Supplementary Material for "Penalized communication-efficient algorithm for quantile regression with high-dimensional and large-scale longitudinal data"

Yanan Song, Haohui Han, Liya Fu*
School of Mathematics and Statistics, Xi'an Jiaotong University

This document contains supplementary material for "Penalized communication-efficient algorithm for quantile regression with high-dimensional and large-scale longitudinal data". Section 1 provides additional definitions and technical preliminaries. Section 2 presents supporting lemmas and their proofs. Section 3 contains detailed proofs of the main results from the article.

1 Preliminary Definitions and Notation

We begin by establishing key notations and definitions used throughout the supplementary material. We denote the global estimating function and its derivative as

$$\boldsymbol{U}_M(\boldsymbol{\beta}) = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{U}_i(\boldsymbol{\beta}), \quad \boldsymbol{D}_M(\boldsymbol{\beta}) = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{D}_i(\boldsymbol{\beta}),$$

where for $1 \le i \le M$,

$$\boldsymbol{U}_i(\boldsymbol{\beta}) = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{A}_i^{-1/2}\boldsymbol{R}_i^{-1}(\boldsymbol{\alpha})\boldsymbol{A}_i^{-1/2}\Big[\boldsymbol{K}_h\big(-\boldsymbol{\epsilon}_i(\boldsymbol{\beta})\big) - \mathbf{1}_{n_i}\cdot\tau\Big],$$
$$\boldsymbol{D}_i(\boldsymbol{\beta}) = \frac{\partial \boldsymbol{U}_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{A}_i^{-1/2}\boldsymbol{R}_i^{-1}(\boldsymbol{\alpha})\boldsymbol{A}_i^{-1/2}\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\boldsymbol{x}_i.$$

For each subset $\mathcal{I}_l$ (with $l = 1, \ldots, L$), we denote their local counterparts as

$$\boldsymbol{U}_{m,l}(\boldsymbol{\beta}) = \frac{1}{m}\sum_{i\in\mathcal{I}_l}\boldsymbol{U}_i(\boldsymbol{\beta}), \quad \boldsymbol{D}_{m,l}(\boldsymbol{\beta}) = \frac{1}{m}\sum_{i\in\mathcal{I}_l}\boldsymbol{D}_i(\boldsymbol{\beta}).$$

Let $\boldsymbol{J}_i(\boldsymbol{\beta}) = E\big[\boldsymbol{U}_i(\boldsymbol{\beta})\big]$ and $\boldsymbol{H}_i(\boldsymbol{\beta}) = E\big[\boldsymbol{D}_i(\boldsymbol{\beta})\big]$ be the expected estimating function and its derivative for the i-th subject, respectively. We define their global averages as

$$\boldsymbol{J}_M(\boldsymbol{\beta}) = E\big[\boldsymbol{U}_M(\boldsymbol{\beta})\big] = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{J}_i(\boldsymbol{\beta}), \quad \boldsymbol{H}_M(\boldsymbol{\beta}) = E\big[\boldsymbol{D}_M(\boldsymbol{\beta})\big] = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{H}_i(\boldsymbol{\beta}),$$

*The authors gratefully acknowledge *National Natural Science Foundation of China (No.12371295)*

and their local averages as

$$\boldsymbol{J}_{m,l}(\boldsymbol{\beta}) = E\big[\boldsymbol{U}_{m,l}(\boldsymbol{\beta})\big] = \frac{1}{m}\sum_{i\in\mathcal{I}_l}\boldsymbol{J}_i(\boldsymbol{\beta}), \quad \boldsymbol{H}_{m,l}(\boldsymbol{\beta}) = E\big[\boldsymbol{D}_{m,l}(\boldsymbol{\beta})\big] = \frac{1}{m}\sum_{i\in\mathcal{I}_l}\boldsymbol{H}_i(\boldsymbol{\beta}).$$

We further define the centered expected estimating function as

$$\Im_i(\boldsymbol{\beta}) = \boldsymbol{J}_i(\boldsymbol{\beta}) - \boldsymbol{J}_i(\boldsymbol{\beta}_\tau), \quad \Im_M(\boldsymbol{\beta}) = \frac{1}{M}\sum_{i=1}^{M}\Im_i(\boldsymbol{\beta}).$$

Finally, we introduce the stochastic process as

$$\wp_M(\boldsymbol{\beta}) = \boldsymbol{U}_M(\boldsymbol{\beta}) - \boldsymbol{U}_M(\boldsymbol{\beta}_\tau),$$

which satisfies:

$$E\big[\wp_M(\boldsymbol{\beta})\big] = \boldsymbol{J}_M(\boldsymbol{\beta}) - \boldsymbol{J}_M(\boldsymbol{\beta}_\tau) = \Im_M(\boldsymbol{\beta}).$$

2 Technical Lemmas

This section contains technical lemmas that support the proofs of the main theorems.

Lemma 1. *Following He et al. (2023)[4], for $\delta \in (0,1]$, define $\iota_\delta \geq 0$ as:*

$$\iota_\delta = \inf\big\{\iota > 0 : E\{\langle \boldsymbol{z}, \boldsymbol{u}\rangle^2 I(|\langle \boldsymbol{z}, \boldsymbol{u}\rangle| > \iota)\} \leq \delta \text{ for all } \boldsymbol{u} \in \mathbb{S}^{p-1}\big\}, \tag{1}$$

where $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}$ satisfying $E(\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}) = \boldsymbol{I}_{p\times p}$ and $\boldsymbol{u} \in \mathbb{S}^{p-1}$ satisfying $E\langle \boldsymbol{u}, \boldsymbol{z}\rangle^2 = 1$. For any $\gamma > 0$, $0 < r < h_l/(4\iota_{0.25})$ and $s \geq 0$, there exists $f_{l,l} \geq 0$ such that

$$\inf_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \frac{\bar{D}^{(s)}_{\widetilde{\mathcal{L}}_{M,l}}(\boldsymbol{\beta}, \boldsymbol{\beta}_\tau)}{\kappa_* C_m^{-1}\|\boldsymbol{\beta} - \boldsymbol{\beta}_\tau\|^2_{\boldsymbol{\Sigma}}} \geq \frac{3n_l}{4} f_{l,l} - \frac{5n_u}{4r}\sqrt{\frac{p f_{u,l} h_l}{m}} - 3n_u\sqrt{\frac{2\gamma\mu_4 f_{u,l}}{h_l m}} - \frac{13\gamma h_l n_u}{48 r^2 m}$$

holds with probability at least $1-e^{-\gamma}$, where $C_m = \tau(1-\tau)\xi_{\max}$ and $\kappa_* = \min_{|u|\leq 1} k(u) > 0$.

Proof. Define the symmetrized Bregman divergence for the surrogate loss functions $\widetilde{\mathcal{L}}^{(s)}_{M,l}(\boldsymbol{\beta})$ on the lth machine at s-th iteration as

$$\begin{aligned}\bar{D}^{(s)}_{\widetilde{\mathcal{L}}_{M,l}}(\boldsymbol{\beta}, \boldsymbol{\beta}_\tau) &= \Big\langle \nabla\widetilde{\mathcal{L}}^{(s)}_{M,l}(\boldsymbol{\beta}) - \nabla\widetilde{\mathcal{L}}^{(s)}_{M,l}(\boldsymbol{\beta}_\tau), \boldsymbol{\beta} - \boldsymbol{\beta}_\tau\Big\rangle \\ &= \Big\langle \nabla\mathcal{L}^{(s)}_{m,l}(\boldsymbol{\beta}) - \nabla\mathcal{L}^{(s)}_{m,l}(\boldsymbol{\beta}_\tau), \boldsymbol{\beta} - \boldsymbol{\beta}_\tau\Big\rangle = \bar{D}^{(s)}_{\mathcal{L}_{m,l}}(\boldsymbol{\beta}, \boldsymbol{\beta}_\tau).\end{aligned}$$

For $\epsilon_{ik} = y_{ik} - \boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\beta}_\tau$, $i = 1, \ldots, M$, $k = 1, \ldots, n_i$, define the event

$$\mathcal{E}_{ik} = \left\{|\epsilon_{ik}| \leq \frac{h_l}{2}\right\} \cap \left\{\frac{|\langle \boldsymbol{x}_{ik}, \boldsymbol{\beta} - \boldsymbol{\beta}_\tau\rangle|}{\|\boldsymbol{\beta} - \boldsymbol{\beta}_\tau\|_{\boldsymbol{\Sigma}}} \leq \frac{h_l}{2r}\right\}.$$

Recall that $\mathbb{B}_{\boldsymbol{\Sigma}}(r) = \{\boldsymbol{\beta} \in \mathbb{R}^p : \|\boldsymbol{\beta} - \boldsymbol{\beta}_\tau\|_{\boldsymbol{\Sigma}} \leq r\}$ is a local neighborhood of $\boldsymbol{\beta}_\tau$ under $\|\cdot\|_{\boldsymbol{\Sigma}}$-norm. For any $\boldsymbol{\beta} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r)$, we have $|y_{ik} - \boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\beta}| \leq h_l$ on $\mathcal{E}_{ik}$. Therefore, conditioned on the event $\mathcal{E}_{ik}$, we have

$$
\begin{aligned}
\bar{D}^{(s)}_{\mathcal{L}_{m,l}}(\boldsymbol{\beta},\boldsymbol{\beta}_\tau) &= \left\langle \frac{1}{m}\sum_{i\in\mathcal{I}_l}\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{W}_i^{(s)}\boldsymbol{x}_i(\boldsymbol{\beta}-\boldsymbol{\beta}_\tau), \boldsymbol{\beta}-\boldsymbol{\beta}_\tau\right\rangle \\
&= \frac{1}{m}\sum_{i\in\mathcal{I}_l}(\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)^{\mathrm{T}}\boldsymbol{x}_i^{\mathrm{T}}\big(\widehat{\boldsymbol{V}}_i^{(s)}\big)^{-1}\boldsymbol{\Lambda}_i^{(s)}\boldsymbol{x}_i(\boldsymbol{\beta}-\boldsymbol{\beta}_\tau) \\
&\geq \frac{\kappa_*}{mh_l}\frac{1}{\tau(1-\tau)\xi_{\max}}\sum_{i\in\mathcal{I}_l}\sum_{k=1}^{n_i}\langle \boldsymbol{x}_{ik},\boldsymbol{\beta}-\boldsymbol{\beta}_\tau\rangle^2 I(\mathcal{E}_{ik}) \\
&= \frac{\kappa_*}{mh_l}\frac{1}{\tau(1-\tau)\xi_{\max}}\sum_{i\in\mathcal{I}_l}\sum_{k=1}^{n_i}\langle \boldsymbol{x}_{ik},\boldsymbol{\beta}-\boldsymbol{\beta}_\tau\rangle^2 \varsigma_{ik} I\left(|\delta_{ik,\boldsymbol{v}}|\leq \frac{h_l}{2r}\right)
\end{aligned}
$$

for any $\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)$, where $\varsigma_{ik}=I(|\epsilon_{ik}|\leq h_l/2)$ and $\delta_{ik,\boldsymbol{v}}=\langle \boldsymbol{z}_{ik},\boldsymbol{v}\rangle$ with $\boldsymbol{z}_{ik}=\boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}_{ik}$, $\boldsymbol{v}=\boldsymbol{\Sigma}^{1/2}\boldsymbol{\zeta}/\|\boldsymbol{\zeta}\|_{\boldsymbol{\Sigma}}\in\mathbb{S}^{p-1}$ and $\boldsymbol{\zeta}=\boldsymbol{\beta}-\boldsymbol{\beta}_\tau$. Note that $E(\delta_{ik,\boldsymbol{v}}^2)=1$.

Following the prove of Lemma C.3 in He et al. (2023)[4], we have

$$
\langle \boldsymbol{z}_{ik},\boldsymbol{v}\rangle^2 I\left(\big|\langle \boldsymbol{z}_{ik},\boldsymbol{v}\rangle\big|\leq\frac{h_l}{2r}\right)\geq \varphi_{h_l/(2r)}\big(\langle \boldsymbol{z}_{ik},\boldsymbol{v}\rangle\big)\geq \langle \boldsymbol{z}_{ik},\boldsymbol{v}\rangle^2 I\left(\big|\langle \boldsymbol{z}_{ik},\boldsymbol{v}\rangle\big|\leq\frac{h_l}{4r}\right),
$$

where $\varphi_R(u)=u^2I(|u|\leq R/2)+\big[u\ \mathrm{sign}(u)-R\big]^2I(R/2<|u|\leq R)$ satisfying $u^2I(|u|\leq R/2)\leq\varphi_R(u)\leq u^2I(|u|\leq R)$, $\varphi_0(u)=0$, and $\varphi_{sR}(su)=s^2\varphi_R(u)$ for any $s>0$ and $R>0$. Define

$$
D_0(\boldsymbol{v})=\frac{1}{mh_l}\sum_{i\in\mathcal{I}_l}\sum_{k=1}^{n_i}\varsigma_{ik}\varphi_{h_l/(2r)}\big(\langle \boldsymbol{z}_{ik},\boldsymbol{v}\rangle\big)=\frac{1}{m}\sum_{i\in\mathcal{I}_l}\sum_{k=1}^{n_i}\aleph_{ik}(\boldsymbol{z}_{ik},\boldsymbol{v}), \tag{2}
$$

where $\aleph_{ik}(\boldsymbol{z}_{ik},\boldsymbol{v})=h_l^{-1}\varsigma_{ik}\varphi_{h_l/(2r)}\big(\langle \boldsymbol{z}_{ik},\boldsymbol{v}\rangle\big)$. At follows, we drive the lower bound of $D_0(\boldsymbol{v})$ uniformly over $\boldsymbol{v}\in\mathbb{S}^{p-1}$.

According to Condition (A5), there exist some constants $f_{u,l}\geq f_{l,l}\geq 0$ such that $f_{l,l}\leq\min_{|u|\leq h_l/2}f_{\epsilon|\boldsymbol{x}}(u)\leq\max_{|u|\leq h_l/2}f_{\epsilon|\boldsymbol{x}}(u)\leq f_{u,l}$ almost surely (over $\boldsymbol{x}$) as long as h_l is sufficiently small. Then, we have

$$
f_{l,l}h_l\leq E(\varsigma_{ik}|\boldsymbol{x}_{ik})=\int_{-h_l/2}^{h_l/2}f_{\epsilon|\boldsymbol{x}}(u)du\leq f_{u,l}h_l \tag{3}
$$

almost surely (over $\boldsymbol{x}$). For $r\leq h_l/(4\iota_{1/4})$ with $\iota_{1/4}$ defined in (1), we have

$$
\begin{aligned}
E\left[\varsigma_{ik}\varphi_{h_l/(2r)}(\delta_{ik,\boldsymbol{v}})\right] &\geq E\left[\varsigma_{ik}\delta_{ik,\boldsymbol{v}}^2 I\left(|\delta_{ik,\boldsymbol{v}}|\leq\frac{h_l}{4r}\right)\right] \\
&\geq f_{l,l}h_l\left\{1-E\left[\delta_{ik,\boldsymbol{v}}^2 I\left(|\delta_{ik,\boldsymbol{v}}|>\frac{h_l}{4r}\right)\right]\right\} \\
&\geq f_{l,l}h_l\Big\{1-\sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}E\big[\langle \boldsymbol{z},\boldsymbol{u}\rangle^2 I(|\langle \boldsymbol{z},\boldsymbol{u}\rangle|\geq\iota_{1/4})\big]\Big\} \\
&\geq \frac{3}{4}f_{l,l}h_l.
\end{aligned} \tag{4}
$$

Together with (2) and (4),

$$
\inf_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}E\big[D_0(\boldsymbol{v})\big]\geq\frac{3}{4}n_lf_{l,l}. \tag{5}
$$

Let $\mathcal{N}_\varepsilon \subseteq \mathbb{S}^{p-1}$ be an ε-net of $\mathbb{S}^{p-1}$, which is a finite set satisfying that: for any $\boldsymbol{u} \in \mathbb{S}^{p-1}$, there exists $\boldsymbol{u}' \in \mathcal{N}_\varepsilon$ such that $\|\boldsymbol{u} - \boldsymbol{u}'\|_2 \leq \varepsilon$. Define

$$\Pi(\boldsymbol{v}) = \sup_{\boldsymbol{v}\in\mathcal{N}_\varepsilon} \Big\{ E\big[D_0(\boldsymbol{v})\big] - D_0(\boldsymbol{v}) \Big\} = \sup_{\boldsymbol{v}\in\mathcal{N}_\varepsilon} \Big\{ \frac{1}{m} \sum_{i\in\mathcal{I}_l} \sum_{k=1}^{n_i} E\big[\aleph_{ik}(\boldsymbol{z}_{ik}, \boldsymbol{v})\big] - \aleph_{ik}(\boldsymbol{z}_{ik}, \boldsymbol{v}) \Big\}.$$

Next, we derive the upper bound of $\Pi(\boldsymbol{v})$.

Since $0 \leq \varphi_R(u) \leq (R/2)^2$ for all $u \in \mathbb{R}$. We have $0 \leq \aleph_{ik}(\boldsymbol{z}_{ik}, \boldsymbol{v}) \leq h_l/(4r)^2$. For $\varsigma_{ik} \in \{0,1\}$, based on (3), we have $E\left[\aleph_{ik}^2(\boldsymbol{z}_{ik}, \boldsymbol{v})\right] \leq h_l^{-1}\mu_4 f_{u,l}$. Define the function

$$g_i(\boldsymbol{z}_i, \boldsymbol{v}) = \frac{(4r)^2}{h_l n_u} \sum_{k=1}^{n_i} E\big[\aleph_{ik}\big(\langle \boldsymbol{z}_{ik}, \boldsymbol{v}\rangle\big)\big] - \aleph_{ik}\big(\langle \boldsymbol{z}_{ik}, \boldsymbol{v}\rangle\big),$$

where $\boldsymbol{z}_i = (\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{in_i})^{\mathrm{T}}$. Let $\mathcal{G}$ be a countable set of functions from $\mathcal{Z}$ to $\mathbb{R}$, and assume that all functions g_i ($i \in \mathcal{I}_l$) in $\mathcal{G}$ are measurable, square-integrable and satisfy $E\big[g_i(\boldsymbol{z}_i, \boldsymbol{v})\big] = 0$. We have $\sup_{g_i\in\mathcal{G}} g_i \leq 1$ and

$$\begin{aligned}\sup_{g_i\in\mathcal{G}} E\big[g_i^2(\boldsymbol{z}_i, \boldsymbol{v})\big] &= \frac{(4r)^4 n_i^2}{(h_l n_u)^2} E\Big\{ \frac{1}{n_i} \sum_{k=1}^{n_i} E\big[\aleph_{ik}\big(\langle \boldsymbol{z}_{ik}, \boldsymbol{v}\rangle\big)\big] - \aleph_{ik}\big(\langle \boldsymbol{z}_{ik}, \boldsymbol{v}\rangle\big) \Big\}^2 \\ &\leq \frac{(4r)^4}{h_l^2} \cdot \frac{1}{n_i} \sum_{k=1}^{n_i} E\Big\{ E\big[\aleph_{ik}\big(\langle \boldsymbol{z}_{ik}, \boldsymbol{v}\rangle\big)\big] - \aleph_{ik}\big(\langle \boldsymbol{z}_{ik}, \boldsymbol{v}\rangle\big) \Big\}^2 \leq \frac{f_{u,l}(4r)^4\mu_4}{h_l^3} = C_{h_l}.\end{aligned}$$

Therefore, $\sup_{g_i\in\mathcal{G}} \sum_{i\in\mathcal{I}_l} E\big[g_i^2(\boldsymbol{z}_i, \boldsymbol{v})\big] \leq mC_{h_l}$. Since

$$\begin{aligned}\sup_{g_i\in\mathcal{G}} \sum_{i\in\mathcal{I}_l} g_i(\boldsymbol{z}_i, \boldsymbol{v}) &= \frac{(4r)^2}{h_l n_u} \sup_{\boldsymbol{v}\in\mathcal{N}_\varepsilon} \sum_{i\in\mathcal{I}_l} \sum_{k=1}^{n_i} E\big[\aleph_{ik}\big(\langle \boldsymbol{z}_{ik}, \boldsymbol{v}\rangle\big) - \aleph_{ik}\big(\langle \boldsymbol{z}_{ik}, \boldsymbol{v}\rangle\big) \\ &= \frac{(4r)^2 m}{h_l n_u} \sup_{\boldsymbol{v}\in\mathcal{N}_\varepsilon} \Big\{ E\big[D_0(\boldsymbol{v})\big] - D_0(\boldsymbol{v}) \Big\} = \frac{(4r)^2 m}{h_l n_u} \Pi(\boldsymbol{v}).\end{aligned}$$

According to a refined Talagrand's inequality in Theorem 7.3 of Bousquet (2003) [2] and the elementary inequality $ab \leq a^2/4 + b^2$ for all $a, b \in \mathbb{R}$,

$$\begin{aligned}\Pi(\boldsymbol{v}) &\leq E\big[\Pi(\boldsymbol{v})\big] + \frac{h_l n_u}{(4r)^2 m} \sqrt{2\gamma \Big[m\frac{f_{u,l}(4r)^4\mu_4}{h_l^3} + 2\frac{(4r)^2 m}{h_l n_u} E\big[\Pi(\boldsymbol{v})\big] \Big]} + \frac{\gamma}{3} \frac{h_l n_u}{(4r)^2 m} \\ &\leq E\big[\Pi(\boldsymbol{v})\big] + n_u \sqrt{2\gamma \frac{\mu_4 f_{u,l}}{h_l m}} + \sqrt{\frac{4\gamma h_l n_u E\big[\Pi(\boldsymbol{v})\big]}{(4r)^2 m}} + \frac{\gamma h_l n_u}{3(4r)^2 m} \\ &\leq E\big[\Pi(\boldsymbol{v})\big] + n_u \sqrt{\frac{2\gamma\mu_4 f_{u,l}}{h_l m}} + \frac{1}{4} E\big[\Pi(\boldsymbol{v})\big] + \frac{\gamma h_l n_u}{4r^2 m} + \frac{\gamma h_l n_u}{48 r^2 m} \\ &\leq \frac{5}{4} E\big[\Pi(\boldsymbol{v})\big] + n_u \sqrt{\frac{2\gamma\mu_4 f_{u,l}}{h_l m}} + \frac{13\gamma h_l n_u}{48 r^2 m} \qquad (6)\end{aligned}$$

holds with probability at least $1 - e^{-\gamma}$ for any $\gamma > 0$.

Now, we derive the upper bound of $E\big[\Pi(\boldsymbol{v})\big]$. Since $\varsigma_{ik} = I(|\epsilon_{ik}| \leq h_l/2) \in \{0,1\}$, we have $\varsigma_{ik}\varphi_{h_l/(2r)}(\delta_{ik,\boldsymbol{v}}) = \varsigma_{ik}^2 \varphi_{h_l/(2r)}(\delta_{ik,\boldsymbol{v}}) = \varphi_{\varsigma_{ik} h_l/(2r)}(\varsigma_{ik}\delta_{ik,\boldsymbol{v}}) = \varphi_{h_l/(2r)}(\varsigma_{ik}\delta_{ik,\boldsymbol{v}})$. Write

$\aleph_{ik}\big(\bar{\boldsymbol{z}}_{ik},\boldsymbol{v}\big)=h_l^{-1}\varphi_{h_l/(2r)}\big(\langle\bar{\boldsymbol{z}}_{ik},\boldsymbol{v}\rangle\big)$ with $\bar{\boldsymbol{z}}_{ik}=\varsigma_{ik}\boldsymbol{z}_{ik}$. Since $\varphi_R(\cdot)$ is R-Lipschitz continuous, $h^{-1}\varphi_{h_l/(2r)}\big(\bar{\boldsymbol{z}}_{ik},\boldsymbol{v}\big)$ is $(2r)^{-1}$-Lipschitz continuous, that is, for any $\boldsymbol{v}$, $\boldsymbol{v}'\in\mathbb{R}^p$,

$$\frac{1}{h_l}\Big|\varphi_{h_l/(2r)}\big(\langle\bar{\boldsymbol{z}}_{ik},\boldsymbol{v}\rangle\big)-\varphi_{h_l/(2r)}\big(\langle\bar{\boldsymbol{z}}_{ik},\boldsymbol{v}'\rangle\big)\Big|\le\frac{1}{2r}\Big|\langle\bar{\boldsymbol{z}}_{ik},\boldsymbol{v}\rangle-\langle\bar{\boldsymbol{z}}_{ik},\boldsymbol{v}'\rangle\Big|. \tag{7}$$

Moreover, for any $\boldsymbol{v}$ such that $\langle\bar{\boldsymbol{z}}_{ik},\boldsymbol{v}\rangle=0$, we have $\aleph_{ik}\big(\bar{\boldsymbol{z}}_{ik},\boldsymbol{v}\big)=0$.

Suppose that $Z_1,\ldots,Z_m$ are m independent Rademacher random variables. According to Rademacher symmetrization, we have

$$E\big[\Pi(\boldsymbol{v})\big]\le 2E\left[\sup_{\boldsymbol{v}\in\mathcal{N}_\varepsilon}\left(\frac{1}{m}\sum_{i\in\mathcal{I}_l}Z_i\aleph_i\big(\langle\bar{\boldsymbol{z}}_i,\boldsymbol{v}\rangle\big)\right)\right], \tag{8}$$

where $\aleph_i\big(\langle\bar{\boldsymbol{z}}_i,\boldsymbol{v}\rangle\big)=\sum_{k=1}^{n_i}\aleph_{ik}\big(\langle\bar{\boldsymbol{z}}_{ik},\boldsymbol{v}\rangle\big)$ and $\bar{\boldsymbol{z}}_i=(\bar{\boldsymbol{z}}_{i1},\ldots,\bar{\boldsymbol{z}}_{in_i})^{\mathrm{T}}$. Define the subset $\mathcal{V}_l\subseteq\mathbb{R}^{N_l}$ ($N_l=\sum_{i\in\mathcal{I}_l}n_i$) as

$$\begin{aligned}\mathcal{V}_l=&\big\{\boldsymbol{\nu}=(\boldsymbol{\nu}_1^{\mathrm{T}},\ldots,\boldsymbol{\nu}_m^{\mathrm{T}})^{\mathrm{T}},\boldsymbol{\nu}_i=(\nu_{i1},\ldots,\nu_{in_i})^{\mathrm{T}}:\\&\nu_{ik}=\langle\bar{\boldsymbol{z}}_{ik},\boldsymbol{v}\rangle,\ i=1,\ldots,m,\ k=1,\ldots,n_i,\ \boldsymbol{v}\in\mathcal{N}_\varepsilon\big\}.\end{aligned}$$

Let $\phi_{ik}(\nu_{ik})=(2r/h_l)\cdot\varphi_{h_l/(2r)}(\nu_{ik})$. Then, ϕ_{ik} is a convex function on its segment value ranges and $\aleph_{ik}\big(\langle\bar{\boldsymbol{z}}_{ik},\boldsymbol{v}\rangle\big)=h_l^{-1}\varphi_{h_l/(2r)}(\nu_{ik})=(1/2r)\phi_{ik}(\nu_{ik})$. Based on (7), we have $|\phi(\nu)-\phi(\nu')|\le|\nu-\nu'|$ for $\nu,\nu'\in\mathbb{R}$. According to the Talagrand's contraction principle (4.20) of Theorem 4.12 in Ledoux and Talagrand (1991) [5],

$$\begin{aligned}2E\left[\sup_{\boldsymbol{v}\in\mathcal{N}_\varepsilon}\left(\frac{1}{m}\sum_{i\in\mathcal{I}_l}Z_i\aleph_i\big(\langle\bar{\boldsymbol{z}}_i,\boldsymbol{v}\rangle\big)\right)\right]&\le\frac{1}{r}E\left\{\sup_{\boldsymbol{\nu}\in\mathcal{V}_l}\left[\frac{1}{m}\sum_{i\in\mathcal{I}_l}Z_i\left(\sum_{k=1}^{n_i}\phi_{ik}(\nu_{ik})\right)\right]\right\}\\&\le\frac{1}{r}E\left\{\sup_{\boldsymbol{\nu}\in\mathcal{V}_l}\left[\frac{1}{m}\sum_{i\in\mathcal{I}_l}Z_i\left(\sum_{k=1}^{n_i}\nu_{ik}\right)\right]\right\}\\&=\frac{1}{r}E\left\{\sup_{\boldsymbol{v}\in\mathcal{N}_\varepsilon}\left[\frac{1}{m}\sum_{i\in\mathcal{I}_l}Z_i\left(\sum_{k=1}^{n_i}\langle\bar{\boldsymbol{z}}_{ik},\boldsymbol{v}\rangle\right)\right]\right\}\\&\le\frac{1}{r}E\left\|\frac{1}{m}\sum_{i\in\mathcal{I}_l}Z_i\left(\sum_{k=1}^{n_i}\varsigma_{ik}\boldsymbol{z}_{ik}\right)\right\|_2.\end{aligned} \tag{9}$$

Let $\hbar_j=m^{-1}\sum_{i\in\mathcal{I}_l}Z_i\sum_{k=1}^{n_i}\varsigma_{ik}z_{ik,j}$. Note that $E\big(Z_i\sum_{k=1}^{n_i}\varsigma_{ik}z_{ik,j}\big)=0$. Based on (3),

$$\begin{aligned}E\big(\hbar_j^2\big)=&E\left[\frac{1}{m}\sum_{i\in\mathcal{I}_l}Z_i\left(\sum_{k=1}^{n_i}\varsigma_{ik}z_{ik,j}\right)\right]^2=\frac{1}{m^2}\sum_{i\in\mathcal{I}_l}E\left[Z_i^2n_i^2\left(\frac{1}{n_i}\sum_{k=1}^{n_i}\varsigma_{ik}z_{ik,j}\right)^2\right]\\\le&\frac{1}{m^2}\sum_{i\in\mathcal{I}_l}n_i\sum_{k=1}^{n_i}E\big[(\varsigma_{ik}z_{ik,j})^2\big]\le\frac{f_{u,l}h_ln_u^2}{m}\end{aligned}$$

for $j=1,\ldots,p$. Then,

$$E\left\|\frac{1}{m}\sum_{i\in\mathcal{I}_l}Z_i\left(\sum_{k=1}^{n_i}\varsigma_{ik}\boldsymbol{z}_{ik}\right)\right\|_2\le\sqrt{\sum_{j=1}^{p}E\big(\hbar_j^2\big)}\le n_u\sqrt{\frac{pf_{u,l}h_l}{m}}. \tag{10}$$

Together with (6) and (8)–(10),

$$\sup_{\boldsymbol{v}\in\mathcal{N}_\varepsilon}\Big\{E\big[D_0(\boldsymbol{v})\big]-D_0(\boldsymbol{v})\Big\}\leq\frac{5n_u}{4r}\sqrt{\frac{pf_{u,l}h_l}{m}}+n_u\sqrt{\frac{2\gamma\mu_4f_{u,l}}{h_lm}}+\frac{13\gamma h_ln_u}{48r^2m}\tag{11}$$

holds with probability at least $1-e^{-\gamma}$. Since for any $\boldsymbol{v}\in\mathbb{S}^{p-1}$, there exists $\boldsymbol{v}'\in\mathcal{N}_\varepsilon$ such that

$$\begin{aligned}&\sup_{\boldsymbol{v}\in\mathbb{S}^{p-1}}\Big\{E\big[D_0(\boldsymbol{v})\big]-D_0(\boldsymbol{v})\Big\}\\&=\sup_{\boldsymbol{v}\in\mathbb{S}^{p-1},\ \boldsymbol{v}'\in\mathcal{N}_\varepsilon}\Big\{E\big[D_0(\boldsymbol{v}')\big]-D_0(\boldsymbol{v}')+E\big[D_0(\boldsymbol{v})\big]-E\big[D_0(\boldsymbol{v}')\big]+D_0(\boldsymbol{v}')-D_0(\boldsymbol{v})\Big\}\\&\leq\sup_{\boldsymbol{v}'\in\mathcal{N}_\varepsilon}\Big\{E\big[D_0(\boldsymbol{v}')\big]-D_0(\boldsymbol{v}')\Big\}\end{aligned}\tag{12}$$

as $\varepsilon\to 0$. Together with (5), (11) and (12),

$$\begin{aligned}\inf_{\boldsymbol{v}\in\mathbb{S}^{p-1}}D_0(\boldsymbol{v})&=\inf_{\boldsymbol{v}\in\mathbb{S}^{p-1}}\Big\{D_0(\boldsymbol{v})-E\big[D_0(\boldsymbol{v})\big]+E\big[D_0(\boldsymbol{v})\big]\Big\}\\&\geq\inf_{\boldsymbol{v}\in\mathbb{S}^{p-1}}E\big[D_0(\boldsymbol{v})\big]+\inf_{\boldsymbol{v}\in\mathbb{S}^{p-1}}\Big\{D_0(\boldsymbol{v})-E\big[D_0(\boldsymbol{v})\big]\Big\}\\&=\inf_{\boldsymbol{v}\in\mathbb{S}^{p-1}}E\big[D_0(\boldsymbol{v})\big]-\sup_{\boldsymbol{v}\in\mathbb{S}^{p-1}}\Big\{-D_0(\boldsymbol{v})+E\big[D_0(\boldsymbol{v})\big]\Big\}\\&\geq\frac{3n_l}{4}f_{l,l}-\frac{5n_u}{4r}\sqrt{\frac{pf_{u,l}h_l}{m}}-n_u\sqrt{\frac{2\gamma\mu_4f_{u,l}}{h_lm}}-\frac{13\gamma h_ln_u}{48r^2m}\end{aligned}$$

holds with probability at least $1-e^{-\gamma}$. Consequently,

$$\begin{aligned}\inf_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\bar{D}^{(s)}_{\mathcal{L}_{m,l}}(\boldsymbol{\beta},\boldsymbol{\beta}_\tau)&\geq\frac{\kappa_*\|\boldsymbol{\zeta}\|^2_{\boldsymbol{\Sigma}}}{\tau(1-\tau)\xi_{\max}}\inf_{\boldsymbol{v}\in\mathbb{S}^{p-1}}D_0(\boldsymbol{v})\\&\geq\frac{\kappa_*\|\boldsymbol{\zeta}\|^2_{\boldsymbol{\Sigma}}}{\tau(1-\tau)\xi_{\max}}\left(\frac{3n_l}{4}f_{l,l}-\frac{5n_u}{4r}\sqrt{\frac{pf_{u,l}h_l}{m}}-3n_u\sqrt{\frac{2\gamma\mu_4f_{u,l}}{h_lm}}-\frac{13\gamma h_ln_u}{48r^2m}\right).\end{aligned}$$

■

Lemma 2. *Given an estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, denote $\widehat{\boldsymbol{R}}_i$ as the estimator of $\boldsymbol{R}_i=\boldsymbol{R}_i(\boldsymbol{\alpha})$ based on $\widehat{\boldsymbol{\beta}}$. Let $\bar{\tau}=\max(1-\tau,\tau)$. Under Conditions (A1)–(A4), for any $\varepsilon>0$, there exist constants $C_\varepsilon, M_0>0$ such that for every $M\geq M_0$, the following hold simultaneously with probability at least $1-\varepsilon$:*

(i)

$$\Big\|\widehat{\boldsymbol{U}}_M(\boldsymbol{\beta})-\boldsymbol{U}_M(\boldsymbol{\beta})\Big\|_\infty\leq C_0C_\varepsilon\sqrt{\frac{p}{M}},$$

where $C_0=(1-\bar{\tau})^{-1}n_uB$ is independent of M and p.

(ii)

$$\|\mathfrak{S}_M(\boldsymbol{\beta})\|_2=\Big\|\boldsymbol{\Sigma}^{-1/2}\Big\{\widehat{\boldsymbol{U}}_M(\boldsymbol{\beta})-\boldsymbol{U}_M(\boldsymbol{\beta})\Big\}\Big\|_2\leq\xi_p^{-1/2}C_0C_\varepsilon\frac{p}{\sqrt{M}},$$

where $\xi_p>0$ is the smallest eigenvalue of $\boldsymbol{\Sigma}=E(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})$.

Proof. Under Conditions (A1)–(A4), with probability at least $1-\varepsilon$:

$$\begin{aligned}
\mathfrak{U}_M &= \left\|\widehat{\boldsymbol{U}}_M(\boldsymbol{\beta}) - \boldsymbol{U}_M(\boldsymbol{\beta})\right\|_\infty \\
&= \max_{1\le j\le p}\left|\frac{1}{M}\sum_{i=1}^{M}\boldsymbol{x}_{i,j}^{\mathrm{T}}\boldsymbol{A}_i^{-1/2}\left(\widehat{\boldsymbol{R}}_i^{-1} - \boldsymbol{R}_i^{-1}\right)\boldsymbol{A}_i^{-1/2}\Big[\boldsymbol{K}_h\big(-\boldsymbol{\epsilon}_i(\boldsymbol{\beta})\big) - \mathbf{1}_{n_i}\cdot\tau\Big]\right| \\
&\le \max_{1\le j\le p}\frac{1}{M}\sum_{i=1}^{M}\|\boldsymbol{x}_{i,j}\|_2\cdot\left\|\boldsymbol{A}_i^{-1/2}\left(\widehat{\boldsymbol{R}}_i^{-1} - \boldsymbol{R}_i^{-1}\right)\boldsymbol{A}_i^{-1/2}\right\|_2\cdot\left\|\boldsymbol{K}\left(-\frac{\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta}}{h}\right) - \mathbf{1}_{n_i}\cdot\tau\right\|_2 \\
&\le \frac{1}{\tau(1-\tau)}\cdot\frac{1}{M}\sum_{i=1}^{M}\|\boldsymbol{x}_{i,j}\|_2\cdot\left\|\widehat{\boldsymbol{R}}_i^{-1} - \boldsymbol{R}_i^{-1}\right\|_2\cdot\sqrt{n_u}\max_{i,k}\left|K\left(-\frac{y_{ik} - \boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\beta}}{h}\right) - \tau\right| \\
&\le \frac{1}{\tau(1-\tau)}\cdot\frac{1}{M}\sum_{i=1}^{M}\sqrt{n_u}\max_{i,k,j}|x_{ik,j}|\cdot C_\varepsilon\sqrt{\frac{p}{M}}\cdot\sqrt{n_u}\cdot\bar{\tau} \\
&\le (1-\bar{\tau})^{-1}n_u B C_\varepsilon\sqrt{\frac{p}{M}} = C_0 C_\varepsilon\sqrt{\frac{p}{M}},
\end{aligned}$$

where the identity $\tau^{-1}(1-\tau)^{-1}\bar{\tau} = (1-\bar{\tau})^{-1}$ follows from the definition of $\bar{\tau}$. With the same probability, we have

$$\begin{aligned}
\|\mathfrak{S}_M(\boldsymbol{\beta})\|_2 &= \left\|\boldsymbol{\Sigma}^{-1/2}\left(\widehat{\boldsymbol{U}}_M(\boldsymbol{\beta}) - \boldsymbol{U}_M(\boldsymbol{\beta})\right)\right\|_2 \le \xi_p^{-1/2}\left\|\widehat{\boldsymbol{U}}_M(\boldsymbol{\beta}) - \boldsymbol{U}_M(\boldsymbol{\beta})\right\|_2 \\
&\le \xi_p^{-1/2}\sqrt{p}\left\|\widehat{\boldsymbol{U}}_M(\boldsymbol{\beta}) - \boldsymbol{U}_M(\boldsymbol{\beta})\right\|_\infty \le \xi_p^{-1/2}C_0C_\varepsilon\sqrt{\frac{p^2}{M}}.
\end{aligned}$$

■

Lemma 3. *Assume that Conditions (A2)–(A7) hold. For any $\gamma > 0$, as long as $h \gtrsim \sqrt{(p+\gamma)/M}$,*

$$\begin{aligned}
\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\left\|\Delta_M(\boldsymbol{\beta})\right\|_2 &= \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\left\|\boldsymbol{\Sigma}^{-1/2}\Big\{\boldsymbol{U}_M(\boldsymbol{\beta}) - \boldsymbol{U}_M(\boldsymbol{\beta}_\tau) - \boldsymbol{H}_M(\boldsymbol{\beta}_\tau)(\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\Big\}\right\|_2 \\
&\lesssim C_1 r\left(\sqrt{\frac{p+\gamma}{Mh}} + r\right)
\end{aligned}$$

holds with probability at least $1-e^{-\gamma}$, where $C_1 > 0$ is a constant depending on $(\tau, L_0, n_u, f_u, \kappa_u, a_1, \mu_3, \xi_{\min})$.

Proof. Similar to the proof of Proposition S5 in Song et al. (2024) [6], we just need to derive the upper bound of

$$\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\left\|\Delta_M(\boldsymbol{\beta})\right\|_2 \le \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\left\|E\big[\Delta_M(\boldsymbol{\beta})\big]\right\|_2 + \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\left\|\Delta_M(\boldsymbol{\beta}) - E\big[\Delta_M(\boldsymbol{\beta})\big]\right\|_2. \quad (13)$$

According to the mean-value theorem of vector-valued functions, for $\boldsymbol{\zeta} = \boldsymbol{\beta} - \boldsymbol{\beta}_\tau$ with $\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)$,

$$\begin{aligned}
E\big[\Delta_M(\boldsymbol{\beta})\big] &= \boldsymbol{\Sigma}^{-1/2}\left\langle\int_0^1\boldsymbol{H}_M\big(\boldsymbol{\beta}_\tau + t\boldsymbol{\zeta}\big)\,dt,\ \boldsymbol{\zeta}\right\rangle - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{H}_M(\boldsymbol{\beta}_\tau)\cdot\boldsymbol{\zeta} \\
&= \left\langle\boldsymbol{\Sigma}^{-1/2}\int_0^1\boldsymbol{H}_M\big(\boldsymbol{\beta}_\tau + t\boldsymbol{\zeta}\big)\,dt\cdot\ \boldsymbol{\Sigma}^{-1/2} - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{H}_M(\boldsymbol{\beta}_\tau)\cdot\boldsymbol{\Sigma}^{-1/2},\ \boldsymbol{\Sigma}^{1/2}\boldsymbol{\zeta}\right\rangle.
\end{aligned} \quad (14)$$

Denote $\boldsymbol{z}_i = \boldsymbol{x}_i \boldsymbol{\Sigma}^{-1/2}$ and $\boldsymbol{\delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\zeta}$, we have

$$\boldsymbol{\Sigma}^{-1/2}\boldsymbol{H}_M(\boldsymbol{\beta}_\tau)\cdot\boldsymbol{\Sigma}^{-1/2} = \frac{1}{M}\sum_{i=1}^{M} E\Big\{\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{A}_i^{-1/2}\boldsymbol{R}_i^{-1}\boldsymbol{A}_i^{-1/2}\boldsymbol{\Lambda}_i(\boldsymbol{\beta}_\tau)\boldsymbol{z}_i\Big\}. \tag{15}$$

Since for $\boldsymbol{z}_{ik}$ $(1 \le i \le M,\ 1 \le k \le n_i)$,

$$\begin{aligned} E\Big\{\boldsymbol{z}_{ik}\Lambda_{ik}(\boldsymbol{\beta}_\tau + t\boldsymbol{\zeta})\boldsymbol{z}_{ik}^{\mathrm{T}}\Big\} &= \frac{1}{h}E\left\{k\left(\frac{\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}\cdot t - \epsilon_{ik}}{h}\right)\cdot \boldsymbol{z}_{ik}\boldsymbol{z}_{ik}^{\mathrm{T}}\right\} \\ &= E\left[\int_{-\infty}^{+\infty} k(u) f_{\epsilon|\boldsymbol{x}}\big(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{\delta}\cdot t - hu\big)du\cdot \boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}\right], \end{aligned} \tag{16}$$

where $\Lambda_{ik}(\cdot)$ is the kth diagonal element of $\boldsymbol{\Lambda}_i(\cdot)$ and $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}$. According to (15) and (16), For $\boldsymbol{\delta}$ which satisfies $\|\boldsymbol{\delta}\|_2 \le r$,

$$\begin{aligned} &\Big\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{H}_M(\boldsymbol{\beta}_\tau + t\boldsymbol{\zeta})\cdot\boldsymbol{\Sigma}^{-1/2} - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{H}_M(\boldsymbol{\beta}_\tau)\cdot\boldsymbol{\Sigma}^{-1/2}\Big\|_2 \\ &= \left\|\frac{1}{M}\sum_{i=1}^{M} E\Big\{\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{A}_i^{-1/2}\boldsymbol{R}_i^{-1}\boldsymbol{A}_i^{-1/2}\big[\boldsymbol{\Lambda}_i(\boldsymbol{\beta}_\tau + t\boldsymbol{\zeta}) - \boldsymbol{\Lambda}_i(\boldsymbol{\beta}_\tau)\big]\boldsymbol{z}_i\Big\}\right\|_2 \\ &\le \frac{1}{\tau(1-\tau)\xi_{\min}}\left\|\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{n_i} E\Big\{\boldsymbol{z}_{ik}\big[\Lambda_{ik}(\boldsymbol{\beta}_\tau + t\boldsymbol{\zeta}) - \Lambda_{ik}(\boldsymbol{\beta}_\tau)\big]\boldsymbol{z}_{ik}^{\mathrm{T}}\Big\}\right\|_2 \\ &\le \frac{n_u}{\tau(1-\tau)\xi_{\min}} E\left\|\int_{-\infty}^{+\infty} k(u)\big[f_{\epsilon|\boldsymbol{x}}\big(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{\delta}\cdot t - hu\big) - f_{\epsilon|\boldsymbol{x}}(-hu)\big]du\cdot\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}\right\|_2 \\ &\le \frac{n_u L_0 t}{C_r}\sup_{\|\boldsymbol{u}\|_2=1}\Big\{E\big(\langle\boldsymbol{z},\boldsymbol{u}\rangle^2|\langle\boldsymbol{z},\boldsymbol{v}\rangle|\big)\Big\}\cdot\|\boldsymbol{\delta}\|_2 \\ &\le \frac{n_u\mu_3 L_0 r t}{C_r}, \end{aligned} \tag{17}$$

where $C_r = \tau(1-\tau)\xi_{\min}$. Together with (14) and (17),

$$\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\Big\|E\big[\Delta_M(\boldsymbol{\beta})\big]\Big\|_2 \le 0.5\ C_r^{-1}L_0 n_u\mu_3 r^2. \tag{18}$$

At follows, we derive the upper bound of $\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\big\|\Delta_M(\boldsymbol{\beta}) - E\big[\Delta_M(\boldsymbol{\beta})\big]\big\|_2$. Define the centralized gradient process

$$\Psi_M(\boldsymbol{\beta}) = \boldsymbol{\Sigma}^{-1/2}\big[\boldsymbol{U}_M(\boldsymbol{\beta}) - \boldsymbol{J}_M(\boldsymbol{\beta})\big]$$

such that $\Delta_M(\boldsymbol{\beta}) - E\big[\Delta_M(\boldsymbol{\beta})\big] = \Psi_M(\boldsymbol{\beta}) - \Psi_M(\boldsymbol{\beta}_\tau)$. For $\boldsymbol{\delta}$ which satisfies $\|\boldsymbol{\delta}\|_2 \le r$,

$$\begin{aligned} \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\Big\|\Delta_M(\boldsymbol{\beta}) - E\big[\Delta_M(\boldsymbol{\beta})\big]\Big\|_2 &= \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\Big\|\Psi_M(\boldsymbol{\beta}) - \Psi_M(\boldsymbol{\beta}_\tau)\Big\|_2 \\ &= \sup_{\|\boldsymbol{\delta}\|_2\le r}\Big\|\Psi_M\big(\boldsymbol{\beta}_\tau + \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}\big) - \Psi_M(\boldsymbol{\beta}_\tau)\Big\|_2. \end{aligned}$$

Let $\Delta_{M,0}(\boldsymbol{\delta}) = \Psi_M\big(\boldsymbol{\beta}_\tau + \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}\big) - \Psi_M(\boldsymbol{\beta}_\tau)$, then $\Delta_{M,0}(\boldsymbol{0}) = \boldsymbol{0}$ and $E\big[\Delta_{M,0}(\boldsymbol{\delta})\big] = \boldsymbol{0}$. We now drive the upper bound of $\sup_{\|\boldsymbol{\delta}\|_2\le r}\Delta_{M,0}(\boldsymbol{\delta})$ based on Theorem A.3 in Spokoiny

(2012) [7]. Take the first-order derivative of $\Delta_{M,0}(\boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$, we have

$$
\begin{aligned}
\nabla\Delta_{M,0}(\boldsymbol{\delta}) &= \boldsymbol{\Sigma}^{-1/2}\Big[\boldsymbol{D}_M\big(\boldsymbol{\beta}_\tau+\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}\big)-\boldsymbol{H}_M\big(\boldsymbol{\beta}_\tau+\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}\big)\Big]\cdot\boldsymbol{\Sigma}^{-1/2}\\
&\le \frac{1}{C_r}\frac{1}{M}\sum_{i=1}^{M}\Big\{\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i\big(\boldsymbol{\beta}_\tau+\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}\big)\boldsymbol{z}_i-E\big[\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i\big(\boldsymbol{\beta}_\tau+\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}\big)\boldsymbol{z}_i\big]\Big\}\\
&= \frac{1}{C_r}\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{n_i}\Big\{\boldsymbol{z}_{ik}\Lambda_{ik}\big(\boldsymbol{\beta}_\tau+\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}\big)\boldsymbol{z}_{ik}^{\mathrm{T}}-E\big[\boldsymbol{z}_{ik}\Lambda_{ik}\big(\boldsymbol{\beta}_\tau+\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}\big)\boldsymbol{z}_{ik}^{\mathrm{T}}\big]\Big\}\\
&= \frac{1}{C_r}\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{n_i}\Big\{k_h\big(\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}-\epsilon_{ik}\big)\boldsymbol{z}_{ik}\boldsymbol{z}_{ik}^{\mathrm{T}}-E\big[k_h\big(\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}-\epsilon_{ik}\big)\boldsymbol{z}_{ik}\boldsymbol{z}_{ik}^{\mathrm{T}}\big]\Big\},
\end{aligned}
$$

where $k_h(\cdot)=(1/h)k(\cdot)$. Write $k_{ik,\delta}=k_h(\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}-\epsilon_{ik})$, then $0\le k_{ik,\delta}\le \kappa_u/h$. According to the proof of Proposition S5 in the supplementary of Song et al. (2024) [6] and the elementary inequality $|e^u-1-u|\le u^2e^{|u|}/2$, for any $\boldsymbol{u},\boldsymbol{u}'\in\mathbb{S}^{p-1}$ and $\lambda\in\mathbb{R}$,

$$
\begin{aligned}
&E\Big[\exp\Big\{\lambda\sqrt{M}\langle\boldsymbol{u},\nabla\Delta_{M,0}(\boldsymbol{\delta})\boldsymbol{u}'\rangle/a_1^2\Big\}\Big]\\
&\le E\left[\exp\left\{\frac{\lambda\sqrt{M}}{a_1^2C_r}\left\langle\boldsymbol{u},\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{n_i}\Big\{k_{ik,\delta}\boldsymbol{z}_{ik}\boldsymbol{z}_{ik}^{\mathrm{T}}-E\big(k_{ik,\delta}\boldsymbol{z}_{ik}\boldsymbol{z}_{ik}^{\mathrm{T}}\big)\Big\}\boldsymbol{u}'\right\rangle\right\}\right]\\
&=\prod_{i=1}^{M}E\left[\exp\left\{\frac{\lambda\sqrt{M}}{a_1^2C_rM}\sum_{k=1}^{n_i}\Big\langle\boldsymbol{u},\Big\{k_{ik,\delta}\boldsymbol{z}_{ik}\boldsymbol{z}_{ik}^{\mathrm{T}}-E\big(k_{ik,\delta}\boldsymbol{z}_{ik}\boldsymbol{z}_{ik}^{\mathrm{T}}\big)\Big\}\boldsymbol{u}'\Big\rangle\right\}\right]\\
&\le\prod_{i=1}^{M}\left\{1+E\left[\frac{\lambda\sqrt{M}}{a_1^2C_rM}\sum_{k=1}^{n_i}\Big\langle\boldsymbol{u},\Big\{k_{ik,\delta}\boldsymbol{z}_{ik}\boldsymbol{z}_{ik}^{\mathrm{T}}-E\big(k_{ik,\delta}\boldsymbol{z}_{ik}\boldsymbol{z}_{ik}^{\mathrm{T}}\big)\Big\}\boldsymbol{u}'\Big\rangle\right]\right.\\
&\quad+\frac{\lambda^2M}{2a_1^4C_r^2M^2}E\left[\left(\sum_{k=1}^{n_i}\Big\langle\boldsymbol{u},\Big\{k_{ik,\delta}\boldsymbol{z}_{ik}\boldsymbol{z}_{ik}^{\mathrm{T}}-E\big(k_{ik,\delta}\boldsymbol{z}_{ik}\boldsymbol{z}_{ik}^{\mathrm{T}}\big)\Big\}\boldsymbol{u}'\Big\rangle\right)^2\right.\\
&\quad\left.\left.\times\exp\left\{\left|\frac{\lambda\sqrt{M}}{a_1^2C_rM}\sum_{k=1}^{n_i}\Big\langle\boldsymbol{u},\Big\{k_{ik,\delta}\boldsymbol{z}_{ik}\boldsymbol{z}_{ik}^{\mathrm{T}}-E\big(k_{ik,\delta}\boldsymbol{z}_{ik}\boldsymbol{z}_{ik}^{\mathrm{T}}\big)\Big\}\boldsymbol{u}'\Big\rangle\right|\right\}\right]\right\}\\
&\le\prod_{i=1}^{M}\left\{1+\frac{\lambda^2Mn_i^2}{2a_1^4C_r^2M^2}E\left[\frac{1}{n_i}\sum_{k=1}^{n_i}\Big\{k_{ik,\delta}\langle\boldsymbol{z}_{ik},\boldsymbol{u}\rangle\langle\boldsymbol{z}_{ik},\boldsymbol{u}'\rangle-E\big(k_{ik,\delta}\langle\boldsymbol{z}_{ik},\boldsymbol{u}\rangle\langle\boldsymbol{z}_{ik},\boldsymbol{u}'\rangle\big)\Big\}^2\right.\right.\\
&\quad\left.\left.\times\exp\left\{\frac{\kappa_u|\lambda|\sqrt{M}n_u}{ha_1^2C_rM}\frac{1}{n_i}\sum_{k=1}^{n_i}\big|\langle\boldsymbol{z}_{ik},\boldsymbol{u}\rangle\langle\boldsymbol{z}_{ik},\boldsymbol{u}'\rangle\big|\right\}\exp\left\{\frac{f_u|\lambda|\sqrt{M}n_u}{a_1^2C_rM}\Big|E\big[\langle\boldsymbol{z}_{ik},\boldsymbol{u}\rangle\langle\boldsymbol{z}_{ik},\boldsymbol{u}'\rangle\big]\Big|\right\}\right]\right\}\\
&\le\prod_{i=1}^{M}\left\{1+\frac{\lambda^2Mn_u^2}{2a_1^4C_r^2M^2}E\left[\frac{1}{n_i}\sum_{k=1}^{n_i}\Big\{k_{ik,\delta}\langle\boldsymbol{z}_{ik},\boldsymbol{u}\rangle\langle\boldsymbol{z}_{ik},\boldsymbol{u}'\rangle-E\big(k_{ik,\delta}\langle\boldsymbol{z}_{ik},\boldsymbol{u}\rangle\langle\boldsymbol{z}_{ik},\boldsymbol{u}'\rangle\big)\Big\}^2\right.\right.\\
&\quad\left.\left.\times\exp\left\{\frac{\kappa_u|\lambda|\sqrt{M}n_u}{ha_1^2C_rM}\big|\langle\boldsymbol{z}_{ik'},\boldsymbol{u}\rangle\langle\boldsymbol{z}_{ik'},\boldsymbol{u}'\rangle\big|\right\}\exp\left\{\frac{f_u|\lambda|\sqrt{M}n_u}{a_1^2C_rM}\right\}\right]\right\},
\end{aligned}
$$

where $k'=\mathrm{argmax}_{1\le k\le n_i}\big|\langle\boldsymbol{z}_{ik},\boldsymbol{u}\rangle\langle\boldsymbol{z}_{ik},\boldsymbol{u}'\rangle\big|$. Let $|\lambda|\le\min\big\{h/(4n_u\kappa_u\sqrt{M}),1/(n_uf_u\sqrt{M})\big\}C_rM$, then $\kappa_u|\lambda|\sqrt{M}n_u/(hC_rM) \le 1/4$ and $f_u|\lambda|\sqrt{M}n_u/(a_1^2C_rM) \le 1$. Consequently,

$$\begin{aligned}
&E\Big[\exp\Big\{\lambda\sqrt{M}\langle \boldsymbol{u}, \nabla\Delta_{M,0}(\boldsymbol{\delta})\boldsymbol{u'}\rangle/a_1^2\Big\}\Big] \\
&\le \prod_{i=1}^M \Bigg\{1 + \frac{\lambda^2 M n_u^2 e}{2a_1^4C_r^2M^2}E\bigg[\frac{1}{n_i}\sum_{k=1}^{n_i}\Big\{k_{ik,\delta}\langle \boldsymbol{z}_{ik}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik}, \boldsymbol{u'}\rangle - E\big(k_{ik,\delta}\langle \boldsymbol{z}_{ik}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik}, \boldsymbol{u'}\rangle\big)\Big\}^2 \\
&\quad \times \exp\Big\{\big|\langle \boldsymbol{z}_{ik'}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik'}, \boldsymbol{u'}\rangle\big|/(4a_1^2)\Big\}\bigg]\Bigg\} \\
&\le \prod_{i=1}^M \Bigg\{1 + \frac{\lambda^2 M n_u^2 e}{a_1^4C_r^2M^2}E\bigg[\frac{1}{n_i}\sum_{k=1}^{n_i}\Big\{\big(k_{ik,\delta}\langle \boldsymbol{z}_{ik}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik}, \boldsymbol{u'}\rangle\big)^2 + \big[E\big(k_{ik,\delta}\langle \boldsymbol{z}_{ik}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik}, \boldsymbol{u'}\rangle\big)\big]^2\Big\} \\
&\quad \times \exp\Big\{\big|\langle \boldsymbol{z}_{ik'}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik'}, \boldsymbol{u'}\rangle\big|/(4a_1^2)\Big\}\bigg]\Bigg\} \\
&= \prod_{i=1}^M \Bigg\{1 + \frac{\lambda^2 M n_u^2 e}{a_1^4C_r^2M^2}E\bigg\{\frac{1}{n_i}\sum_{k=1}^{n_i}\big(k_{ik,\delta}\langle \boldsymbol{z}_{ik}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik}, \boldsymbol{u'}\rangle\big)^2 e^{|\langle \boldsymbol{z}_{ik'}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik'}, \boldsymbol{u'}\rangle|/(4a_1^2)}\bigg\} \\
&\quad + \frac{\lambda^2 M n_u^2 e}{a_1^4C_r^2M^2}\Big\{E\big(k_{ik,\delta}\langle \boldsymbol{z}_{ik}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik}, \boldsymbol{u'}\rangle\big)\Big\}^2 E\Big[e^{|\langle \boldsymbol{z}_{ik'}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik'}, \boldsymbol{u'}\rangle|/(4a_1^2)}\Big]\Bigg\} \\
&\le \prod_{i=1}^M \Bigg\{1 + \frac{\lambda^2 M n_u^2 e}{a_1^4C_r^2M^2}\frac{1}{n_i}\sum_{k=1}^{n_i}E\Big\{\big(k_{ik,\delta}\langle \boldsymbol{z}_{ik}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik}, \boldsymbol{u'}\rangle\big)^2 e^{|\langle \boldsymbol{z}_{ik'}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik'}, \boldsymbol{u'}\rangle|/(4a_1^2)}\Big\} \\
&\quad + \frac{\lambda^2 M n_u^2 e}{a_1^4C_r^2M^2}f_u E\Big[e^{\langle \boldsymbol{z}_{ik'}, \boldsymbol{u}\rangle^2/(2\cdot 4a_1^2) + \langle \boldsymbol{z}_{ik'}, \boldsymbol{u'}\rangle^2/(2\cdot 4a_1^2)}\Big]\Bigg\} \\
&\le \prod_{i=1}^M \Bigg\{1 + \frac{\lambda^2 M n_u^2 e\kappa_u}{a_1^4C_r^2M^2h}E\Big\{\big(\langle \boldsymbol{z}_{ik'}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik'}, \boldsymbol{u'}\rangle\big)^2 e^{|\langle \boldsymbol{z}_{ik'}, \boldsymbol{u}\rangle\langle \boldsymbol{z}_{ik'}, \boldsymbol{u'}\rangle|/(4a_1^2)}\Big\} + 3\frac{\lambda^2 M n_u^2 e f_u}{a_1^4C_r^2M^2}\Bigg\} \\
&\le \prod_{i=1}^M \Bigg\{1 + \frac{\lambda^2 M n_u^2 e}{a_1^4C_r^2M^2}\frac{\kappa_u}{h}\Big[E\langle \boldsymbol{z}_{ik'}, \boldsymbol{u}\rangle^4 e^{\langle \boldsymbol{z}_{ik'}, \boldsymbol{u}\rangle^2/(4a_1^2)}\Big]^{1/2}\Big[E\langle \boldsymbol{z}_{ik'}, \boldsymbol{u'}\rangle^4 e^{\langle \boldsymbol{z}_{ik'}, \boldsymbol{u'}\rangle^2/(4a_1^2)}\Big]^{1/2} \\
&\quad + 3\frac{\lambda^2 M n_u^2 e f_u}{a_1^4C_r^2M^2}\Bigg\} \\
&\le \prod_{i=1}^M \Bigg\{1 + 8\frac{\lambda^2 M n_u^2 e\kappa_u}{a_1^4C_r^2M^2h}16a_1^4 + 3\frac{\lambda^2 M n_u^2 e f_u}{a_1^4C_r^2M^2}\Bigg\} \le \prod_{i=1}^M \Bigg\{1 + \frac{C^2 M n_u^2\lambda^2}{2C_r^2M^2h}\Bigg\} \\
&\le \prod_{i=1}^M \exp\Bigg\{\frac{C^2 n_u^2\lambda^2}{2C_r^2Mh}\Bigg\} = \exp\Bigg\{\frac{C^2 n_u^2\lambda^2}{2C_r^2h}\Bigg\},
\end{aligned}$$

where $C > 0$ depends on (f_u, κ_u).

Let $v_0 = Cn_uC_r^{-1}h^{-1/2}$ and $g = \min\big\{h/(4n_u\kappa_u), 1/(n_uf_u)\big\}C_r\sqrt{M/2}$. After applying Theorem A.3 of Spokoiny (2013) [8] to the process $\big\{\sqrt{M}\Delta_{M,0}(\boldsymbol{\delta})/a_1^2, \boldsymbol{\delta} \in \mathbb{B}^p(r)\big\}$ with $\mathbb{B}^p(r) = \big\{\boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}\|_2 \le r\big\}$, we have

$$\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \Big\|\Delta_M(\boldsymbol{\beta}) - E\big[\Delta_M(\boldsymbol{\beta})\big]\Big\|_2 \le 6\ C_r^{-1}Cn_ura_1^2\sqrt{\frac{4p+2\gamma}{Mh}} \tag{19}$$

with probability at least $1 - e^{-\gamma}$ for any $\gamma \ge 0$ as long as $h \ge 8\ C_r^{-1}\kappa_un_u\sqrt{(2p+\gamma)/M}$

and $M \geq 4\ C_r^{-2} f_u^2 n_u^2 (2p+\gamma)$. Together with (13), (18) and (19), for any $\gamma \geq 0$,

$$\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \left\|\Delta_M(\boldsymbol{\beta})\right\|_2 \leq 0.5\ C_r^{-1} n_u \mu_3 L_0 r^2 + 6\ C_r^{-1} C n_u a_1^2 r \sqrt{\frac{4p+2\gamma}{Mh}}$$

holds with probability at least $1-e^{-\gamma}$.

■

Lemma 4. *Assume that Conditions (A2)–(A7) hold. For any $\gamma > 0$, as long as $M \gtrsim p+\gamma$,*

$$\left\|\mathfrak{T}_M(\boldsymbol{\beta}_\tau)\right\|_2 = \left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{U}_M(\boldsymbol{\beta}_\tau)\right\|_2 \lesssim C_2\left(\sqrt{\frac{p+\gamma}{M}} + h^2\right)$$

holds with probability at least $1-e^{-\gamma}$, where $C_2 > 0$ is a constant depending on $(\tau, L_0, n_u, \kappa_2, a_1, h, \xi_{\min})$.

Proof. Assume that $\boldsymbol{x}_{ik}$ $(1 \leq i \leq M, 1 \leq k \leq n_i)$ satisfies Condition (A7). Let $\varpi_{ik} = K\big(-(y_{ik} - \boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\beta}_\tau)/h\big) - \tau$ and $\boldsymbol{\varpi}_i = (\varpi_{i1}, \ldots, \varpi_{in_i})^{\mathrm{T}} = \boldsymbol{K}\big(-(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta}_\tau)/h\big) - \mathbf{1}_{n_i}\cdot\tau$. Denote $\boldsymbol{\Omega}_i = \boldsymbol{A}_i^{-1/2}\boldsymbol{R}_i^{-1}\boldsymbol{A}_i^{-1/2}$. According to He et al. (2023)[4], for any $\varepsilon \in (0,1)$, there exists an ε-net $\mathcal{N}_\varepsilon$ of the unit sphere $\mathbb{S}^{p-1}$ with cardinality $|\mathcal{N}_\varepsilon| \leq (1+2/\varepsilon)^p$ such that

$$\begin{aligned}
&\left\|\boldsymbol{\Sigma}^{-1/2}\Big\{\boldsymbol{U}_M(\boldsymbol{\beta}_\tau) - E\big[\boldsymbol{U}_M(\boldsymbol{\beta}_\tau)\big]\Big\}\right\|_2 = \left\|\frac{1}{M}\sum_{i=1}^M\Big\{\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\varpi}_i - E\big[\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\varpi}_i\big]\Big\}\right\|_2 \\
&\leq (1-\varepsilon)^{-1}\max_{\boldsymbol{u}\in\mathcal{N}_\varepsilon}\left\langle \boldsymbol{u}, \frac{1}{M}\sum_{i=1}^M\Big\{\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\varpi}_i - E\big[\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\varpi}_i\big]\Big\}\right\rangle.
\end{aligned}$$

Let $\mathcal{X}_{\boldsymbol{u},i} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\varpi}_i$ for each direction $\boldsymbol{u} \in \mathcal{N}_\varepsilon$, we have

$$\begin{aligned}
\sum_{i=1}^M E\big(\mathcal{X}_{\boldsymbol{u},i}\big)^2 &= \sum_{i=1}^M E\big(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\varpi}_i\big)^2 \leq \frac{1}{C_r^2}\sum_{i=1}^M E\big(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\varpi}_i\big)^2 \\
&\leq \frac{1}{C_r^2}\sum_{i=1}^M E\left(\sum_{k=1}^{n_i}\langle\boldsymbol{u}, \boldsymbol{z}_{ik}\varpi_{ik}\rangle\right)^2 \leq \frac{n_u^2}{C_r^2}\sum_{i=1}^M E\big[\langle\boldsymbol{u},\boldsymbol{z}\rangle^2\varpi^2\big],
\end{aligned}$$

where $\varpi = K(-\epsilon/h) - \tau$ with ϵ the random variable of $\epsilon_{ik} = y_{ik} - \boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\beta}_\tau$. At follows, we derive the upper bound for $E(\varpi^2|\boldsymbol{x})$, which involves calculating $E\big[K^2(-\epsilon/h)|\boldsymbol{x}\big]$ and $E\big[K(-\epsilon/h)|\boldsymbol{x}\big]$. Note that

$$\begin{aligned}
E\left[K^2\left(\frac{-\epsilon}{h}\right)\middle|\boldsymbol{x}\right] &= \int_{-\infty}^{+\infty} K^2\left(\frac{-t}{h}\right) f_{\epsilon|\boldsymbol{x}}(t)dt = h\int_{-\infty}^{+\infty} K^2(u) f_{\epsilon|\boldsymbol{x}}(-uh)\,du \\
&= -\int_{-\infty}^{+\infty} K^2(u)dF_{\epsilon|\boldsymbol{x}}(-uh) = 2\int_{-\infty}^{+\infty} K(u)k(u)F_{\epsilon|\boldsymbol{x}}(-uh)\,du. \quad (20)
\end{aligned}$$

Let

$$0 < u_k = \int_{-\infty}^{+\infty} uk(u)K(u)du = \int_0^{+\infty} k(u)(1-k(u))du \leq \kappa_1. \quad (21)$$

Since $F_{\epsilon|\boldsymbol{x}}(0) = P(\epsilon \le 0|\boldsymbol{x}) = \tau$ and

$$
\begin{aligned}
F_{\epsilon|\boldsymbol{x}}(-uh) &= F_{\epsilon|\boldsymbol{x}}(0) + \int_0^{-uh} f_{\epsilon|\boldsymbol{x}}(t)dt \\
&= \tau + (-uh)\, f_{\epsilon|\boldsymbol{x}}(0) + \int_0^{-uh} \left[f_{\epsilon|\boldsymbol{x}}(t) - f_{\epsilon|\boldsymbol{x}}(0)\right] dt. \qquad (22)
\end{aligned}
$$

Based on (20)–(22) and Conditions (A5) and (A6), we have

$$
\begin{aligned}
E\left[K^2\left(-\frac{\epsilon}{h}\right)\bigg|\boldsymbol{x}\right] &= 2\tau\int_{-\infty}^{+\infty} k(u)K(u)du - 2hf_{\epsilon|\boldsymbol{x}}(0)\int_{-\infty}^{+\infty} uk(u)K(u)du \\
&\quad + 2\int_{-\infty}^{+\infty} k(u)K(u)\int_0^{-uh}[f_{\epsilon|\boldsymbol{x}}(t) - f_{\epsilon|\boldsymbol{x}}(0)]dtdu \\
&\le \tau - 2hf_{\epsilon|\boldsymbol{x}}(0)u_k + L_0h^2\int_{-\infty}^{+\infty} u^2k(u)K(u)du \\
&\le \tau + L_0\kappa_2h^2. \qquad (23)
\end{aligned}
$$

Similarly to (22),

$$
E\left[K\left(\frac{-\epsilon}{h}\right)\bigg|\boldsymbol{x}\right] = \tau + \int_{-\infty}^{+\infty} k(u)\int_0^{-uh}\left[f_{\epsilon|\boldsymbol{x}}(t) - f_{\epsilon|\boldsymbol{x}}(0)\right]dtdu.
$$

Based on the Lipschitz condition of $f_{\epsilon|\boldsymbol{x}}(\cdot)$,

$$
\left|E\left[K\left(\frac{-\epsilon}{h}\right)\bigg|\boldsymbol{x}\right] - \tau\right| \le 0.5L_0\kappa_2h^2.
$$

Therefore,

$$
\tau - 0.5L_0\kappa_2h^2 \le E\left[K\left(\frac{-\epsilon}{h}\right)\bigg|\boldsymbol{x}\right] \le \tau + 0.5L_0\kappa_2h^2. \qquad (24)
$$

Together with (23) and (24),

$$
\begin{aligned}
E(\varpi^2|\boldsymbol{x}) &= E\left[K^2\left(\frac{-\epsilon}{h}\right)\bigg|\boldsymbol{x}\right] - 2\tau E\left[K\left(\frac{-\epsilon}{h}\right)\bigg|\boldsymbol{x}\right] + \tau^2 \\
&\le \tau + L_0\kappa_2h^2 - 2\tau\big(\tau - 0.5L_0\kappa_2h^2\big) + \tau^2 \\
&= \tau(1-\tau) + (\tau+1)L_0\kappa_2h^2 = C_\tau^2.
\end{aligned}
$$

Therefore, $\sum_{i=1}^M E\big(\mathcal{X}_{\boldsymbol{u},i}\big)^2 \le C_r^{-2}Mn_u^2C_\tau^2$. Since $|\varpi| \le \max(1-\tau,\tau) = \bar{\tau}$. Based on Condition (A7), for $\iota \ge 3$, we have $P(|x_j| \ge a_0\sigma_{jj}^{1/2}\gamma) \le e^{-\gamma}$, where a_0 is a constant depending on a_1. Therefore, for all $\gamma \ge 0$,

$$
\begin{aligned}
E|\varpi x_j|^\iota &= E_{\boldsymbol{x}}\left[|x_j|^\iota E\big(|\varpi|^\iota|\boldsymbol{x}\big)\right] \le \bar{\tau}^{\iota-2}E_{\boldsymbol{x}}\left[|x_j|^\iota E(\varpi^2|\boldsymbol{x})\right] \\
&\le \bar{\tau}^{\iota-2}C_\tau^2a_0^\iota\sigma_{jj}^{\iota/2}\int_0^{+\infty}\iota\gamma^{\iota-1}P(|x_j| \ge a_0\sigma_{jj}^{1/2}\gamma)d\gamma \\
&\le \bar{\tau}^{\iota-2}C_\tau^2a_0^\iota\sigma_{jj}^{\iota/2}\iota\int_0^{+\infty}\gamma^{\iota-1}e^{-\gamma}d\gamma = \bar{\tau}^{\iota-2}C_\tau^2a_0^\iota\sigma_{jj}^{\iota/2}\iota! \\
&\le \frac{\iota!}{2}C_\tau^2a_0^2\sigma_{jj}\left[2\bar{\tau}a_0\sigma_{jj}^{1/2}\right]^{\iota-2}.
\end{aligned}
$$

Thus, $E\big(\big|\langle \boldsymbol{u}, \boldsymbol{z}_{ik}\varpi_{ik}\rangle\big|^{\iota}\big) \le (\iota!/2)C_\tau^2 a_0^2(2\bar{\tau}a_0)^{\iota-2}$ for $\iota \ge 3$. This implies that for all integers $\iota \ge 3$,

$$\sum_{i=1}^{M} E\big(|\mathcal{X}_{\boldsymbol{u},i}|^{\iota}\big) = \sum_{i=1}^{M} E\Big(\big|\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\varpi}_i\big|^{\iota}\Big) \le \frac{1}{C_r^{\iota}}\sum_{i=1}^{M} E\bigg(\Big|\sum_{k=1}^{n_i}\langle \boldsymbol{u}, \boldsymbol{z}_{ik}\varpi_{ik}\rangle\Big|^{\iota}\bigg)$$
$$\le \frac{\iota!}{2}MC_r^{-2}n_u^2C_\tau^2a_0^2(2\bar{\tau}a_0n_uC_r^{-1})^{\iota-2}.$$

According to the Bernstein's inequality in Theorem 2.10 of Boucheron (2013) [1], for all $\gamma > 0$,

$$\max_{\boldsymbol{u}\in\mathcal{N}_\varepsilon}\left\langle \boldsymbol{u}, \frac{1}{M}\sum_{i=1}^{M}\Big\{\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\varpi}_i - E\big[\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\varpi}_i\big]\Big\}\right\rangle \le C_r^{-1}a_0n_u\left[C_\tau\sqrt{\frac{2\gamma}{M}} + \bar{\tau}\frac{2\gamma}{M}\right]$$

holds with probability at least $1 - e^{-\gamma}$. Applying a union bound over all vectors $\boldsymbol{u} \in \mathcal{N}_\varepsilon$,

$$\Big\|\boldsymbol{\Sigma}^{-1/2}\Big\{\boldsymbol{U}_M(\boldsymbol{\beta}_\tau) - E\big[\boldsymbol{U}_M(\boldsymbol{\beta}_\tau)\big]\Big\}\Big\|_2 \le \frac{a_0n_u}{(1-\varepsilon)C_r}\left[C_\tau\sqrt{\frac{2\gamma}{M}} + \bar{\tau}\frac{2\gamma}{M}\right]$$

holds with probability at least $1 - e^{\log(1+2/\varepsilon)p-\gamma}$. After taking $\varepsilon = 2/(e^2-1)$ and replacing γ by $2p+\gamma$, we have

$$\Big\|\boldsymbol{\Sigma}^{-1/2}\Big\{\boldsymbol{U}_M(\boldsymbol{\beta}_\tau) - E\big[\boldsymbol{U}_M(\boldsymbol{\beta}_\tau)\big]\Big\}\Big\|_2 \le 1.46\ C_r^{-1}a_0n_u\left[C_\tau\sqrt{\frac{4p+2\gamma}{M}} + \bar{\tau}\frac{4p+2\gamma}{M}\right] \quad (25)$$

with probability at least$1 - e^{-\gamma}$. Also because

$$\begin{aligned}\omega_M^* &= \Big\|\boldsymbol{\Sigma}^{-1/2}E\big[\boldsymbol{U}_M(\boldsymbol{\beta}_\tau)\big]\Big\|_2 = \Big\|\frac{1}{M}\sum_{i=1}^{M}E\big[\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\varpi}_i\big]\Big\|_2\\ &\le \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\left|\frac{1}{M}\sum_{i=1}^{M}\frac{n_u}{C_r}E\Big\{\big[K(-\epsilon/h)-\tau\big]\langle \boldsymbol{z}, \boldsymbol{u}\rangle\Big\}\right|\\ &\le \frac{n_u}{C_r}\sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\left\{E_{\boldsymbol{z}}\left|\int_{-\infty}^{+\infty}k(u)\int_0^{-hu}\big[f_{\epsilon|\boldsymbol{x}}(t) - f_{\epsilon|\boldsymbol{x}}(0)\big]dt\cdot du\cdot \boldsymbol{z}^{\mathrm{T}}\boldsymbol{u}\right|\right\}\\ &\le 0.5\ C_r^{-1}n_uL_0\kappa_2h^2. \end{aligned} \quad (26)$$

Together with (25) and (26), as long as $M \gtrsim p + \gamma$,

$$\begin{aligned}\Big\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{U}_M(\boldsymbol{\beta}_\tau)\Big\|_2 &\le \Big\|\boldsymbol{\Sigma}^{-1/2}\Big\{\boldsymbol{U}_M(\boldsymbol{\beta}_\tau) - E\big[\boldsymbol{U}_M(\boldsymbol{\beta}_\tau)\big]\Big\}\Big\|_2 + \Big\|\boldsymbol{\Sigma}^{-1/2}E\big[\boldsymbol{U}_M(\boldsymbol{\beta}_\tau)\big]\Big\|_2\\ &\le 1.46\ C_r^{-1}a_0n_u\left[C_\tau\sqrt{\frac{4p+2\gamma}{M}} + \bar{\tau}\frac{4p+2\gamma}{M}\right] + 0.5\ C_r^{-1}n_uL_0\kappa_2h^2\\ &\lesssim C_2\left(\sqrt{\frac{p+\gamma}{M}} + h^2\right)\end{aligned}$$

holds with probability at least $1 - e^{-\gamma}$, where $C_2 > 0$ is a constant depending on $(\tau, L_0, n_u, \kappa_2, a_1, h, \xi_{\min})$. This completes the proof. ■

Lemma 5. *Under Conditions (A2)–(A6),*

$$\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \left\|\mathfrak{F}_M(\boldsymbol{\beta})\right\|_2 = \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \left\|\boldsymbol{\Sigma}^{-1/2}\Big[\boldsymbol{H}_M(\boldsymbol{\beta}) - \boldsymbol{H}_M(\boldsymbol{\beta}_\tau)\Big](\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2 \le C_3 r^2,$$

where $C_3 > 0$ is a constant depending only on $(\tau, L_0, n_u, \mu_3, \xi_{\min})$.

Proof. For $\boldsymbol{\zeta} = \boldsymbol{\beta} - \boldsymbol{\beta}_\tau$, $\boldsymbol{\delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\zeta}$, and $\boldsymbol{v} = \boldsymbol{\delta}/\|\boldsymbol{\delta}\|_2$, we have

$$\begin{aligned}
&\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \left\|\boldsymbol{\Sigma}^{-1/2}\Big[\boldsymbol{H}_M(\boldsymbol{\beta}) - \boldsymbol{H}_M(\boldsymbol{\beta}_\tau)\Big]\boldsymbol{\zeta}\right\|_2 \\
&= \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \left\|\frac{1}{M}\sum_{i=1}^M E\left\{\boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\big[\boldsymbol{\Lambda}_i(\boldsymbol{\beta}) - \boldsymbol{\Lambda}_i(\boldsymbol{\beta}_\tau)\big]\boldsymbol{x}_i\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\zeta}\right\}\right\|_2 \\
&= \sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \left|\frac{1}{M}\sum_{i=1}^M E\left[\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i \mathrm{diag}\left(\frac{1}{h}\left\{k\left(\frac{\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}-\epsilon_{ik}}{h}\right) - k\left(\frac{-\epsilon_{ik}}{h}\right)\right\}_{k=1}^{n_i}\right)\boldsymbol{z}_i\boldsymbol{\delta}\right]\right| \\
&\le \sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \frac{1}{M}\sum_{i=1}^M \frac{1}{C_r}\sum_{k=1}^{n_i} E\left|\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_{ik}\int_{-\infty}^{\infty} k(u)\left[f_{\epsilon|\boldsymbol{x}}(\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta} - hu) - f_{\epsilon|\boldsymbol{x}}(-hu)\right]du\cdot \boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}\right| \\
&\le \frac{n_u L_0}{C_r}\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\sup_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{S}^{p-1}} E\big(|\langle\boldsymbol{z},\boldsymbol{u}\rangle|\langle\boldsymbol{z},\boldsymbol{v}\rangle^2\big)\cdot\|\boldsymbol{\delta}\|_2^2 \quad \text{(by Conditions (A5) and (A6))} \\
&\le \frac{n_u L_0}{C_r}\mu_3\cdot r^2 = C_3 r^2.
\end{aligned}$$

■

Lemma 6. *Under Conditions (A2)–(A6),*

$$\begin{aligned}
\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \|\mathfrak{F}_{M,l}(\boldsymbol{\beta})\|_2 &= \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \left\|\boldsymbol{\Sigma}^{-1/2}\left[\boldsymbol{H}_M(\boldsymbol{\beta}) - \boldsymbol{H}_{m,l}(\boldsymbol{\beta})\right](\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2 \\
&\le \left[C_4(1-L^{-1}) + C_5|h_l - h|\right] r,
\end{aligned}$$

where $C_4 > 0$ is a constant depending only on $(\tau, n_u, f_u, \xi_{\min})$, $C_5 > 0$ is a constant depending only on $(\tau, n_u, L_0, \kappa_1, \xi_{\min})$.

Proof. Define $\boldsymbol{H}_{i,h}(\boldsymbol{\beta}) = E\left[\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\boldsymbol{x}_i\right]$, where the kth diagonal element of $\boldsymbol{\Lambda}_i(\boldsymbol{\beta})$ is $k\left((\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta} - \epsilon_{ik})/h\right)/h$. Then, we have the decomposition:

$$\begin{aligned}
&\left\|\boldsymbol{\Sigma}^{-1/2}\Big[\boldsymbol{H}_M(\boldsymbol{\beta}) - \boldsymbol{H}_{m,l}(\boldsymbol{\beta})\Big](\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2 \\
&\le \left\|\boldsymbol{\Sigma}^{-1/2}\left[\frac{1}{M}\sum_{i=1}^M \boldsymbol{H}_{i,h}(\boldsymbol{\beta}) - \frac{1}{m}\sum_{i\in\mathcal{I}_l}\boldsymbol{H}_{i,h}(\boldsymbol{\beta})\right](\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2 \\
&\quad + \left\|\boldsymbol{\Sigma}^{-1/2}\frac{1}{m}\sum_{i\in\mathcal{I}_l}\Big[\boldsymbol{H}_{i,h}(\boldsymbol{\beta}) - \boldsymbol{H}_{i,h_l}(\boldsymbol{\beta})\Big](\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2. \qquad (27)
\end{aligned}$$

Define the coefficients $a_{i,l} = 1/M - 1/m$ for $i \in \mathcal{I}_l$ and $a_{i,l} = 1/M$ for $i \notin \mathcal{I}_l$. Then, $\sum_{i=1}^M a_{i,l} = 0$ and $\sum_{i=1}^M |a_{i,l}| = \sum_{i\in\mathcal{I}_l}(1/m - 1/M) + \sum_{i\notin\mathcal{I}_l} 1/M = 2(1 - m/M)$. Thus,

$$\frac{1}{M}\sum_{i=1}^M \boldsymbol{H}_{i,h}(\boldsymbol{\beta}) - \frac{1}{m}\sum_{i\in\mathcal{I}_l}\boldsymbol{H}_{i,h}(\boldsymbol{\beta}) = \sum_{i=1}^M a_{i,l}\boldsymbol{H}_{i,h}(\boldsymbol{\beta}). \qquad (28)$$

Next, we bound the norm of each term in (27). First,

$$\begin{aligned}
&\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{H}_{i,h}(\boldsymbol{\beta})(\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2 \\
&= \sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)} \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \left| E\left[\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i \mathrm{diag}\left(\left\{\frac{1}{h}k\left(\frac{\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}-\epsilon_{ik}}{h}\right)\right\}_{k=1}^{n_i}\right)\boldsymbol{z}_i\right]\boldsymbol{\delta}\right| \\
&\le \sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)} \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \left|\frac{1}{C_r}\sum_{k=1}^{n_i} E\left[\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_{ik}\int_{-\infty}^{\infty} k(u) f_{\epsilon|\boldsymbol{x}}(\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}-hu)\,du\cdot \boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}\right]\right| \\
&\le \frac{f_u n_u}{C_r}\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)} \sup_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{S}^{p-1}} E\left(|\langle \boldsymbol{z},\boldsymbol{u}\rangle\langle \boldsymbol{z},\boldsymbol{v}\rangle|\right)\cdot\|\boldsymbol{\delta}\|_2 \le \frac{f_u n_u r}{C_r}.
\end{aligned} \tag{29}$$

From (28) and (29), we obtain:

$$\begin{aligned}
&\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \left\|\boldsymbol{\Sigma}^{-1/2}\left[\frac{1}{M}\sum_{i=1}^{M}\boldsymbol{H}_{i,h}(\boldsymbol{\beta})-\frac{1}{m}\sum_{i\in\mathcal{I}_l}\boldsymbol{H}_{i,h}(\boldsymbol{\beta})\right](\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2 \\
&\le \sum_{i=1}^{M}|a_{i,l}|\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{H}_{i,h}(\boldsymbol{\beta})(\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2 \\
&\le 2\left(1-\frac{m}{M}\right)\frac{f_u n_u r}{C_r} = C_4\left(1-\frac{m}{M}\right)r,
\end{aligned} \tag{30}$$

where $C_4 = 2\,C_r^{-1}f_u n_u$.

For the second term in (27), we have:

$$\begin{aligned}
&\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \left\|\boldsymbol{\Sigma}^{-1/2}\frac{1}{m}\sum_{i\in\mathcal{I}_l}\Big[\boldsymbol{H}_{i,h}(\boldsymbol{\beta})-\boldsymbol{H}_{i,h_l}(\boldsymbol{\beta})\Big](\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2 \\
&\le \sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)} \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \Bigg|\frac{1}{m}\sum_{i\in\mathcal{I}_l}\frac{1}{C_r}\sum_{k=1}^{n_i} E\Bigg[\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_{ik}\int_{-\infty}^{\infty} k(u)\big[f_{\epsilon|\boldsymbol{x}}(\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}-hu) \\
&\qquad - f_{\epsilon|\boldsymbol{x}}(\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}-h_l u)\big]\,du\cdot\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}\Bigg]\Bigg| \\
&\le \frac{n_u}{C_r}L_0\int_{-\infty}^{\infty}|u|k(u)\,du\cdot|h_l-h|\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)} \sup_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{S}^{p-1}} E\big(|\langle \boldsymbol{z},\boldsymbol{u}\rangle\langle \boldsymbol{z},\boldsymbol{v}\rangle|\big)\cdot\|\boldsymbol{\delta}\|_2 \\
&\le C_r^{-1}n_u L_0\kappa_1|h_l-h|r = C_5\,|h_l-h|r,
\end{aligned} \tag{31}$$

where $C_5 = C_r^{-1}n_u L_0\kappa_1$. Combining (30) and (31), we conclude that

$$\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \left\|\boldsymbol{\Sigma}^{-1/2}\Big[\boldsymbol{H}_M(\boldsymbol{\beta})-\boldsymbol{H}_{m,l}(\boldsymbol{\beta})\Big](\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2 \le \left[C_4(1-L^{-1})+C_5|h_l-h|\right]r.$$

■

Lemma 7. *Assume that Conditions (A2)–(A7) hold. For any $\gamma>0$, as long as $m\gtrsim p+\gamma$,*

$$\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\left\|\mathfrak{H}_{m,l}(\boldsymbol{\beta})\right\|_2 = \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\left\|\boldsymbol{\Sigma}^{-1/2}\Big[\boldsymbol{D}_{m,l}(\boldsymbol{\beta})-\boldsymbol{H}_{m,l}(\boldsymbol{\beta})\Big](\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2 \lesssim C_6\,r\sqrt{\frac{p+\gamma}{m}}$$

holds with probability at least $1-e^{-\gamma}$, where $C_6>0$ is a constant depending on $(\tau, n_u, f_u, a_1, \xi_{\min})$.

Proof. The proof is similar to that of Lemma 4. For any $\varepsilon \in (0,1)$, consider an ε-net $\mathcal{N}_\varepsilon$ of the unit sphere $\mathbb{S}^{p-1}$ with cardinality $|\mathcal{N}_\varepsilon| \le (1+2/\varepsilon)^p$. Then,

$$\begin{aligned}
&\left\| \boldsymbol{\Sigma}^{-1/2}\Big[\boldsymbol{D}_{m,l}(\boldsymbol{\beta}) - \boldsymbol{H}_{m,l}(\boldsymbol{\beta})\Big](\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2 \\
&= \left\| \frac{1}{m}\sum_{i\in\mathcal{I}_l}\Big\{\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\boldsymbol{z}_i - E\left[\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\boldsymbol{z}_i\right]\Big\}\boldsymbol{\delta}\right\|_2 \\
&\le (1-\varepsilon)^{-1}\max_{\boldsymbol{u}\in\mathcal{N}_\varepsilon}\left\langle \boldsymbol{u}, \frac{1}{m}\sum_{i\in\mathcal{I}_l}\Big\{\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\boldsymbol{z}_i - E\left[\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\boldsymbol{z}_i\right]\Big\}\boldsymbol{\delta}\right\rangle.
\end{aligned}$$

From Condition (A7), $P(|z_j| \ge a_0\gamma) \le e^{-\gamma}$, where z_j is the jth component of $\boldsymbol{z}$ and a_0 is a constant depending on a_1. Thus, for any integer $\iota \ge 2$,

$$E\left(|z_j|^\iota\right) = a_0^\iota \int_0^\infty \iota\gamma^{\iota-1}P(|z_j|\ge a_0\gamma)\,d\gamma \le a_0^\iota\int_0^\infty \iota\gamma^{\iota-1}e^{-\gamma}\,d\gamma = a_0^\iota \iota!,$$

which implies that for any $\boldsymbol{u}\in\mathcal{N}_\varepsilon$, $E\left(|\langle \boldsymbol{z},\boldsymbol{u}\rangle|^\iota\right) \le a_0^\iota \iota!$.

Define $\mathcal{X}_{\boldsymbol{u},i} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\boldsymbol{z}_i\boldsymbol{\delta}$ for each direction $\boldsymbol{u}\in\mathcal{N}_\varepsilon$. We bound the second moment:

$$\begin{aligned}
\max_{\boldsymbol{u}\in\mathcal{N}_\varepsilon}\sum_{i\in\mathcal{I}_l}E\left(\mathcal{X}_{\boldsymbol{u},i}\right)^2 &\le \sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\max_{\boldsymbol{u}\in\mathcal{N}_\varepsilon}\sum_{i\in\mathcal{I}_l}E\left[\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\boldsymbol{z}_i\boldsymbol{\delta}\right]^2 \\
&\le \frac{1}{C_r^2}\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\max_{\boldsymbol{u}\in\mathcal{N}_\varepsilon}\sum_{i\in\mathcal{I}_l}E\left[\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\boldsymbol{z}_i\boldsymbol{\delta}\right]^2 \\
&\le \frac{1}{C_r^2}\sum_{i\in\mathcal{I}_l}\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\max_{\boldsymbol{u}\in\mathcal{N}_\varepsilon}E\left[\sum_{k=1}^{n_i}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_{ik}\int_{-\infty}^{\infty}k(u)f_{\epsilon|\boldsymbol{x}}(\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}-h_lu)\,du\cdot\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}\right]^2 \\
&\le \frac{mf_u^2n_u^2}{C_r^2}\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\max_{\boldsymbol{u},\boldsymbol{v}\in\mathcal{N}_\varepsilon}E\big(\langle\boldsymbol{z},\boldsymbol{u}\rangle\langle\boldsymbol{z},\boldsymbol{v}\rangle\big)^2\cdot\|\boldsymbol{\delta}\|_2^2 \\
&\le \frac{mf_u^2n_u^2r^2}{C_r^2}\cdot 4!\,a_0^4.
\end{aligned}$$

For $\iota \ge 3$, the ιth moment is bounded by:

$$\begin{aligned}
\max_{\boldsymbol{u}\in\mathcal{N}_\varepsilon}\sum_{i\in\mathcal{I}_l}E\left(\mathcal{X}_{\boldsymbol{u},i}\right)^\iota &\le \sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\max_{\boldsymbol{u}\in\mathcal{N}_\varepsilon}\sum_{i\in\mathcal{I}_l}E\left[\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\boldsymbol{z}_i\boldsymbol{\delta}\right]^\iota \\
&\le \frac{1}{C_r^\iota}\sum_{i\in\mathcal{I}_l}\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\max_{\boldsymbol{u}\in\mathcal{N}_\varepsilon}E\left[\sum_{k=1}^{n_i}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_{ik}\int_{-\infty}^{\infty}k(u)f_{\epsilon|\boldsymbol{x}}(\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}-h_lu)\,du\cdot\boldsymbol{z}_{ik}^{\mathrm{T}}\boldsymbol{\delta}\right]^\iota \\
&\le \frac{mf_u^\iota n_u^\iota}{C_r^\iota}\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\max_{\boldsymbol{u},\boldsymbol{v}\in\mathcal{N}_\varepsilon}E\big(|\langle\boldsymbol{z},\boldsymbol{u}\rangle\langle\boldsymbol{z},\boldsymbol{v}\rangle|^\iota\big)\cdot\|\boldsymbol{\delta}\|_2^\iota \\
&\le \frac{mf_u^\iota n_u^\iota r^\iota}{C_r^\iota}(2\iota)!\,a_0^{2\iota} \\
&\le \frac{\iota!}{2}\cdot\frac{mf_u^2n_u^2r^2 4! a_0^4}{C_r^2}\left[\frac{f_un_ura_0^2}{C_r}\left(\frac{(2\iota)!}{12\iota!}\right)^{1/(\iota-2)}\right]^{\iota-2}.
\end{aligned}$$

Applying Bernstein's inequality [1], we obtain that with probability at least $1-e^{-\gamma}$,

$$\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\max_{\boldsymbol{u}\in\mathcal{N}_\varepsilon}\left\langle \boldsymbol{u}, \frac{1}{m}\sum_{i\in\mathcal{I}_l}\left\{\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\boldsymbol{z}_i - E\left[\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\boldsymbol{z}_i\right]\right\}\boldsymbol{\delta}\right\rangle$$
$$\le \frac{f_u n_u a_0^2 r}{C_r}\left[4\sqrt{\frac{3\gamma}{m}} + \left(\frac{(2\iota)!}{12\iota!}\right)^{1/(\iota-2)}\frac{\gamma}{m}\right] \lesssim Cr\left(\sqrt{\frac{\gamma}{m}}+\frac{\gamma}{m}\right).$$

Following the proof of Lemma 4, after choosing an appropriate ε, we can conclude that

$$\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\left\|\boldsymbol{\Sigma}^{-1/2}\Big[\boldsymbol{D}_{m,l}(\boldsymbol{\beta})-\boldsymbol{H}_{m,l}(\boldsymbol{\beta})\Big](\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2 \lesssim C_6 r\sqrt{\frac{p+\gamma}{m}}$$

holds with probability at least $1-e^{-\gamma}$, where $C_6>0$ is a constant depending on $(\tau, n_u, f_u, a_1, \xi_{\min})$. ■

Lemma 8. *Assume that Conditions (A1)–(A4) and (A6) hold. For any $\varepsilon>0$, there exist constants $C_\varepsilon, M_0>0$ such that for every $M\ge M_0$,*

$$\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\left\|\mathfrak{E}_{m,l}(\boldsymbol{\beta})\right\|_2 = \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\left\|\boldsymbol{\Sigma}^{-1/2}\left[\widehat{\boldsymbol{D}}_{m,l}(\boldsymbol{\beta})-\boldsymbol{D}_{m,l}(\boldsymbol{\beta})\right](\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2 \le C_7 C_\varepsilon\frac{r}{h_l}\sqrt{\frac{p^3}{M}}$$

holds with probability at least $1-\varepsilon$, where $C_7=\tau^{-1}(1-\tau)^{-1}\xi_p^{-1}\kappa_u n_u B^2$.

Proof. Under Conditions (A1)–(A4) and (A6),

$$\sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)}\left\|\boldsymbol{\Sigma}^{-1/2}\left[\widehat{\boldsymbol{D}}_{m,l}(\boldsymbol{\beta})-\boldsymbol{D}_{m,l}(\boldsymbol{\beta})\right](\boldsymbol{\beta}-\boldsymbol{\beta}_\tau)\right\|_2$$
$$=\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\left|\frac{1}{m}\sum_{i\in\mathcal{I}_l}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{A}_i^{-1/2}\left[\widehat{\boldsymbol{R}}_i^{-1}-\boldsymbol{R}_i^{-1}\right]\boldsymbol{A}_i^{-1/2}\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\boldsymbol{z}_i\boldsymbol{\delta}\right|$$
$$\le\frac{1}{m}\sum_{i\in\mathcal{I}_l}\frac{1}{\tau(1-\tau)}\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\left|\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}_i^{\mathrm{T}}\left[\widehat{\boldsymbol{R}}_i^{-1}-\boldsymbol{R}_i^{-1}\right]\operatorname{diag}\left(\left\{\frac{1}{h_l}k\left(-\frac{y_{ik}-\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\beta}}{h_l}\right)\right\}_{k=1}^{n_i}\right)\boldsymbol{z}_i\boldsymbol{\delta}\right|$$
$$\le\frac{1}{\tau(1-\tau)}\frac{1}{m}\sum_{i\in\mathcal{I}_l}\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\left\{\|\boldsymbol{z}_i\boldsymbol{u}\|_2\cdot\left\|\widehat{\boldsymbol{R}}_i^{-1}-\boldsymbol{R}_i^{-1}\right\|_2\cdot\frac{\kappa_u}{h_l}\cdot\|\boldsymbol{z}_i\boldsymbol{\delta}\|_2\right\}$$
$$\le\frac{\kappa_u C_\varepsilon}{\tau(1-\tau)h_l}\sqrt{\frac{p}{M}}\cdot\frac{1}{m}\sum_{i\in\mathcal{I}_l}\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\left(\|\boldsymbol{z}_i\boldsymbol{u}\|_2\cdot\|\boldsymbol{z}_i\boldsymbol{\delta}\|_2\right)$$
$$\le\frac{\kappa_u C_\varepsilon}{\tau(1-\tau)h_l}\sqrt{\frac{p}{M}}\cdot\frac{1}{m}\sum_{i\in\mathcal{I}_l}\sup_{\boldsymbol{\delta}\in\mathbb{B}^p(r)}\left(\|\boldsymbol{z}_i\|_2^2\cdot\|\boldsymbol{\delta}\|_2\right)$$
$$\le\frac{\kappa_u C_\varepsilon r}{\tau(1-\tau)h_l}\sqrt{\frac{p}{M}}\cdot\frac{1}{m}\sum_{i\in\mathcal{I}_l}\|\boldsymbol{z}_i\|_2^2=\frac{\kappa_u C_\varepsilon r}{\tau(1-\tau)h}\sqrt{\frac{p}{M}}\cdot\frac{1}{m}\sum_{i\in\mathcal{I}_l}\left\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}_i\right\|_2^2$$
$$\le\frac{\kappa_u C_\varepsilon r}{\tau(1-\tau)h_l}\sqrt{\frac{p}{M}}\cdot\frac{1}{m}\sum_{i\in\mathcal{I}_l}\left\|\boldsymbol{\Sigma}^{-1/2}\right\|_2^2\cdot\|\boldsymbol{x}_i\|_2^2\le\frac{\kappa_u C_\varepsilon r}{\tau(1-\tau)\xi_p h_l}\sqrt{\frac{p}{M}}\cdot\frac{1}{m}\sum_{i\in\mathcal{I}_l}\|\boldsymbol{x}_i\|_2^2$$
$$\le\frac{\kappa_u C_\varepsilon r}{\tau(1-\tau)\xi_p h_l}\sqrt{\frac{p}{M}}\cdot n_u p\,B^2$$
$$=\frac{\kappa_u n_u B^2 C_\varepsilon r}{\tau(1-\tau)\xi_p h_l}\sqrt{\frac{p^3}{M}}=C_6 C_\varepsilon\frac{r}{h_l}\sqrt{\frac{p^3}{M}}.$$

holds with probability at least $1-\varepsilon$. ■

Lemma 9. *For any $\gamma > 0$, $0 < r < h_l/(4\iota_{0.25})$ and $s \geq 0$, there exists $f_{l,l} \geq 0$ such that*

$$\inf_{\boldsymbol{\beta} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r) \cap \mathbb{C}_{\boldsymbol{\Sigma}}(d)} \frac{\bar{D}^{(s)}_{\widetilde{\mathcal{L}}_{M,l}}(\boldsymbol{\beta}, \boldsymbol{\beta}_\tau)}{\kappa_* C_m^{-1} \|\boldsymbol{\beta} - \boldsymbol{\beta}_\tau\|_{\boldsymbol{\Sigma}}^2}$$
$$\geq \frac{3n_l}{4} f_{l,l} - f_{u,l}^{1/2} n_u \left[5B\sqrt{\frac{2dh_l \log(2p)}{r^2 m}} - 3\sqrt{\frac{2\gamma m_4}{h_l m}}\right] - \frac{13\gamma h_l n_u}{48 r^2 m}$$

holds with probability at least $1 - e^{-\gamma}$.

Proof. Define the parameter set $\mathcal{B}_0(r_1, r_2) = \{\boldsymbol{\zeta} \in \mathbb{R}^p : \|\boldsymbol{\zeta}\|_1 \leq r_1, \ \|\boldsymbol{\zeta}\|_{\boldsymbol{\Sigma}} \leq r_2\}$ for $r_1, r_2 > 0$. Let $\boldsymbol{\zeta} = \boldsymbol{\beta} - \boldsymbol{\beta}_\tau \in \mathcal{B}_0(4d^{1/2}r, \ r)$ for $\boldsymbol{\beta} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r) \cap \mathbb{C}_{\boldsymbol{\Sigma}}(d)$. From the proof of Lemma 1, it's sufficient to derive the upper bound of

$$E\left\{\sup_{\boldsymbol{\zeta} \in \mathcal{B}_0(4d^{1/2}r, \ r)} \left[\frac{1}{m}\sum_{i \in \mathcal{I}_l} Z_i \left(\sum_{k=1}^{n_i} \langle \varsigma_{ik} \boldsymbol{x}_{ik}, \boldsymbol{\zeta}/\|\boldsymbol{\zeta}\|_{\boldsymbol{\Sigma}} \rangle\right)\right]\right\}$$
$$\leq 4d^{1/2} E\left\|\frac{1}{m}\sum_{i \in \mathcal{I}_l} Z_i \left(\sum_{k=1}^{n_i} \varsigma_{ik} \boldsymbol{x}_{ik}\right)\right\|_\infty = 4d^{1/2} E\left\{E_Z\left\|\frac{1}{m}\sum_{i \in \mathcal{I}_l} Z_i \left(\sum_{k=1}^{n_i} \varsigma_{ik} \boldsymbol{x}_{ik}\right)\right\|_\infty\right\},$$

where E_Z stands for the (conditional) expectation over $Z_1, \ldots, Z_M$ given all the remaining random variables. Based on the Hoeffding's moment inequality, we have

$$E_Z\left\|\frac{1}{m}\sum_{i \in \mathcal{I}_l} Z_i \left(\sum_{k=1}^{n_i} \varsigma_{ik} \boldsymbol{x}_{ik}\right)\right\|_\infty \leq \max_{1 \leq j \leq p}\left\{\frac{1}{m}\sum_{i \in \mathcal{I}_l}\left(\sum_{k=1}^{n_i} \varsigma_{ik} x_{ik,j}\right)^2\right\}^{1/2} \sqrt{\frac{2\log(2p)}{m}}$$
$$\leq \max_{1 \leq j \leq p}\left(\frac{n_u}{m}\sum_{i \in \mathcal{I}_l}\sum_{k=1}^{n_i} \varsigma_{ik}^2 x_{ik,j}^2\right)^{1/2} \sqrt{\frac{2\log(2p)}{m}}$$
$$\leq \left(\frac{n_u}{m}\sum_{i \in \mathcal{I}_l}\sum_{k=1}^{n_i} \varsigma_{ik}\right)^{1/2} B\sqrt{\frac{2\log(2p)}{m}}.$$

Since $E(\varsigma_{ik}|\boldsymbol{x}_{ik}) \leq f_{u,l} h_l$, we have

$$E\left\{\sup_{\boldsymbol{\zeta} \in \mathcal{B}_0(4d^{1/2}r, \ r)} \left[\frac{1}{m}\sum_{i \in \mathcal{I}_l} Z_i \left(\sum_{k=1}^{n_i} \langle \varsigma_{ik} \boldsymbol{x}_{ik}, \boldsymbol{\zeta}/\|\boldsymbol{\zeta}\|_{\boldsymbol{\Sigma}} \rangle\right)\right]\right\} \leq 4n_u f_{u,l}^{1/2} B\sqrt{\frac{2dh_l \log(2p)}{m}}.$$

After using the similar arguments in the rest proof as that in the proof of Lemma 1, we obtained the claimed bound. ■

Lemma 10. *Assume that Conditions (A1)–(A7) hold. For any $\gamma > 0$ and $r > 0$,*

$$\Xi_M(r) = \sup_{\boldsymbol{\beta} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r) \cap \mathbb{C}_{\boldsymbol{\Sigma}}(d)} \left\|\wp_M(\boldsymbol{\beta}) - E\left[\wp_M(\boldsymbol{\beta})\right]\right\|_\infty$$
$$\leq \left[\frac{c_3}{h}\sqrt{\frac{2d \ \log \ (2p)}{M}} + c_4\sqrt{\frac{\log \ (2p) + \gamma}{Mh}} + c_5 d^{1/2}\frac{\log \ (2p) + \gamma}{Mh}\right] \cdot r$$

holds with probability at least $1 - e^{-\gamma}$, where $c_3 = 20\, C_r^{-1} n_u \kappa_u B^2$, $c_4 = C_r^{-1}(2\kappa_u f_u \sigma_u)^{1/2} \mu_4^{1/4} n_u$ with $\sigma_u = \max_{1 \leq j \leq p} \sigma_{jj}$, and $c_5 = (52/3) C_r^{-1} n_u \kappa_u B^2$.

Proof. Note that $\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\beta} - y_{ik} = \boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta} - \epsilon_{ik}$, where $\epsilon_{ik} = y_{ik} - \boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\beta}_\tau$. Write $\boldsymbol{x}_{i,j} = (x_{i1,j}, \ldots, x_{in_i,j})^{\mathrm{T}}$, we have

$$
\begin{aligned}
&\sup_{\boldsymbol{\beta}\in\mathcal{B}(r_1,r_2)} \Big\| \wp_M(\boldsymbol{\beta}) - E\big[\wp_M(\boldsymbol{\beta})\big] \Big\|_\infty \\
&= \max_{1\le j\le p} \sup_{\boldsymbol{\zeta}\in\mathcal{B}_0(r_1,r_2)} \left| \frac{1}{M}\sum_{i=1}^{M} (1-\mathbb{E}) \left\{ \boldsymbol{x}_{i,j}^{\mathrm{T}}\boldsymbol{\Omega}_i \left[\boldsymbol{K}\left(\frac{\boldsymbol{x}_i\boldsymbol{\zeta} - \boldsymbol{\epsilon}_i}{h}\right) - \boldsymbol{K}\left(\frac{-\boldsymbol{\epsilon}_i}{h}\right) \right] \right\} \right| \\
&\triangleq \max_{1\le j\le p} \sup_{\boldsymbol{\zeta}\in\mathcal{B}_0(r_1,r_2)} \left| \frac{1}{M}\sum_{i=1}^{M} (1-\mathbb{E})\, \psi_{i,j}(\boldsymbol{\zeta}) \right| \\
&\triangleq \max_{1\le j\le p} \varUpsilon_{M,j},
\end{aligned}
$$

where $\mathcal{B}(r_1,r_2) = \big\{\boldsymbol{\beta}\in\mathbb{R}^p : \|\boldsymbol{\beta}-\boldsymbol{\beta}_\tau\|_1 \le r_1, \|\boldsymbol{\beta}-\boldsymbol{\beta}_\tau\|_{\boldsymbol{\Sigma}} \le r_2\big\}$, $(1-\mathbb{E})\psi(\cdot) = \psi(\cdot) - E[\psi(\cdot)]$, $\boldsymbol{K}\big((\boldsymbol{x}_i\boldsymbol{\zeta} - \boldsymbol{\epsilon}_i)/h\big) = \big(K\left((\boldsymbol{x}_{i1}^{\mathrm{T}}\boldsymbol{\zeta} - \boldsymbol{\epsilon}_{i1})/h\right), \ldots, K\left((\boldsymbol{x}_{in_i}^{\mathrm{T}}\boldsymbol{\zeta} - \boldsymbol{\epsilon}_{in_i})/h\right)\big)^{\mathrm{T}}$, and $\boldsymbol{K}(-\boldsymbol{\epsilon}_i/h) = \left(K\left(-\epsilon_{i1}/h\right), \ldots, K\left(-\epsilon_{in_i}/h\right)\right)^{\mathrm{T}}$. At follows, we derive the upper bound of $\varUpsilon_{M,j}$.

Since $k(\cdot) = K'(\cdot)$ is uniformly bounded, we have

$$
\left| K\left(\frac{\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta} - \epsilon_{ik}}{h}\right) - K\left(\frac{-\epsilon_{ik}}{h}\right) \right| \le \frac{\kappa_u}{h}\left|\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta}\right|.
$$

Therefore, for $\boldsymbol{\zeta}\in\mathcal{B}_0(r_1,r_2)$, we have

$$
\begin{aligned}
\sup_{\boldsymbol{\zeta}\in\mathcal{B}_0(r_1,r_2)} \left|\psi_{i,j}(\boldsymbol{\zeta})\right| &\le \sup_{\boldsymbol{\zeta}\in\mathcal{B}_0(r_1,r_2)} \left\{ \left\|\boldsymbol{x}_{i,j}\right\|_2 \cdot \left\|\boldsymbol{\Omega}_i\right\|_2 \cdot \left\| \boldsymbol{K}\left(\frac{\boldsymbol{x}_i\boldsymbol{\zeta} - \boldsymbol{\epsilon}_i}{h}\right) - \boldsymbol{K}\left(\frac{-\boldsymbol{\epsilon}_i}{h}\right) \right\|_2 \right\} \\
&\le \sup_{\boldsymbol{\zeta}\in\mathcal{B}_0(r_1,r_2)} \left\{ \left\|\boldsymbol{x}_{i,j}\right\|_2 \cdot \left\|\boldsymbol{\Omega}_i\right\|_2 \cdot \sqrt{n_i}\frac{\kappa_u}{h} \cdot \|\boldsymbol{x}_{ik}\|_\infty \|\boldsymbol{\zeta}\|_1 \right\} \\
&\le \sqrt{n_i} \max_{1\le k\le n_i} |x_{ik,j}| \cdot \frac{1}{C_r} \cdot \sqrt{n_i}\frac{\kappa_u}{h} \cdot \max_{1\le j\le p} |x_{ik,j}| r_1 \\
&\le \frac{n_u\kappa_u B^2 r_1}{C_r h}
\end{aligned}
$$

and

$$
\begin{aligned}
\varsigma_{ik}(\boldsymbol{\zeta}) &= E_{\epsilon_{ik}|\boldsymbol{x}_{ik}} \left[K\left(\frac{\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta} - \epsilon_{ik}}{h}\right) - K\left(\frac{-\epsilon_{ik}}{h}\right) \right]^2 \\
&= \int_{-\infty}^{\infty} \left[K\left(\frac{\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta} - u}{h}\right) - K\left(\frac{-u}{h}\right) \right]^2 f_{\epsilon_{ik}|\boldsymbol{x}_{ik}}(u)du \\
&= h\int_{-\infty}^{\infty} \left[K\left(\frac{\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta}}{h} + v\right) - K\left(v\right) \right]^2 f_{\epsilon_{ik}|\boldsymbol{x}_{ik}}(-vh)dv \\
&= \frac{1}{h}(\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta})^2 \int_{-\infty}^{\infty} \left[\int_0^1 k\left(v + \frac{w\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta}}{h}\right) dw \right]^2 f_{\epsilon_{ik}|\boldsymbol{x}_{ik}}(-vh)dv \\
&\le \frac{f_u}{h}(\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta})^2 \int_{-\infty}^{\infty} \left[\int_0^1 k\left(v + \frac{w\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta}}{h}\right) dw \right]^2 dv \\
&\le \frac{f_u}{h}(\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta})^2 \left\{ \int_0^1 \left[\int_{-\infty}^{\infty} k^2\left(v + \frac{w\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta}}{h}\right) dv \right]^{1/2} dw \right\}^2 \\
&\le \frac{f_u}{h}(\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta})^2 \kappa_u.
\end{aligned}
$$

Denote $\boldsymbol{K}_i(\boldsymbol{\beta}) = \boldsymbol{K}\big((\boldsymbol{x}_i\boldsymbol{\zeta} - \boldsymbol{\epsilon}_i)/h\big)$ and $\boldsymbol{K}_i(\boldsymbol{\beta}_\tau) = \boldsymbol{K}\left(-\boldsymbol{\epsilon}_i/h\right)$ for the sake of notation, we have

$$\begin{aligned}
E\big[\psi_{i,j}^2(\boldsymbol{\zeta})\big] &= E\Big[\boldsymbol{x}_{i,j}^{\mathrm{T}}\boldsymbol{\Omega}_i(\boldsymbol{\alpha})\big(\boldsymbol{K}_i(\boldsymbol{\beta}) - \boldsymbol{K}_i(\boldsymbol{\beta}_\tau)\big)\big(\boldsymbol{K}_i(\boldsymbol{\beta}) - \boldsymbol{K}_i(\boldsymbol{\beta}_\tau)\big)^{\mathrm{T}}\boldsymbol{\Omega}_i(\boldsymbol{\alpha})\boldsymbol{x}_{i,j}\Big] \\
&= E_{\boldsymbol{x}_i}\Big\{\boldsymbol{x}_{i,j}^{\mathrm{T}}\boldsymbol{\Omega}_i(\boldsymbol{\alpha})E_{\boldsymbol{\epsilon}_i|\boldsymbol{x}_i}\big[\big(\boldsymbol{K}_i(\boldsymbol{\beta}) - \boldsymbol{K}_i(\boldsymbol{\beta}_\tau)\big)\big(\boldsymbol{K}_i(\boldsymbol{\beta}) - \boldsymbol{K}_i(\boldsymbol{\beta}_\tau)\big)^{\mathrm{T}}\big]\boldsymbol{\Omega}_i(\boldsymbol{\alpha})\boldsymbol{x}_{i,j}\Big\} \\
&\le E_{\boldsymbol{x}_i}\Big\{\boldsymbol{x}_{i,j}^{\mathrm{T}}\boldsymbol{\Omega}_i(\boldsymbol{\alpha})E_{\boldsymbol{\epsilon}_i|\boldsymbol{x}_i}\big[\big(\boldsymbol{K}_i(\boldsymbol{\beta}) - \boldsymbol{K}_i(\boldsymbol{\beta}_\tau)\big)^{\mathrm{T}}\big(\boldsymbol{K}_i(\boldsymbol{\beta}) - \boldsymbol{K}_i(\boldsymbol{\beta}_\tau)\big)\big]\boldsymbol{I}_{n_i}\boldsymbol{\Omega}_i(\boldsymbol{\alpha})\boldsymbol{x}_{i,j}\Big\} \\
&\le \frac{1}{C_r^2}E_{\boldsymbol{x}_i}\left[\boldsymbol{x}_{i,j}^{\mathrm{T}}\boldsymbol{x}_{i,j}\sum_{k=1}^{n_i}\varsigma_{ik}(\boldsymbol{\zeta})\right] \le \frac{f_u\kappa_u}{hC_r^2}E_{\boldsymbol{x}_i}\left[\boldsymbol{x}_{i,j}^{\mathrm{T}}\boldsymbol{x}_{i,j}\sum_{k=1}^{n_i}(\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta})^2\right] \\
&= \frac{f_u\kappa_u}{hC_r^2}E_{\boldsymbol{x}_i}\big[\boldsymbol{x}_{i,j}^{\mathrm{T}}\boldsymbol{x}_{i,j}\cdot\boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_i\boldsymbol{\zeta}\big] \le \frac{f_u\kappa_u}{hC_r^2}\left[E(\boldsymbol{x}_{i,j}^{\mathrm{T}}\boldsymbol{x}_{i,j})^2\right]^{1/2}\left[E(\boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_i\boldsymbol{\zeta})^2\right]^{1/2} \\
&\le \frac{f_u\kappa_u n_u^2\mu_4^{1/2}\sigma_{jj}r_2^2}{hC_r^2},
\end{aligned} \tag{32}$$

where $E(\boldsymbol{x}_{i,j}^{\mathrm{T}}\boldsymbol{x}_{i,j})^2 = E\big\|\boldsymbol{z}_i\boldsymbol{\Sigma}^{1/2}\boldsymbol{e}_j\big\|_2^4 \le n_u^2\mu_4\big\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{e}_j\big\|_2^4 = n_u^2\mu_4\sigma_{jj}^2$ with the j-th diagonal element σ_{jj} of $\boldsymbol{\Sigma}$, and $E(\boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_i\boldsymbol{\zeta})^2 \le n_u^2\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{\zeta}\|_2^4 \le n_u^2r_2^4$.

According to the Bousquet's version of Talagrand's inequality in Bousquet (2003) [2], for any $\gamma > 0$, we have

$$\varUpsilon_{M,j} \le \frac{5}{4}E(\varUpsilon_{M,j}) + \frac{f_u^{1/2}\mu_4^{1/4}\sigma_{jj}^{1/2}\kappa_u^{1/2}n_ur_2}{C_r}\sqrt{\frac{2\gamma}{Mh}} + \left(4 + \frac{1}{3}\right)\frac{n_u\kappa_uB^2r_1\gamma}{C_rMh} \tag{33}$$

with probability at least $1 - e^{-\gamma}$.

At follows, we derive the upper bound of $E(\varUpsilon_{M,j})$. Let $Z_1, \ldots, Z_M$ be i.i.d. Rademacher random variables. Based on the Rademacher symmetrization,

$$\begin{aligned}
E(\varUpsilon_{M,j}) &\le 2E\left\{\sup_{\boldsymbol{\zeta}\in\mathcal{B}_0(r_1,r_2)}\left|\frac{1}{M}\sum_{i=1}^M Z_i\psi_{i,j}(\boldsymbol{\zeta})\right|\right\} \\
&= 2E\left\{E_Z\left[\sup_{\boldsymbol{\zeta}\in\mathcal{B}_0(r_1,r_2)}\left|\frac{1}{M}\sum_{i=1}^M Z_i\psi_{i,j}(\boldsymbol{\zeta})\right|\right]\right\},
\end{aligned} \tag{34}$$

where the expectation E_Z is taken with respect to $Z_1, \ldots, Z_M$. Write $\psi_{ij}(\boldsymbol{\zeta}) = \boldsymbol{x}_{i,j}^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\varphi}_i(\boldsymbol{x}_i\boldsymbol{\zeta})$ with $\boldsymbol{\varphi}_i(\boldsymbol{x}_i\boldsymbol{\zeta}) = \big(\varphi_{i1}(\boldsymbol{x}_{i1}^{\mathrm{T}}\boldsymbol{\zeta}), \ldots, \varphi_{in_i}(\boldsymbol{x}_{in_i}^{\mathrm{T}}\boldsymbol{\zeta})\big)^{\mathrm{T}}$,where $\varphi_{ik}(\cdot)$ satisfies $\varphi_{ik}(0) = 0$ and $|\varphi_{ik}(u) - \varphi_{ik}(v)| \le \kappa_uh^{-1}|u - v|$. According to the Talagrand's contraction principle in Ledoux and Talagrand (1991)[5], we have

$$\begin{aligned}
&E_Z\left\{\sup_{\boldsymbol{\zeta}\in\mathcal{B}_0(r_1,r_2)}\left|\frac{1}{M}\sum_{i=1}^M Z_i\psi_{i,j}(\boldsymbol{\zeta})\right|\right\} \\
&= E_Z\left\{\sup_{\boldsymbol{\zeta}\in\mathcal{B}_0(r_1,r_2)}\left|\frac{1}{M}\sum_{i=1}^M Z_i\Big[\boldsymbol{x}_{i,j}^{\mathrm{T}}\boldsymbol{\Omega}_i\boldsymbol{\varphi}_i(\boldsymbol{x}_i\boldsymbol{\zeta})\Big]\right|\right\} \\
&\le \frac{2\kappa_u}{C_rh}\max_{1\le i\le M, 1\le k\le n_i}|x_{ik,j}|\cdot E_Z\left\{\sup_{\boldsymbol{\zeta}\in\mathcal{B}_0(r_1,r_2)}\left|\frac{1}{M}\sum_{i=1}^M Z_i\left(\sum_{k=1}^{n_i}\boldsymbol{x}_{ik}^{\mathrm{T}}\boldsymbol{\zeta}\right)\right|\right\} \\
&\le \frac{2\kappa_uBr_1}{C_rh}E_Z\left\|\frac{1}{M}\sum_{i=1}^M Z_i\left(\sum_{k=1}^{n_i}\boldsymbol{x}_{ik}\right)\right\|_\infty.
\end{aligned} \tag{35}$$

Based on the Hoeffding's moment inequality,

$$E_Z \left\| \frac{1}{M} \sum_{i=1}^{M} Z_i \left(\sum_{k=1}^{n_i} \boldsymbol{x}_{ik} \right) \right\|_\infty \leq \max_{1\leq j\leq p} \left(\frac{1}{M} \sum_{i=1}^{M} \left(\sum_{k=1}^{n_i} x_{ik,j} \right)^2 \right)^{1/2} \sqrt{\frac{2\log(2p)}{M}}$$
$$\leq \max_{1\leq j\leq p} \left(\frac{n_u^2}{M} \sum_{i=1}^{M} \left(\frac{1}{n_i} \sum_{k=1}^{n_i} x_{ik,j}^2 \right) \right)^{1/2} \sqrt{\frac{2\log(2p)}{M}}. \tag{36}$$

Together with (34)–(36), for any $j = 1, \ldots, p$,

$$E(\Upsilon_{M,j}) \leq \frac{4 n_u \kappa_u B^2 r_1}{C_r h} \sqrt{\frac{2\log(2p)}{M}}. \tag{37}$$

Take $r_1 = 4d^{1/2} r$, $r_2 = r$, and $\gamma' = \log(2p) + \gamma$. Combine with (32), (33), and (37), we have the claimed bound. ■

3 Proofs of Main Results

3.1 Proof of Theorem 1

Theorem 1 (Restatement of Theorem 1 from main text)**.** *Assume that Conditions (A1)–(A7) hold. Let $r_0 \gtrsim r_* > 0$. For any $\gamma, \varepsilon > 0$, suppose the bandwidths $h_l \geq h > 0$ satisfy $\max(r_0, \sqrt{(p+\gamma)/m}) \lesssim h_l \lesssim 1$ and $\sqrt{(p+\gamma)/M} \lesssim h$. Then, conditioned on the event $\mathcal{E}_0(r_0) \cap \mathcal{E}_*(\bar{r}_*) \cap \mathcal{E}_*(\tilde{r}_*)$, there exist constants $C_\varepsilon, M_0 > 0$ such that for all $M \geq M_0$, the one-step distributed estimator $\widehat{\boldsymbol{\beta}}_l^{(1)}$ computed on the l-th machine satisfies*

$$\|\widehat{\boldsymbol{\beta}}_l^{(1)} - \boldsymbol{\beta}_\tau\|_{\boldsymbol{\Sigma}} \lesssim C \left(\sqrt{\frac{p+\gamma}{Mh}} + \sqrt{\frac{p+\gamma}{m}} + \frac{C_\varepsilon}{h_l} \sqrt{\frac{p^3}{M}} + h_l + 1 - L^{-1} \right) \cdot r_0 + r_*$$

with probability at least $1 - 3e^{-\gamma} - \varepsilon$, where $C > 0$ is a constant depending on $(\tau, n_u, f_u, \kappa_u, \mu_3, a_1, L_0, B, \xi_p, \xi_{\min})$.

Proof. Assume the event $\mathcal{E}_0(r_0) = \{\widehat{\boldsymbol{\beta}}^{(0)} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_0)\}$ holds. Set $r_{\text{loc}} = h_l/(4\iota_{0.25})$. Let $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_l^{(1)}$ be the one-step estimator that minimizes $\widetilde{\mathcal{L}}_{M,l}^{(0)}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ on the lth machine. Define an intermediate estimator $\widehat{\boldsymbol{\beta}}_\upsilon = \boldsymbol{\beta}_\tau + \upsilon(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_\tau)$, where $\upsilon = \sup\{u \in [0,1] : \boldsymbol{\beta}_\tau + u\,(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_\tau) \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_{\text{loc}})\}$ is the largest value of $u \in (0,1]$ such that the convex combination of $\boldsymbol{\beta}_\tau$ and $\widehat{\boldsymbol{\beta}}$: $(1-\upsilon)\boldsymbol{\beta}_\tau + \upsilon\widehat{\boldsymbol{\beta}}$ falls into $\mathbb{B}_{\boldsymbol{\Sigma}}(r_0)$. If $\widehat{\boldsymbol{\beta}} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_{\text{loc}})$, $\upsilon = 1$, and $\widehat{\boldsymbol{\beta}}_\upsilon = \widehat{\boldsymbol{\beta}}$; otherwise, if $\widehat{\boldsymbol{\beta}} \notin \mathbb{B}_{\boldsymbol{\Sigma}}(r_{\text{loc}})$, $\upsilon \in (0,1)$, and $\widehat{\boldsymbol{\beta}}_\upsilon$ falls onto the boundary of $\mathbb{B}_{\boldsymbol{\Sigma}}(r_{\text{loc}})$, that is, $\widehat{\boldsymbol{\beta}}_\upsilon \in \partial\mathbb{B}_{\boldsymbol{\Sigma}}(r_{\text{loc}}) = \{\boldsymbol{\beta} \in \mathbb{R}^p : \|\boldsymbol{\beta} - \boldsymbol{\beta}_\tau\|_{\boldsymbol{\Sigma}} = r_{\text{loc}}\}$.

Based on Lemma F.2 in Fan et al. (2018) [3] and the first-order optimality condition $\nabla\widetilde{\mathcal{L}}_{M,l}^{(0)}(\widehat{\boldsymbol{\beta}}) = \mathbf{0}$, we have

$$\bar{D}_{\widetilde{\mathcal{L}}_{M,l}}^{(0)}(\widehat{\boldsymbol{\beta}}_\upsilon, \boldsymbol{\beta}_\tau) \leq \upsilon \bar{D}_{\widetilde{\mathcal{L}}_{M,l}}^{(0)}(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}_\tau) = -\upsilon \left\langle \nabla\widetilde{\mathcal{L}}_{M,l}^{(0)}(\boldsymbol{\beta}_\tau), \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_\tau \right\rangle$$
$$\leq \left\| \boldsymbol{\Sigma}^{-1/2} \nabla\widetilde{\mathcal{L}}_{M,l}^{(0)}(\boldsymbol{\beta}_\tau) \right\|_2 \cdot \left\| \boldsymbol{\Sigma}^{1/2}(\widehat{\boldsymbol{\beta}}_\upsilon - \boldsymbol{\beta}_\tau) \right\|_2. \tag{38}$$

Based on Lemma 2, given any $\gamma > 0$, as long as $(p+\gamma)/m \lesssim h_l \lesssim 1$,

$$\bar{D}^{(0)}_{\widetilde{\mathcal{L}}_{M,l}}(\boldsymbol{\beta}, \boldsymbol{\beta}_\tau) \geq 0.5\ C_m^{-1} \kappa_* n_l f_{l,l} \cdot \|\boldsymbol{\beta} - \boldsymbol{\beta}_\tau\|^2_{\boldsymbol{\Sigma}} \tag{39}$$

holds uniformly over all $\boldsymbol{\beta} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_{\text{loc}})$ with probability at least $1 - e^{-\gamma}$.

Note that $\nabla \widetilde{\mathcal{L}}^{(0)}_{M,l}(\boldsymbol{\beta}) = \nabla \mathcal{L}^{(0)}_{m,l}(\boldsymbol{\beta}) - \nabla \mathcal{L}^{(0)}_{m,l}(\widehat{\boldsymbol{\beta}}^{(0)}) + \nabla \mathcal{L}^{(0)}_{M}(\widehat{\boldsymbol{\beta}}^{(0)})$. Since

$$\nabla \mathcal{L}^{(0)}_{m,l}(\boldsymbol{\beta}_\tau) = \widehat{\boldsymbol{U}}_{m,l}(\widehat{\boldsymbol{\beta}}^{(0)}) - \widehat{\boldsymbol{D}}_{m,l}(\widehat{\boldsymbol{\beta}}^{(0)})(\widehat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau).$$

We can decompose $\nabla \widetilde{\mathcal{L}}^{(0)}_{M,l}(\boldsymbol{\beta}_\tau)$ as

$$\begin{aligned}
\nabla \widetilde{\mathcal{L}}^{(0)}_{M,l}(\boldsymbol{\beta}_\tau) &= \widehat{\boldsymbol{U}}_M(\widehat{\boldsymbol{\beta}}^{(0)}) - \widehat{\boldsymbol{D}}_{m,l}(\widehat{\boldsymbol{\beta}}^{(0)})(\widehat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau) \\
&= \widehat{\boldsymbol{U}}_M(\widehat{\boldsymbol{\beta}}^{(0)}) - \boldsymbol{U}_M(\widehat{\boldsymbol{\beta}}^{(0)}) + \boldsymbol{U}_M(\widehat{\boldsymbol{\beta}}^{(0)}) - \boldsymbol{U}_M(\boldsymbol{\beta}_\tau) + \boldsymbol{U}_M(\boldsymbol{\beta}_\tau) \\
&\quad - \widehat{\boldsymbol{D}}_{m,l}(\widehat{\boldsymbol{\beta}}^{(0)})(\widehat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau) \\
&= \widehat{\boldsymbol{U}}_M(\widehat{\boldsymbol{\beta}}^{(0)}) - \boldsymbol{U}_M(\widehat{\boldsymbol{\beta}}^{(0)}) \\
&\quad + \boldsymbol{U}_M(\widehat{\boldsymbol{\beta}}^{(0)}) - \boldsymbol{U}_M(\boldsymbol{\beta}_\tau) - \boldsymbol{H}_M(\boldsymbol{\beta}_\tau)(\widehat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau) \\
&\quad + \boldsymbol{H}_M(\boldsymbol{\beta}_\tau) - \boldsymbol{H}_M(\widehat{\boldsymbol{\beta}}^{(0)})(\widehat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau) \\
&\quad + \boldsymbol{H}_M(\widehat{\boldsymbol{\beta}}^{(0)}) - \boldsymbol{H}_{m,l}(\widehat{\boldsymbol{\beta}}^{(0)})(\widehat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau) \\
&\quad + \boldsymbol{H}_{m,l}(\widehat{\boldsymbol{\beta}}^{(0)}) - \boldsymbol{D}_{m,l}(\widehat{\boldsymbol{\beta}}^{(0)})(\widehat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau) \\
&\quad + \boldsymbol{D}_{m,l}(\widehat{\boldsymbol{\beta}}^{(0)}) - \widehat{\boldsymbol{D}}_{m,l}(\widehat{\boldsymbol{\beta}}^{(0)})(\widehat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau) \\
&\quad + \boldsymbol{U}_M(\boldsymbol{\beta}_\tau).
\end{aligned}$$

According to Lemmas 2–8, as long as $h \gtrsim \sqrt{(p+\gamma)/M}$ and $m \gtrsim p + \gamma$,

$$\begin{aligned}
\left\|\boldsymbol{\Sigma}^{-1/2} \nabla \widetilde{\mathcal{L}}^{(0)}_{M,l}(\boldsymbol{\beta}_\tau)\right\|_2 &\leq \left\|\mathfrak{S}_M(\widehat{\boldsymbol{\beta}}^{(0)})\right\|_2 + \left\|\Delta_M(\widehat{\boldsymbol{\beta}}^{(0)})\right\|_2 + \left\|\mathfrak{T}_M(\boldsymbol{\beta}_\tau)\right\|_2 \\
&\quad + \left\|\mathfrak{F}_M(\widehat{\boldsymbol{\beta}}^{(0)})\right\|_2 + \left\|\mathfrak{F}_{M,l}(\widehat{\boldsymbol{\beta}}^{(0)})\right\|_2 \\
&\quad + \left\|\mathfrak{H}_{m,l}(\widehat{\boldsymbol{\beta}}^{(0)})\right\|_2 + \left\|\mathfrak{E}_{m,l}(\widehat{\boldsymbol{\beta}}^{(0)})\right\|_2 \\
&\lesssim \xi_p^{-1/2} C_0 C_\varepsilon \sqrt{\frac{p^2}{M}} + C_1\ r \left(\sqrt{\frac{p+\gamma}{Mh}} + r\right) \\
&\quad + C_2 \left(\sqrt{\frac{p+\gamma}{M}} + h^2\right) + C_3\ r^2 + \left[C_4(1 - L^{-1}) + C_5|h_l - h|\right] r \\
&\quad + C_6\ r\sqrt{\frac{p+\gamma}{m}} + C_7 C_\varepsilon \frac{r}{h_l}\sqrt{\frac{p^3}{M}}
\end{aligned}$$

holds with probability at least $1 - 3e^{-\gamma} - 2\varepsilon$. Then, conditioned on the event $\mathcal{E}_0(r_0) \cap \mathcal{E}_*(\bar{r}_*) \cap \mathcal{E}_*(\tilde{r}_*)$, as long as $h_l \geq h > 0$ and $h_l \gtrsim r_0$,

$$\left\|\boldsymbol{\Sigma}^{-1/2} \nabla \widetilde{\mathcal{L}}^{(0)}_{M,l}(\boldsymbol{\beta}_\tau)\right\|_2 \leq C \left(\sqrt{\frac{p+\gamma}{Mh}} + \sqrt{\frac{p+\gamma}{m}} + \frac{C_\varepsilon}{h_l}\sqrt{\frac{p^3}{M}} + h_l + 1 - L^{-1}\right) \cdot r_0 + r_*, \tag{40}$$

holds with probability at least $1-2e^{-\gamma}-\varepsilon$, where $r_* = O_p(p/\sqrt{M}+h^2)$.

Together with (38), (39), and (40), conditioned on the event $\mathcal{E}_0(r_0)\cap\mathcal{E}_*(\bar{r}_*)\cap\mathcal{E}_*(\tilde{r}_*)$,

$$\begin{aligned}\left\|\widehat{\boldsymbol{\beta}}_\upsilon-\boldsymbol{\beta}_\tau\right\|_{\boldsymbol{\Sigma}} &\le 2\,C_m(\kappa_* n_l f_{l,l})^{-1}\left\|\boldsymbol{\Sigma}^{-1/2}\nabla\widetilde{\mathcal{L}}^{(0)}_{M,l}(\boldsymbol{\beta}_\tau)\right\|_2\\ &\le 2\,C_m(\kappa_* n_l f_{l,l})^{-1}\left\{C\left(\sqrt{\frac{p+\gamma}{Mh}}+\sqrt{\frac{p+\gamma}{m}}+\frac{C_\varepsilon}{h_l}\sqrt{\frac{p^3}{M}}+h_l+1-L^{-1}\right)\cdot r_0+r_*\right\}\end{aligned} \tag{41}$$

holds with probability at least $1-3e^{-\gamma}-\varepsilon$. Let the bandwidth $h_l\ge h>0$ satisfy $1\gtrsim h_l\gtrsim\max(r_0,r_*)$ and $\sqrt{(p+\gamma)/(Mh)}+\sqrt{(p+\gamma)/m}+C_\varepsilon h_l^{-1}\sqrt{p^3/M}+h_l+1-L^{-1}\lesssim 1$, so that the right hand side of (41) is strictly less than r_{loc}. Then, the intermediate estimator $\widehat{\boldsymbol{\beta}}_\upsilon$ falls into the interior of the local region $\mathbb{B}_{\boldsymbol{\Sigma}}(r_{\mathrm{loc}})$ with high probability conditioned on the event $\mathcal{E}_0(r_0)\cap\mathcal{E}_*(\bar{r}_*)\cap\mathcal{E}_*(\tilde{r}_*)$. Note that if $\widehat{\boldsymbol{\beta}}\notin\mathbb{B}_{\boldsymbol{\Sigma}}(r_{\mathrm{loc}})$, $\widehat{\boldsymbol{\beta}}_\upsilon$ lies on the boundary of $\mathbb{B}_{\boldsymbol{\Sigma}}(r_{\mathrm{loc}})$, which is a contradiction to (41). Therefore, $\widehat{\boldsymbol{\beta}}=\widehat{\boldsymbol{\beta}}_\upsilon\in\mathbb{B}_{\boldsymbol{\Sigma}}(r_{\mathrm{loc}})$ and the bound (41) also applies to $\widehat{\boldsymbol{\beta}}$.

■

3.2 Proof of Theorem 2

Theorem 2 (Restatement of Theorem 2 from main text)**.** *Assume the conditions of Theorem 1 hold. For any $\gamma,\varepsilon>0$, conditioned on the event $\mathcal{E}_0(r_0)\cap\mathcal{E}_*(\bar{r}_*)\cap\mathcal{E}_*(\tilde{r}_*)$, there exist constants $C_\varepsilon, M_0>0$ such that for all $M\ge M_0$, the S-step distributed estimator $\widehat{\boldsymbol{\beta}}_l^{(S)}$ computed on the l-th machine satisfies*

$$\|\widehat{\boldsymbol{\beta}}_l^{(S)}-\boldsymbol{\beta}_\tau\|_{\boldsymbol{\Sigma}}\lesssim\vartheta^S\cdot r_0+r_*\lesssim r_*$$

with probability at least $1-(2S+1)e^{-\gamma}-S\varepsilon$, provided that $\vartheta^{S-1}\cdot r_0\lesssim r_*$. Here, $\vartheta=\vartheta(p,m,M,L,h_l,h,\gamma,C_\varepsilon)=C\{\sqrt{(p+\gamma)/(Mh)}+\sqrt{(p+\gamma)/m}+C_\varepsilon h_l^{-1}\sqrt{p^3/M}+h_l+1-L^{-1}\}$, and the number of iterations satisfies $S\gtrsim\log(r_0/r_*)/\log(1/h_l)$.

Proof. Given the s-step ($s\ge 0$) estimator $\widehat{\boldsymbol{\beta}}_l^{(s)}$ of $\boldsymbol{\beta}$ on the lth machine and denote $\widehat{\boldsymbol{\beta}}_l^{(s)}=\widehat{\boldsymbol{\beta}}^{(s)}$, we have

$$\begin{aligned}\boldsymbol{\Sigma}^{-1/2}\nabla\widetilde{\mathcal{L}}^{(s)}_{M,l}(\boldsymbol{\beta}_\tau)&=\mathfrak{S}_M\big(\widehat{\boldsymbol{\beta}}^{(s)}\big)+\Delta_M\big(\widehat{\boldsymbol{\beta}}^{(s)}\big)-\mathfrak{F}_M\big(\widehat{\boldsymbol{\beta}}^{(s)}\big)\\&\quad+\mathfrak{F}_{M,l}\big(\widehat{\boldsymbol{\beta}}^{(s)}\big)-\mathfrak{H}_{m,l}\big(\widehat{\boldsymbol{\beta}}^{(s)}\big)+\mathfrak{E}_{m,l}\big(\widehat{\boldsymbol{\beta}}^{(s)}\big)+\mathfrak{T}_M(\boldsymbol{\beta}_\tau).\end{aligned}$$

Therefore,

$$\begin{aligned}\left\|\boldsymbol{\Sigma}^{-1/2}\nabla\widetilde{\mathcal{L}}^{(s)}_{M,l}(\boldsymbol{\beta}_\tau)\right\|_2&\le\left\|\mathfrak{S}_M\big(\widehat{\boldsymbol{\beta}}^{(s)}\big)\right\|_2+\left\|\Delta_M\big(\widehat{\boldsymbol{\beta}}^{(s)}\big)\right\|_2+\left\|\mathfrak{F}_M\big(\widehat{\boldsymbol{\beta}}^{(s)}\big)\right\|_2+\left\|\mathfrak{F}_{M,l}\big(\widehat{\boldsymbol{\beta}}^{(s)}\big)\right\|_2\\&\quad+\left\|\mathfrak{H}_{m,l}\big(\widehat{\boldsymbol{\beta}}^{(s)}\big)\right\|_2+\left\|\mathfrak{E}_{m,l}\big(\widehat{\boldsymbol{\beta}}^{(s)}\big)\right\|_2+\left\|\mathfrak{T}_M(\boldsymbol{\beta}_\tau)\right\|_2.\end{aligned} \tag{42}$$

Define the "good" events for the sequence of iterates $\{\widehat{\boldsymbol{\beta}}^{(s)}\}_{s=0}^S$:

$$\mathcal{E}_s(r_s)=\left\{\widehat{\boldsymbol{\beta}}^{(s)}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r_s)\right\},\ \ s=0,1,\ldots,S,$$

where $r_0 \geq r_1 \geq \cdots \geq r_S > 0$ are some sequence of radius to be determined. Let $r_{\text{loc}} = h_l/(4\gamma_{0.25})$ be the same local radius defined in the proof of Theorem 1. According to Lemma 1, as long as $(p+\gamma)/m \lesssim h_l \lesssim 1$,

$$\bar{D}^{(s)}_{\widetilde{\mathcal{L}}_{M,l}}(\boldsymbol{\beta}, \boldsymbol{\beta}_\tau) \geq 0.5\; C_m^{-1}\kappa_* n_l f_{l,l} \cdot \|\boldsymbol{\beta} - \boldsymbol{\beta}_\tau\|_{\boldsymbol{\Sigma}}^2 \tag{43}$$

holds uniformly over all $\boldsymbol{\beta} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_{\text{loc}})$ with probability at least $1 - e^{-\gamma}$.

Let $\widehat{\boldsymbol{\beta}}^{(s+1)} = \widehat{\boldsymbol{\beta}}_l^{(s+1)}$ be the update of $\boldsymbol{\beta}$ on the lth machine in the $(s+1)$th iteration. Define an intermediate estimator $\widehat{\boldsymbol{\beta}}_\upsilon^{(s+1)} = \boldsymbol{\beta}_\tau + \upsilon(\widehat{\boldsymbol{\beta}}^{(s+1)} - \boldsymbol{\beta}_\tau)$ (a convex combination of $\widehat{\boldsymbol{\beta}}^{(s+1)}$ and $\boldsymbol{\beta}_\tau$). Assume that $\widehat{\boldsymbol{\beta}}_\upsilon^{(s)} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_{\text{loc}})$. Based on (38), (42), and (43), conditioned on the event $\mathcal{E}_0(r_0) \cap \mathcal{E}_*(\bar{r}_*) \cap \mathcal{E}_*(\tilde{r}_*)$,

$$\begin{aligned}
\left\|\widehat{\boldsymbol{\beta}}_\upsilon^{(s+1)} - \boldsymbol{\beta}_\tau\right\|_{\boldsymbol{\Sigma}} &\leq \bar{\kappa}^{-1}\left\|\boldsymbol{\Sigma}^{-1/2}\nabla\widetilde{\mathcal{L}}^{(s)}_{M,l}(\boldsymbol{\beta}_\tau)\right\|_2 \\
&\leq \bar{\kappa}^{-1}\Big(\left\|\mathfrak{S}_M(\widehat{\boldsymbol{\beta}}^{(s)})\right\|_2 + \left\|\Delta_M(\widehat{\boldsymbol{\beta}}^{(s)})\right\|_2 + \left\|\mathfrak{F}_M(\widehat{\boldsymbol{\beta}}^{(s)})\right\|_2 \\
&\quad + \left\|\mathfrak{F}_{M,l}(\widehat{\boldsymbol{\beta}}^{(s)})\right\|_2 + \left\|\mathfrak{H}_{m,l}(\widehat{\boldsymbol{\beta}}^{(s)})\right\|_2 + \left\|\mathfrak{E}_{m,l}(\widehat{\boldsymbol{\beta}}^{(s)})\right\|_2 + \|\mathfrak{T}_M(\boldsymbol{\beta}_\tau)\|_2\Big) \\
&\leq \bar{\kappa}^{-1}\Big(\left\|\Delta_M(\widehat{\boldsymbol{\beta}}^{(s)})\right\|_2 + \left\|\mathfrak{F}_M(\widehat{\boldsymbol{\beta}}^{(s)})\right\|_2 + \left\|\mathfrak{F}_{M,l}(\widehat{\boldsymbol{\beta}}^{(s)})\right\|_2 \\
&\quad + \left\|\mathfrak{H}_{m,l}(\widehat{\boldsymbol{\beta}}^{(s)})\right\|_2 + \left\|\mathfrak{E}_{m,l}(\widehat{\boldsymbol{\beta}}^{(s)})\right\|_2 + r_*\Big),
\end{aligned} \tag{44}$$

where $\bar{\kappa} = 0.5\; C_m^{-1}\kappa_* n_l f_{l,l}$. Define the event

$$\begin{aligned}
\mathcal{F}(r) = \Bigg\{ \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \Big\{ &\left\|\Delta_M(\boldsymbol{\beta})\right\|_2 + \left\|\mathfrak{F}_M(\boldsymbol{\beta})\right\|_2 + \left\|\mathfrak{F}_{M,l}(\boldsymbol{\beta})\right\|_2 \\
&+ \left\|\mathfrak{H}_{m,l}(\boldsymbol{\beta})\right\|_2 + \left\|\mathfrak{E}_{m,l}(\boldsymbol{\beta})\right\|_2 \Big\} \leq \vartheta(\gamma) \cdot r \Bigg\},
\end{aligned}$$

where $\vartheta(\gamma) = C\{\sqrt{(p+\gamma)/(Mh)} + \sqrt{(p+\gamma)/m} + C_\varepsilon h_l^{-1}\sqrt{p^3/M} + h_l + 1 - L^{-1}\}$ (for some $C > 0$), such that $\mathcal{F}(r)$ holds with probability at least $1 - 2e^{-\gamma} - \varepsilon$ for $0 < r \lesssim h_l$.

Let $\mathcal{E}_{\text{loc}}$ be the event where the local strong convexity in (43) holds. From (44), at the first iteration, conditioned on the event $\mathcal{E}_0(r_0) \cap \mathcal{E}_*(\bar{r}_*) \cap \mathcal{E}_*(\tilde{r}_*) \cap \mathcal{E}_{\text{loc}} \cap \mathcal{F}(r_0)$,

$$\left\|\widehat{\boldsymbol{\beta}}_\upsilon^{(1)} - \boldsymbol{\beta}_\tau\right\|_{\boldsymbol{\Sigma}} \leq r_1 := \bar{\kappa}^{-1}\vartheta(\gamma) \cdot r_0 + \bar{\kappa}^{-1} r_*. \tag{45}$$

From the constraints on (h_l, h, r_0, r_*), we have $\bar{\kappa}^{-1}\vartheta(\gamma) < 1$, $r_1 < r_{\text{loc}} \asymp h_l$, and $r_1 \leq r_0$. This implies $\widehat{\boldsymbol{\beta}}^{(1)} = \widehat{\boldsymbol{\beta}}_\upsilon^{(1)} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_{\text{loc}})$, which in turn certifies the event $\mathcal{E}_1(r_1) = \left\{\widehat{\boldsymbol{\beta}}^{(1)} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_1)\right\}$.

Next, we assume that for some $s \geq 1$, $\widehat{\boldsymbol{\beta}}^{(s)} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_s)$, where $r_s := \bar{\kappa}^{-1}\vartheta(\gamma) \cdot r_{s-1} + \bar{\kappa}^{-1} r_* \leq r_{s-1}$ and $r_\iota < r_{\text{loc}}$ for all $\iota = 1, \ldots, s$. According to (44), at the $(s+1)$th-iteration, conditioned on the event $\mathcal{E}_s(r_s) \cap \mathcal{E}_*(\bar{r}_*) \cap \mathcal{E}_*(\tilde{r}_*) \cap \mathcal{E}_{\text{loc}} \cap \mathcal{F}(r_s)$,

$$\left\|\widehat{\boldsymbol{\beta}}_\upsilon^{(s+1)} - \boldsymbol{\beta}_\tau)\right\|_{\boldsymbol{\Sigma}} \leq r_{s+1} := \bar{\kappa}^{-1}\vartheta(\gamma) \cdot r_s + \bar{\kappa}^{-1} r_*. \tag{46}$$

Note that $r_{s+1} \le \bar{\kappa}^{-1}\vartheta(\gamma)\cdot r_{s-1} + \bar{\kappa}^{-1} r_* = r_s < r_{\rm loc}$. Therefore, $\widehat{\boldsymbol{\beta}}_v^{(s+1)}$ falls into the interior of $\mathbb{B}_{\boldsymbol{\Sigma}}(r_{\rm loc})$, which implies $\widehat{\boldsymbol{\beta}}^{(s+1)} = \widehat{\boldsymbol{\beta}}_v^{(s+1)} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_{\rm loc})$ and certificates the event $\mathcal{E}_{s+1}(r_{s+1}) = \left\{\widehat{\boldsymbol{\beta}}^{(s+1)} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_{s+1})\right\}$. Consequently, $\widehat{\boldsymbol{\beta}}^{(s+1)} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_{s+1}) \subset \mathbb{B}_{\boldsymbol{\Sigma}}(r_s)$ and the bound (46) also applies to $\widehat{\boldsymbol{\beta}}^{(s+1)}$ with $r_{s+1} \le r_s$.

Repeat the above arguments until $s = S \ge 1$. Since for every $1 \le s \le S$, $\mathcal{E}_s(r_s)$ holds under the event $\mathcal{E}_{s-1}(r_{s-1}) \cap \mathcal{E}_*(\bar{r}_*) \cap \mathcal{E}_*(\tilde{r}_*) \cap \mathcal{E}_{\rm loc} \cap \mathcal{F}(r_{s-1})$. Conditioned on the event $\mathcal{E}_0(r_0) \cap \mathcal{E}_*(\bar{r}_*) \cap \mathcal{E}_*(\tilde{r}_*) \cap \mathcal{E}_{\rm loc} \cap \{\cap_{s=0}^{S-1} \mathcal{F}(r_{s-1})\}$, $\widehat{\boldsymbol{\beta}}^{(S)}$ satisfies the bounds

$$\left\|\widehat{\boldsymbol{\beta}}^{(S)} - \boldsymbol{\beta}_\tau\right\|_{\boldsymbol{\Sigma}} \le \bar{\kappa}^{-1}\vartheta(\gamma)\cdot r_{S-1} + \bar{\kappa}^{-1} r_* = r_S \le r_{S-1}. \tag{47}$$

Since $r_s = \{\bar{\kappa}^{-1}\vartheta(\gamma)\}^s r_0 + [1 - \{\bar{\kappa}^{-1}\vartheta(\gamma)\}^s] / [1 - \bar{\kappa}^{-1}\vartheta(\gamma)]\, \bar{\kappa}^{-1} r_*$ for $s = 1, \ldots, S$. Then $S = \lceil \log(r_0/r_*)/\log(\bar{\kappa}/\vartheta(\gamma))\rceil + 1$ is the smallest integer such that $\{\bar{\kappa}^{-1}\vartheta(\gamma)\}^{S-1} r_0 \le r_*$.

Finally, together with (43) and (45)–(47), conditioned on the event $\mathcal{E}_0(r_0) \cap \mathcal{E}_*(\bar{r}_*) \cap \mathcal{E}_*(\tilde{r}_*)$,

$$\left\|\widehat{\boldsymbol{\beta}}^{(S)} - \boldsymbol{\beta}_\tau\right\|_{\boldsymbol{\Sigma}} \le \bar{\kappa}^{-1}\vartheta(\gamma)\cdot r_* + \frac{1}{\bar{\kappa} - \vartheta(\gamma)} r_* \lesssim r_*$$

holds with probability at least $1 - (2S+1)e^{-\gamma} - S\varepsilon$. This completes the proof. ∎

3.3 Proof of Theorem 3

Theorem 3 (Restatement of Theorem 3 from main text)**.** *Assume that Conditions (A1)–(A7) hold, and $r_0, \lambda_* > 0$. For any $\gamma, \varepsilon > 0$, there exists a constant $C_\varepsilon > 0$ such that the bandwidths $h_l \ge h > 0$ and the regularization parameter $\lambda = 2.5(\lambda_* + \varsigma) > 0$ satisfy $d^{1/2}\lambda \lesssim h_l \lesssim 1$, and*

$$\varsigma \asymp \max\left\{\frac{r_0}{h}\sqrt{\frac{d\,(\log\, p + \gamma)}{M}}, d^{-1/2}\left(\sqrt{\frac{p+\gamma}{m}} + \frac{C_\varepsilon}{h_l}\sqrt{\frac{p^3}{M}} + h_l + 1 - L^{-1}\right) r_0 + d^{-1/2}h^2\right\}.$$

Then, conditioned on the event $\mathcal{E}_0(r_0) \cap \mathcal{E}_*(\bar{\lambda}_*) \cap \mathcal{E}_*(\tilde{\lambda}_*)$, there exists a constant $M_0 > 0$ such that for all $M \ge M_0$, the one-step regularized estimator computed on the l-th machine satisfies $\widetilde{\boldsymbol{\beta}}_l^{(1)} \in \mathbb{C}_{\boldsymbol{\Sigma}}(d)$, and with probability at least $1 - 3e^{-\gamma} - \varepsilon$,

$$\|\widetilde{\boldsymbol{\beta}}_l^{(1)} - \boldsymbol{\beta}_\tau\|_{\boldsymbol{\Sigma}} \lesssim C\left(\frac{d}{h}\sqrt{\frac{\log\, p + \gamma}{M}} + \sqrt{\frac{p+\gamma}{m}} + \frac{C_\varepsilon}{h_l}\sqrt{\frac{p^3}{M}} + h_l + 1 - L^{-1}\right) r_0 + d^{1/2}\lambda_* + h^2,$$

where $C > 0$ is a constant depending on $(\tau, n_u, f_u, \kappa_u, \mu_3, a_1, B, \xi_p, \xi_{\min})$.

Proof. Denote $\widetilde{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}_l^{(1)}$ and $\widetilde{\boldsymbol{\zeta}} = \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\tau$. Based on the first-order optimality condition, there exists a subgradient $\widetilde{\boldsymbol{g}} \in \partial\|\widetilde{\boldsymbol{\beta}}\|_1$ such that $\nabla\widetilde{\mathcal{L}}_{M,l}^{(0)}(\widetilde{\boldsymbol{\beta}}) + \lambda\cdot\widetilde{\boldsymbol{g}} = \boldsymbol{0}$ and $\widetilde{\boldsymbol{g}}^{\rm T}\widetilde{\boldsymbol{\beta}} = \|\widetilde{\boldsymbol{\beta}}\|_1$. Therefore,

$$\langle\widetilde{\boldsymbol{g}}, \boldsymbol{\beta}_\tau - \widetilde{\boldsymbol{\beta}}\rangle \le \|\boldsymbol{\beta}_\tau\|_1 - \|\widetilde{\boldsymbol{\beta}}\|_1 = \|\boldsymbol{\beta}_{\tau,\mathcal{A}}\|_1 - \|\widetilde{\boldsymbol{\zeta}}_{\mathcal{A}^c}\|_1 - \|\widetilde{\boldsymbol{\zeta}}_{\mathcal{A}} + \boldsymbol{\beta}_{\tau,\mathcal{A}}\|_1 \le \|\widetilde{\boldsymbol{\zeta}}_{\mathcal{A}}\|_1 - \|\widetilde{\boldsymbol{\zeta}}_{\mathcal{A}^c}\|_1.$$

This implies that

$$\begin{aligned} 0 \le \bar{D}_{\widetilde{\mathcal{L}}_{M,l}}^{(0)}(\widetilde{\boldsymbol{\beta}}, \boldsymbol{\beta}_\tau) &= \left\langle\nabla\widetilde{\mathcal{L}}_{M,l}^{(0)}(\widetilde{\boldsymbol{\beta}}) - \nabla\widetilde{\mathcal{L}}_{M,l}^{(0)}(\boldsymbol{\beta}_\tau),\ \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\tau\right\rangle \\ &= \lambda\cdot\langle\widetilde{\boldsymbol{g}},\ \boldsymbol{\beta}_\tau - \widetilde{\boldsymbol{\beta}}\rangle - \langle\nabla\widetilde{\mathcal{L}}_{M,l}^{(0)}(\boldsymbol{\beta}_\tau),\ \widetilde{\boldsymbol{\zeta}}\rangle \\ &\le \lambda\cdot(\|\widetilde{\boldsymbol{\zeta}}_{\mathcal{A}}\|_1 - \|\widetilde{\boldsymbol{\zeta}}_{\mathcal{A}^c}\|_1) - \langle\nabla\widetilde{\mathcal{L}}_{M,l}^{(0)}(\boldsymbol{\beta}_\tau),\ \widetilde{\boldsymbol{\zeta}}\rangle, \end{aligned} \tag{48}$$

where $\nabla \widetilde{\mathcal{L}}_{M,l}^{(0)}(\boldsymbol{\beta}_\tau) = \nabla \mathcal{L}_{m,l}^{(0)}(\boldsymbol{\beta}_\tau) - \nabla \mathcal{L}_{m,l}^{(0)}(\widetilde{\boldsymbol{\beta}}^{(0)}) + \nabla \mathcal{L}_{M}^{(0)}(\widetilde{\boldsymbol{\beta}}^{(0)})$. Decompose $\nabla \widetilde{\mathcal{L}}_{M,l}^{(0)}(\boldsymbol{\beta}_\tau)$ as

$$\begin{aligned}
&\nabla \widetilde{\mathcal{L}}_{M,l}^{(0)}(\boldsymbol{\beta}_\tau) \\
&= \widehat{\boldsymbol{U}}_M(\widetilde{\boldsymbol{\beta}}^{(0)}) - \widehat{\boldsymbol{D}}_{m,l}(\widetilde{\boldsymbol{\beta}}^{(0)})(\widetilde{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau) \\
&= \Big\{\widehat{\boldsymbol{U}}_M(\widetilde{\boldsymbol{\beta}}^{(0)}) - \boldsymbol{U}_M(\widetilde{\boldsymbol{\beta}}^{(0)})\Big\} + \Big\{\wp_M(\widetilde{\boldsymbol{\beta}}^{(0)}) - \Im_M(\widetilde{\boldsymbol{\beta}}^{(0)})\Big\} \\
&\quad + \Big\{\boldsymbol{U}_M(\boldsymbol{\beta}_\tau) - \boldsymbol{J}_M(\boldsymbol{\beta}_\tau)\Big\} + \boldsymbol{J}_M(\boldsymbol{\beta}_\tau) + \Big\{\boldsymbol{J}_M(\widetilde{\boldsymbol{\beta}}^{(0)}) - \boldsymbol{J}_M(\boldsymbol{\beta}_\tau) - \boldsymbol{H}_M(\boldsymbol{\beta}_\tau)(\widetilde{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau)\Big\} \\
&\quad + \Big\{\boldsymbol{H}_M(\boldsymbol{\beta}_\tau) - \boldsymbol{H}_M(\widetilde{\boldsymbol{\beta}}^{(0)})\Big\}(\widetilde{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau) + \Big\{\boldsymbol{H}_M(\widetilde{\boldsymbol{\beta}}^{(0)}) - \boldsymbol{H}_{m,l}(\widetilde{\boldsymbol{\beta}}^{(0)})\Big\}(\widetilde{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau) \\
&\quad + \Big\{\boldsymbol{H}_{m,l}(\widetilde{\boldsymbol{\beta}}^{(0)}) - \boldsymbol{D}_{m,l}(\widetilde{\boldsymbol{\beta}}^{(0)})\Big\}(\widetilde{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau) + \Big\{\boldsymbol{D}_{m,l}(\widetilde{\boldsymbol{\beta}}^{(0)}) - \widehat{\boldsymbol{D}}_{m,l}(\widetilde{\boldsymbol{\beta}}^{(0)})\Big\}(\widetilde{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}_\tau).
\end{aligned}$$

For $r > 0$, define

$$\begin{aligned}
\mathfrak{I}_M(r) &= \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \big\|E[\Delta_M(\boldsymbol{\beta})]\big\|_2, \quad \mathfrak{H}_M(r) = \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \big\|\mathfrak{F}_M(\boldsymbol{\beta})\big\|_2, \\
\mathfrak{H}_{M,l}(r) &= \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \big\|\mathfrak{F}_{M,l}(\boldsymbol{\beta})\big\|_2, \quad \mathfrak{G}_{m,l}(r) = \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \big\|\mathfrak{H}_{m,l}(\boldsymbol{\beta})\big\|_2, \\
\mathfrak{D}_{m,l}(r) &= \sup_{\boldsymbol{\beta}\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)} \big\|\mathfrak{E}_{m,l}(\boldsymbol{\beta})\big\|_2.
\end{aligned}$$

Based on the Hölder's inequality, under the condition of the event $\mathcal{E}_*(\bar{\lambda}_*) \cap \mathcal{E}_*(\tilde{\lambda}_*) \cap \mathcal{E}_0(r_0)$,

$$\begin{aligned}
\Big|\big\langle \nabla \widetilde{\mathcal{L}}_{M,l}^{(0)}(\boldsymbol{\beta}_\tau),\ \widetilde{\boldsymbol{\zeta}}\big\rangle\Big| \le \big\{\varXi_M(r_0) + \lambda_*\big\} \cdot \big\|\widetilde{\boldsymbol{\zeta}}\big\|_1 \\
+ \big\{\mathfrak{I}_M(r_0) + \omega_M^* + \mathfrak{H}_M(r_0) + \mathfrak{H}_{M,l}(r_0) \\
+ \mathfrak{G}_{m,l}(r_0) + \mathfrak{D}_{m,l}(r_0)\big\} \cdot \big\|\widetilde{\boldsymbol{\zeta}}\big\|_{\boldsymbol{\Sigma}}.
\end{aligned} \tag{49}$$

Take $\lambda = 2.5\ (\lambda_* + \varsigma)$ with ς satisfying

$$\varsigma \ge \max\Big\{\varXi_M(r_0), \big[\mathfrak{I}_M(r_0) + \omega_M^* + \mathfrak{H}_M(r_0) + \mathfrak{H}_{M,l}(r_0) + \mathfrak{G}_{m,l}(r_0) + \mathfrak{D}_{m,l}(r_0)\big] d^{-1/2}\Big\}, \tag{50}$$

so that

$$\begin{aligned}
&\varXi_M(r_0) + \lambda_* \le 0.4\ \lambda, \\
&\mathfrak{I}_M(r_0) + \omega_M^* + \mathfrak{H}_M(r_0) + \mathfrak{H}_{M,l}(r_0) + \mathfrak{G}_{m,l}(r_0) + \mathfrak{D}_{m,l}(r_0) \le 0.4\ d^{1/2}\lambda.
\end{aligned} \tag{51}$$

Together with (48), (49), and (51), we have

$$0 \le 1.4\ \big\|\widetilde{\boldsymbol{\zeta}}_{\mathcal{A}}\big\|_1 - 0.6\ \big\|\widetilde{\boldsymbol{\zeta}}_{\mathcal{A}^c}\big\|_1 + 0.4\ d^{1/2}\big\|\widetilde{\boldsymbol{\zeta}}\big\|_{\boldsymbol{\Sigma}}.$$

Therefore,

$$\big\|\widetilde{\boldsymbol{\zeta}}\big\|_1 \le \frac{10}{3}\ \big\|\widetilde{\boldsymbol{\zeta}}_{\mathcal{A}}\big\|_1 + \frac{2}{3}\ d^{1/2}\big\|\widetilde{\boldsymbol{\zeta}}\big\|_{\boldsymbol{\Sigma}} \le 4\ d^{1/2}\big\|\widetilde{\boldsymbol{\zeta}}\big\|_{\boldsymbol{\Sigma}}.$$

This implies that $\widetilde{\boldsymbol{\beta}} \in \mathbb{C}_{\boldsymbol{\Sigma}}(d)$.

Assume the event $\mathcal{E}_*(\bar{\lambda}_*) \cap \mathcal{E}_*(\tilde{\lambda}_*) \cap \mathcal{E}_0(\lambda_0)$ occurs. Define $\widetilde{\boldsymbol{\beta}}_\upsilon = \boldsymbol{\beta}_\tau + \upsilon(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_\tau)$ with $0 < \upsilon \le 1$. For $r_{\text{loc}} = h_l/(4\iota_{0.25})$, we have $\widetilde{\boldsymbol{\beta}}_\upsilon \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_{\text{loc}}) \cap \mathbb{C}_{\boldsymbol{\Sigma}}(d)$ under the requirement (50) on ς. Based on (48), we have

$$\begin{aligned}
\bar{D}_{\widetilde{\mathcal{L}}_{M,l}}^{(0)}(\widetilde{\boldsymbol{\beta}}_\upsilon, \boldsymbol{\beta}_\tau) &\le \upsilon \cdot \bar{D}_{\widetilde{\mathcal{L}}_{M,l}}^{(0)}(\widetilde{\boldsymbol{\beta}}, \boldsymbol{\beta}_\tau) \le \upsilon\,\Big(1.4\ \lambda\big\|\widetilde{\boldsymbol{\zeta}}_{\mathcal{A}}\big\|_1 + 0.4\ d^{1/2}\lambda\big\|\widetilde{\boldsymbol{\zeta}}\big\|_{\boldsymbol{\Sigma}}\Big) \\
&\le 1.8\ d^{1/2}\lambda\ \big\|\widetilde{\boldsymbol{\beta}}_\upsilon - \boldsymbol{\beta}_\tau\big\|_{\boldsymbol{\Sigma}}.
\end{aligned}$$

From Lemma 9, as long as $(d\ \log\ p+\gamma)/m \lesssim h_l \lesssim 1$,

$$\bar{D}^{(0)}_{\widetilde{\mathcal{L}}_{M,l}}(\widetilde{\boldsymbol{\beta}}_\upsilon, \boldsymbol{\beta}_\tau) \geq 0.5\kappa_* n_l f_{l,l} C_m^{-1} \big\|\widetilde{\boldsymbol{\beta}}_\upsilon - \boldsymbol{\beta}_\tau\big\|^2_{\boldsymbol{\Sigma}}$$

holds with probability at least $1-e^{-\gamma}$. Therefore,

$$\big\|\widetilde{\boldsymbol{\beta}}_\upsilon - \boldsymbol{\beta}_\tau\big\|_{\boldsymbol{\Sigma}} \leq 3.6 C_m (\kappa_* n_l f_{l,l})^{-1} d^{1/2} \lambda. \tag{52}$$

At follows, we choose a sufficiently large λ (or ς) satisfying (50).

According to Lemma 10,

$$\Xi_M(r_0) \lesssim \frac{r_0}{h}\sqrt{\frac{d\ (\log\ p+\gamma)}{M}}$$

holds with probability at least $1-e^{-\gamma}$. Based on Lemmas 3–8, as long as $r_0 \lesssim h_l$ and $h_l \geq h$,

$$\begin{aligned}
&\mathfrak{I}_M(r_0) + \omega_M^* + \mathfrak{H}_M(r_0) + \mathfrak{H}_{M,l}(r_0) + \mathfrak{G}_{m,l}(r_0) + \mathfrak{D}_{m,l}(r_0) \\
&\lesssim C_1\ r_0^2 + C_2\ h^2 + C_3\ r_0^2 + \left[C_4(1-L^{-1}) + C_5|h_l - h|\right] r_0 \\
&\quad + C_6\ r_0 \sqrt{\frac{p+\gamma}{m}} + C_7\ r_0 \frac{C_\varepsilon}{h_l}\sqrt{\frac{p^3}{M}} \\
&\lesssim C\left(\sqrt{\frac{p+\gamma}{m}} + \frac{C_\varepsilon}{h_l}\sqrt{\frac{p^3}{M}} + h_l + 1 - L^{-1}\right) r_0 + C' h^2
\end{aligned}$$

holds with probability at least $1-e^{-\gamma}-\varepsilon$ (for some constants $C, C' > 0$). Therefore, after choosing a sufficiently large ς, which is of the order

$$\varsigma \asymp \max\left\{\frac{r_0}{h}\sqrt{\frac{d\ (\log\ p+\gamma)}{M}}, d^{-1/2}\left(\sqrt{\frac{p+\gamma}{m}} + \frac{C_\varepsilon}{h_l}\sqrt{\frac{p^3}{M}} + h_l + 1 - L^{-1}\right) r_0 + d^{-1/2} h^2\right\},$$

(50) holds with high probability. Then, conditioned on the event $\mathcal{E}_*(\bar{\lambda}_*) \cap \mathcal{E}_*(\tilde{\lambda}_*) \cap \mathcal{E}_0(\lambda_0)$, $\widetilde{\boldsymbol{\beta}}_\upsilon$ satisfies (52) with probability at least $1-3e^{-\gamma}-\varepsilon$. Based on the above choice of ς, the right-hand side of (52) is strictly less than $r_{\rm loc}$ provided that $h_l > 14.4\iota_{0.25}(\kappa_* n_l f_{l,l})^{-1} C_m d^{1/2}\lambda$. This implies that $\widetilde{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}_\upsilon \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_{\rm loc})$; and $\widetilde{\boldsymbol{\beta}}$ satisfies the bound (52). ■

3.4 Proof of Theorem 4

Theorem 4 (Restatement of Theorem 4 from main text)**.** *Assume the Conditions in Theorem 3 hold. For any $\gamma, \varepsilon > 0$ and a constant $C_\varepsilon > 0$, choose the local and global bandwidths as $h_l \asymp (C_\varepsilon^2 p^3/M)^{1/4}$ and $h \asymp \{d\ (\log\ p+\gamma)/M\}^{1/4}$. For $r_0, \lambda_* > 0$, let $r_* = d^{1/2}\lambda_*$ and $\lambda_s = 2.5(\lambda_* + \varsigma_s) > 0$ for $s \geq 2$ with*

$$\varsigma_s \asymp \max\left\{\vartheta^s d^{-1/2} r_0 + \vartheta d^{-1/2}(r_* + h^2),\ \sqrt{\frac{\log\ p+\gamma}{M}}\right\},$$

where $\vartheta \asymp C\max\{d^3(\log\ p+\gamma)/M, C_\varepsilon^2 p^3/M\}^{1/4} + \sqrt{(p+\gamma)/m} + 1 - L^{-1}$. Assume $m \gtrsim p+\gamma$, $M \gtrsim \max\{d^3(\log\ p+\gamma), C_\varepsilon^2 p^3\}$, $r_0 \lesssim \min\{1, (C_\varepsilon^2 p^3/M)^{1/4}\}$, and $r_* \lesssim h_l \lesssim 1$.

Then, conditioned on the event $\mathcal{E}_0(r_0) \cap \mathcal{E}_*(\bar{\lambda}_*) \cap \mathcal{E}_*(\tilde{\lambda}_*)$, there exist a constant $M_0 > 0$ such that for all $M \geq M_0$, the S-step regularized estimator $\widetilde{\boldsymbol{\beta}}_l^{(S)}$ computed on the l-th machine satisfies

$$\|\widetilde{\boldsymbol{\beta}}_l^{(S)} - \boldsymbol{\beta}_\tau\|_{\boldsymbol{\Sigma}} \lesssim d^{1/2}\lambda_* + h^2 \quad \textit{and} \quad \|\widetilde{\boldsymbol{\beta}}_l^{(S)} - \boldsymbol{\beta}_\tau\|_1 \lesssim d\lambda_* + d^{1/2}h^2$$

with probability at least $1 - 3Se^{-\gamma} - S\varepsilon$, provided that the number of iterations satisfies $S \gtrsim \log(r_0/r_*)/\log(1/\vartheta)$.

Proof. We carry out the whole proof under the condition of the event $\mathcal{E}_0(\lambda_0) \cap \mathcal{E}_*(\bar{\lambda}_*) \cap \mathcal{E}_*(\tilde{\lambda}_*)$ for some prespecified $r_0, \lambda_* > 0$; and write $r_* = d^{1/2}\lambda_*$. Recall that in the proof of Theorem 3, the regularization parameter λ_1 needs to be sufficiently large to obtain the desired error bound for the first iterate $\widetilde{\boldsymbol{\beta}}^{(1)} = \widetilde{\boldsymbol{\beta}}_l^{(1)}$ on the lth machine. Given $\gamma > 0$, let $\lambda_1 = 2.5\ (\lambda_* + \varsigma_1)$, where $\varsigma_1 > 0$ is of the order

$$\varsigma_1 \asymp \max\left\{\frac{r_0}{h}\sqrt{\frac{d\ (\log\ p + \gamma)}{M}}, d^{-1/2}\left(\sqrt{\frac{p+\gamma}{m}} + \frac{C_\varepsilon}{h_l}\sqrt{\frac{p^3}{M}} + h_l + 1 - L^{-1}\right) r_0 + d^{-1/2}h^2\right\}.$$

As long as the bandwidth $h_l \gtrsim \max(r_0, r_*) > 0$ satisfy

$$r_* + \max\left\{\frac{dr_0}{h}\sqrt{\frac{\log\ p + \gamma}{M}}, \left(\sqrt{\frac{p+\gamma}{m}} + \frac{C_\varepsilon}{h_l}\sqrt{\frac{p^3}{M}} + h_l + 1 - L^{-1}\right) r_0 + h^2\right\} \lesssim h_l \lesssim 1,$$

the first iterate $\widetilde{\boldsymbol{\beta}}^{(1)}$ satisfies $\widetilde{\boldsymbol{\beta}}^{(1)} \in \mathbb{C}_{\boldsymbol{\Sigma}}(d)$; and with probability at least $1 - 2e^{-\gamma} - \varepsilon$,

$$\begin{aligned}\|\widetilde{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_\tau\|_{\boldsymbol{\Sigma}} \leq C &\left(\frac{d}{h}\sqrt{\frac{\log\ p + \gamma}{M}} + \sqrt{\frac{p+\gamma}{m}} + \frac{C_\varepsilon}{h_l}\sqrt{\frac{p^3}{M}} + h_l + 1 - L^{-1}\right) r_0 \\ &+ C'\left(r_* + h^2\right) = r_1.\end{aligned}$$

Let

$$\begin{aligned}\vartheta &= \vartheta\ (d, p, m, M, L, h_l, h, \gamma, C_\varepsilon) \\ &= C\left\{\frac{d}{h}\sqrt{\frac{\log\ p + \gamma}{M}} + \sqrt{\frac{p+\gamma}{m}} + \frac{C_\varepsilon}{h_l}\sqrt{\frac{p^3}{M}} + h_l + 1 - L\right\}\end{aligned}$$

be the contraction factor. After setting $h_l \asymp (C_\varepsilon^2\ p^3/M)^{1/4}$ and $h \asymp \{d\ (\log\ p + \gamma)/M\}^{1/4}$, we have

$$\vartheta \asymp C\left\{d^{3/4}\ \left(\frac{\log\ p + \gamma}{M}\right)^{1/4} + \left(\frac{C_\varepsilon^2\ p^3}{M}\right)^{1/4} + \sqrt{\frac{p+\gamma}{m}} + 1 - L\right\}$$

and

$$\varsigma_1 \asymp \max\left\{\vartheta d^{-1/2} r_0,\ \sqrt{\frac{\log\ p + \gamma}{M}}\right\}.$$

Therefore, if $r_0 \lesssim \min\{1, (C_\varepsilon^2 p^3/M)^{1/4}\}$, $d^{1/2}\varsigma_1 \lesssim h_l$. To ensure the contraction factor ϑ strictly less than 1, we need $m \gtrsim p + \gamma$ and $M \gtrsim\ \max\{d^3(\log\ p + \gamma), C_\varepsilon^2 p^3\}$. Then, the one-step estimator reduces the estimation error of $\widetilde{\boldsymbol{\beta}}^{(0)}$ by a factor of ϑ.

For $s \geq 2$, define the events $\mathcal{E}_s(r_s) := \{\widetilde{\boldsymbol{\beta}}^{(s)} \in \mathbb{B}_{\boldsymbol{\Sigma}}(r_s) \cap \mathbb{C}_{\boldsymbol{\Sigma}}(d)\}$, where

$$r_s := \vartheta r_{s-1} + C'(r_* + h^2) = \vartheta^s r_0 + C' \frac{1-\vartheta^s}{1-\vartheta}(r_* + h^2). \tag{53}$$

Set $\lambda_s = 3\,\varsigma_s$ at iteration $s \geq 2$ with

$$\varsigma_s \asymp \max\left\{\vartheta d^{-1/2} r_{s-1},\ \sqrt{\frac{\log\ p + \gamma}{M}}\right\}. \tag{54}$$

Then, based on (53) and (54), for $s \geq 2$,

$$\varsigma_s \asymp \max\left\{\vartheta^s d^{-1/2} r_0 + \vartheta d^{-1/2}(r_* + h^2),\ \sqrt{\frac{\log\ p + \gamma}{M}}\right\}.$$

Note that under the conditions on r_0 and (m, M), $d^{1/2}\varsigma_s \lesssim h_l$ for $s \geq 2$. According to Theorem 3, conditioned on the event $\mathcal{E}_{s-1}(r_{s-1}) \cap \mathcal{E}_*(\bar{\lambda}_*) \cap \mathcal{E}_*(\tilde{\lambda}_*)$, the s-step iterate $\widetilde{\boldsymbol{\beta}}^{(s)}$ satisfies $\widetilde{\boldsymbol{\beta}}^{(s)} \in \mathbb{C}_{\boldsymbol{\Sigma}}(d)$; and with probability at least $1 - 3e^{-\gamma} - \varepsilon$,

$$\|\widetilde{\boldsymbol{\beta}}^{(s)} - \boldsymbol{\beta}_\tau\|_{\boldsymbol{\Sigma}} \leq \vartheta r_{s-1} + C'(r_* + h^2) = r_s = \vartheta^s r_0 + C' \frac{1-\vartheta^s}{1-\vartheta}(r_* + h^2).$$

Let $S = \lceil \log(r_0/r_*)/\log(1/\vartheta) \rceil$ be the smallest integer such that $\vartheta^S r_0 \leq r_*$. Under the condition $\mathcal{E}_0(\lambda_0) \cap \mathcal{E}_*(\bar{\lambda}_*) \cap \mathcal{E}_*(\tilde{\lambda}_*)$, the Sth iterate $\widetilde{\boldsymbol{\beta}}^{(S)}$ satisfies the error bounds

$$\|\widetilde{\boldsymbol{\beta}}^{(S)} - \boldsymbol{\beta}_\tau\|_{\boldsymbol{\Sigma}} \lesssim d^{1/2}\lambda^* + h^2 \quad \text{and} \quad \|\widehat{\boldsymbol{\beta}}^{(S)} - \boldsymbol{\beta}_\tau\|_1 \lesssim d\lambda^* + d^{1/2}h^2$$

with probability at least $1 - 3Se^{-\gamma} - S\varepsilon$ based on the union bound over $s = 1, 2, \ldots, S$. This completes the proof of the theorem. ■

References

[1] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Oxford Univ. Press, 2013.

[2] O. Bousquet. Concentration inequalities for sub-additive functions using the entropy method. *In Stochastic Inequalities and Applications. Progress in Probability*, volume 56, pages 213–247, Birkhäuser,Basel, 2003.

[3] Jianqing Fan, Han Liu, Qiang Sun, and Tong Zhang. I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error. *The Annals of Statistics*, 46(2):814–841, 2018.

[4] Xuming He, Xiaoou Pan, Kean Ming Tan, and Wen-Xin Zhou. Smoothed quantile regression with large-scale inference. *Journal of Econometrics*, 232(2):367–388, 2023.

[5] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes.* Springer-Verlag, 1991.

[6] Yanan Song, Haohui Han, Liya Fu, and Ting Wang. Penalized weighted smoothed quantile regression for high-dimensional longitudinal data. *Statistics in Medicine*, 43(10):2007–2042, 2024.

[7] V. Spokoiny. Parametric estimation. Finite sample theory. *The Annals of Statistics*, 40:2877–2909, 2012.

[8] V. Spokoiny. Bernstein-von mises theorem for growing parameter dimension. *Available at arXiv:1302.3430*, 2013.