \sum_{j\in\mathcal{A}_n} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2 \leq c_3\lambda_n s_n^{1/2} q_n^{1/2}\{\sum_{j\in\mathcal{A}_n} n^{-1}||\Theta_j(\eta_j - \eta_j^*)||_2^2\}^{1/2}$$
$$\leq c_3\lambda_n s_n^{1/2} q_n^{1/2}\{\sum_{j=1}^{p_n} n^{-1}||\Theta_j(\eta_j - \eta_j^*)||_2^2\}^{1/2} \leq c_4\lambda_n s_n^{1/2} q_n^{1/2}||\breve{\nu}||_2,$$

with probability tending to one, for some constants $c_3, c_4 > 0$, where the last inequality is by Lemma 6. Then, it follows that we have $||\breve{\nu}||_2 \leq c_4\lambda_n s_n^{1/2} q_n^{1/2}$, which further entails that

$$||\nu||_2 = ||\Lambda^{-1}\breve{\nu}||_2 \leq \lambda_{\max}(\Lambda^{-1})||\breve{\nu}||_2 = \big\{\lambda_{\min}(\Lambda)\big\}^{-1}||\breve{\nu}||_2 \leq c_5\lambda_n s_n^{a/2+1/2} q_n^{1/2},$$

with probability tending to one, for some constant $c_5 > 0$, which completes the proof of part $1)$. On the other hand, (S3.46) also implies that $\sum_{j\in\mathcal{A}_n} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2 \geq \sum_{j\in\mathcal{A}_n^c} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2$. Together with Lemma 6, we have

$$||\breve{\nu}||_1 \leq c_5 s_n^{1/2} \sum_{j=1}^{p_n} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2 \leq 2c_5 s_n^{1/2} \sum_{j\in\mathcal{A}_n} n^{-1/2}||\Theta_j(\eta_j - \eta_j^*)||_2$$
$$\leq 2c_5 s_n^{1/2} q_n^{1/2}\{\sum_{j=1}^{p_n} n^{-1}||\Theta_j(\eta_j - \eta_j^*)||_2^2\}^{1/2} \leq c_6 s_n^{1/2} q_n^{1/2}||\breve{\nu}||_2 \leq c_7\lambda_n s_n q_n,$$

with probability tending to one, for some constants $c_5, c_6, c_7 > 0$, which further entails that

$$||\nu||_1 = ||\Lambda^{-1}\breve{\nu}||_1 \leq \lambda_{\max}(\Lambda^{-1})||\breve{\nu}||_1 = \big\{\lambda_{\min}(\Lambda)\big\}^{-1}||\breve{\nu}||_1 \leq c_8\lambda_n s_n^{a/2+1} q_n,$$

with probability tending to one, for some constant $c_8 > 0$, which completes the proof of part $2)$. □

Proof of Theorem 2. First, we have

$$\big|||\hat{T}^*||_\infty - ||T^*||_\infty\big| \le ||\hat{T}^* - T^*||_\infty = \max_{l\le h_n s_n} |n^{-1/2}\sum_{i=1}^{n}(\hat{S}_{il} - S_{il})| \le c_0 g(n), \tag{S3.47}$$

with probability tending to one, for some constant $c_0 > 0$, where $g(n) = n^{\beta-1/2}\rho_n s_n^{3a/2} + \lambda_n n^{\beta/2} q_n s_n^{a+1} + n^{\beta/2+1/2} q_n s_n^{-\delta}\log s_n + n^{\beta}\rho_n q_n s_n^{3a/2-\delta}\log s_n$, and the last inequality is by Lemma 8. Second, we have

$$\big|||\hat{T}_e^*||_\infty - ||T_e^*||_\infty\big| \le ||\hat{T}_e^* - T_e^*||_\infty = \max_{l\le h_n s_n} |n^{-1/2}\sum_{i=1}^{n} e_i(\hat{S}_{il} - S_{il})|. \tag{S3.48}$$

Since $\{e_1,\ldots,e_n\}$ is a set of independent and identically distributed standard normal random variables independent of the data, by the Hoeffding inequality, we have that for any $l = 1,\ldots,h_n s_n$ and $t > 0$,

$$P_e\{|n^{-1/2}\sum_{i=1}^{n} e_i(\hat{S}_{il} - S_{il})| \ge t\} \le 2\exp\Big[-t^2/\{2n^{-1}\sum_{i=1}^{n}(\hat{S}_{il} - S_{il})^2\}\Big],$$

where $P_e(\cdot)$ means the probability with respect to e. Then, the union bound inequality yields,

$$P_e\{\max_{l\le h_n s_n} |n^{-1/2}\sum_{i=1}^{n} e_i(\hat{S}_{il} - S_{il})| \ge t\} \le 2h_n s_n \exp\Big[-t^2/\{2\max_{l\le h_n s_n} n^{-1}\sum_{i=1}^{n}(\hat{S}_{il} - S_{il})^2\}\Big].$$

Together with Lemma 8, it is easy to see that

$$P_e\{\max_{l\le h_n s_n} |n^{-1/2}\sum_{i=1}^{n} e_i(\hat{S}_{il} - S_{il})| \le c_1 f(n)\} \to 1,$$

with probability tending to one, for some constant $c_1 > 0$, where

$$f(n) = \lambda_n n^{3\beta/2} q_n s_n^{a/2+1} + n^{\beta} q_n s_n^{-\delta} \log s_n + \rho_n s_n^{3a/2} n^{3\beta/2-1/2} (\log n)^{1/2} + \lambda_n \rho_n q_n s_n^{2a+1} n^{2\beta-1/2} + n^{3\beta/2-1/2} \rho_n q_n s_n^{3a/2-\delta} \log s_n.$$

Together with (S3.48), we obtain

$$P_e\Big\{\Big| ||\hat{T}_e^*||_\infty - ||T_e^*||_\infty \Big| \le c_1 f(n)\Big\} \to 1. \tag{S3.49}$$

Moreover, we have

$$g(n)\{\log(p_n s_n)\}^{1/2} \sim g(n) n^{\beta/2} = n^{3\beta/2-1/2} \rho_n s_n^{3a/2} + \lambda_n n^{\beta} q_n s_n^{a+1} + n^{\beta+1/2} q_n s_n^{-\delta} \log s_n + n^{3\beta/2} \rho_n q_n s_n^{3a/2-\delta} \log s_n = o(1), \tag{S3.50}$$

under (A5), (B1)–(B3). In addition, we also have

$$f(n)\{\log(p_n s_n)\}^{1/2} \sim f(n) n^{\beta/2} = \lambda_n n^{2\beta} q_n s_n^{a/2+1} + n^{3\beta/2} q_n s_n^{-\delta} \log s_n + \rho_n s_n^{3a/2} n^{2\beta-1/2} (\log n)^{1/2} + \lambda_n \rho_n q_n s_n^{2a+1} n^{5\beta/2-1/2} + n^{2\beta-1/2} \rho_n q_n s_n^{3a/2-\delta} \log s_n = o(1), \tag{S3.51}$$

under (A5), (B1)–(B3). Furthermore, (A4) implies that

$$E(S_{il}^2) \ge c_2, \tag{S3.52}$$

for some universal constant $c_2 > 0$, and (A2) implies that

$$||S_{il}||_{\phi_2} \le c_3, \tag{S3.53}$$

for some universal constant $c_3 > 0$. Hence, by combining (S3.47), (S3.49), (S3.50), (S3.51), (S3.52), (S3.53) with (A5), we have

$$\lim_{n\to\infty} \sup_{\alpha\in(0,1)} \left| P\{||\hat{T}^*||_\infty \le c_B(\alpha)\} - (1-\alpha) \right| = 0,$$

by Lemma H.7 in Ning and Liu (2017), which completes the proof. □

Proof of Theorem 3. First, for the decorrelated Wald test, by definition, we have

$$\begin{aligned} \hat{W}^* &= n^{1/2}\hat{\Lambda}_{\mathcal{H}_n}^{-1}\hat{I}_{\mathcal{H}_n|\mathcal{H}_n^c}\acute{\eta}_{\mathcal{H}_n} = n^{1/2}\hat{\Lambda}_{\mathcal{H}_n}^{-1}(\hat{I}_{\mathcal{H}_n\mathcal{H}_n} - \hat{w}'\hat{I}_{\mathcal{H}_n^c\mathcal{H}_n})\acute{\eta}_{\mathcal{H}_n} \\ &= n^{1/2}\hat{\Lambda}_{\mathcal{H}_n}^{-1}(n^{-1}\Theta_{\mathcal{H}_n}'\Theta_{\mathcal{H}_n} - n^{-1}\hat{w}'\Theta_{\mathcal{H}_n^c}'\Theta_{\mathcal{H}_n})\acute{\eta}_{\mathcal{H}_n}, \end{aligned} \tag{S3.54}$$

where

$$\begin{aligned} \acute{\eta}_{\mathcal{H}_n} &= \hat{\eta}_{\mathcal{H}_n} - \{\partial\hat{S}(\hat{\eta}_{\mathcal{H}_n}, \hat{\eta}_{\mathcal{H}_n^c})/\partial\eta_{\mathcal{H}_n}\}^{-1}\hat{S}(\hat{\eta}_{\mathcal{H}_n}, \hat{\eta}_{\mathcal{H}_n^c}) \\ &= -\{\hat{\Lambda}_{\mathcal{H}_n}^{-1}(n^{-1}\Theta_{\mathcal{H}_n}'\Theta_{\mathcal{H}_n} - n^{-1}\hat{w}'\Theta_{\mathcal{H}_n^c}'\Theta_{\mathcal{H}_n})\}^{-1}n^{-1/2}\hat{T}^*. \end{aligned} \tag{S3.55}$$

Hence, by combining (S3.54) with (S3.55), it is easy to see that

$$\hat{W}^* = -\hat{T}^*. \tag{S3.56}$$

Second, for the decorrelated likelihood ratio test, by definition, we have

$$\hat{L}^* = \hat{\Lambda}_{\mathcal{H}_n}^{-2}\mathrm{diag}\{(\Theta_{\mathcal{H}_n^c}\hat{w} - \Theta_{\mathcal{H}_n})'(\Theta_{\mathcal{H}_n^c}\hat{w} - \Theta_{\mathcal{H}_n})/n\}\hat{\Upsilon}, \tag{S3.57}$$

where $\hat{\Upsilon} = (\hat{\Upsilon}_1, \ldots, \hat{\Upsilon}_{(j-1)s_n+k}, \ldots, \hat{\Upsilon}_{h_n s_n})'$ whose $\{(j-1)s_n + k\}$'th element is

$$\hat{\Upsilon}_{(j-1)s_n+k} = 2n\{\hat{L}_{jk}(0,0,\hat{\eta}_{\mathcal{H}_n^c}) - \hat{L}_{jk}(\grave{\eta}_{jk},0,\hat{\eta}_{\mathcal{H}_n^c})\} = ||Y - \Theta_{\mathcal{H}_n^c}\hat{\eta}_{\mathcal{H}_n^c}||_2^2 - ||Y - \Theta_{\mathcal{H}_n^c}\hat{\eta}_{\mathcal{H}_n^c} - \grave{\eta}_{jk}(\Theta_{\mathcal{H}_n,(j-1)s_n+k} - \Theta_{\mathcal{H}_n^c}\hat{w}_{(j-1)s_n+k})||_2^2,$$

where $\Theta_{\mathcal{H}_n,(j-1)s_n+k}$ represents the $\{(j-1)s_n+k\}$'th column of $\Theta_{\mathcal{H}_n}$. Hence, we have

$$\begin{aligned}\hat{\Upsilon}_{(j-1)s_n+k} &= ||Y - \Theta_{\mathcal{H}_n^c}\hat{\eta}_{\mathcal{H}_n^c}||_2^2 - ||Y - \Theta_{\mathcal{H}_n^c}\hat{\eta}_{\mathcal{H}_n^c} - \grave{\eta}_{jk}(\Theta_{\mathcal{H}_n,(j-1)s_n+k} - \Theta_{\mathcal{H}_n^c}\hat{w}_{(j-1)s_n+k})||_2^2 \\ &= 2\grave{\eta}_{jk}(\Theta_{\mathcal{H}_n,(j-1)s_n+k} - \Theta_{\mathcal{H}_n^c}\hat{w}_{(j-1)s_n+k})'(Y - \Theta_{\mathcal{H}_n^c}\hat{\eta}_{\mathcal{H}_n^c}) - \\ &\quad \grave{\eta}_{jk}^2(\Theta_{\mathcal{H}_n,(j-1)s_n+k} - \Theta_{\mathcal{H}_n^c}\hat{w}_{(j-1)s_n+k})'(\Theta_{\mathcal{H}_n,(j-1)s_n+k} - \Theta_{\mathcal{H}_n^c}\hat{w}_{(j-1)s_n+k}),\end{aligned}$$

with $\grave{\eta}_{jk} = \{(\Theta_{\mathcal{H}_n,(j-1)s_n+k} - \Theta_{\mathcal{H}_n^c}\hat{w}_{(j-1)s_n+k})'(\Theta_{\mathcal{H}_n,(j-1)s_n+k} - \Theta_{\mathcal{H}_n^c}\hat{w}_{(j-1)s_n+k})\}^{-1} \cdot \{(\Theta_{\mathcal{H}_n,(j-1)s_n+k} - \Theta_{\mathcal{H}_n^c}\hat{w}_{(j-1)s_n+k})'(Y - \Theta_{\mathcal{H}_n^c}\hat{\eta}_{\mathcal{H}_n^c})\}$. Then

$$\hat{\Upsilon}_{(j-1)s_n+k} = \{(\Theta_{\mathcal{H}_n,(j-1)s_n+k} - \Theta_{\mathcal{H}_n^c}\hat{w}_{(j-1)s_n+k})'(\Theta_{\mathcal{H}_n,(j-1)s_n+k} - \Theta_{\mathcal{H}_n^c}\hat{w}_{(j-1)s_n+k})\}^{-1} \cdot \{(\Theta_{\mathcal{H}_n,(j-1)s_n+k} - \Theta_{\mathcal{H}_n^c}\hat{w}_{(j-1)s_n+k})'(Y - \Theta_{\mathcal{H}_n^c}\hat{\eta}_{\mathcal{H}_n^c})\}^2. \quad \text{(S3.58)}$$

Then, by combining (S3.57) with (S3.58), one has $\hat{L}_{(j-1)s_n+k} = \hat{T}^{*2}_{(j-1)s_n+k}$. Combining with (S3.56) yields $||\hat{T}^*||_\infty = ||\hat{W}^*||_\infty = ||\hat{L}^*||_\infty^{1/2}$, which completes the proof. □

References

Loh, P.-L. and M. J. Wainwright (2015). Regularized m-estimators with nonconvexity: statistical and algorithmic theory for local optima. *Journal of Machine Learning Research 71*, 559–616.

Ning, Y. and H. Liu (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics 45*, 158–195.