Supplementary Material for "Estimation and Inference for Density-convoluted Support Vector Machine with Streaming Data"

Haochen Rao, Xu Guo, Heng Lian and Haobo Qi

This supplementary material contains all the proof of the main theoretical results in the main text. In addition, we demonstrate the technical lemmas used in the main proof. Some illustrations and instructions of the methodology are also provided.

S1 Proof of Main Theorem

In this appendix, we prove the theorems, lemmas and propositions in Section 3.

Proof of Lemma 1. Denote $\widetilde{K}(u) = \int_{-\infty}^{u} K(t)$. For any unit vector $\boldsymbol{\nu} \in \mathbb{R}^p$ we have

$$\begin{aligned}\boldsymbol{\nu}^\top \boldsymbol{S}_h(\boldsymbol{\beta}^0) &= \mathbb{E}\left\{y\boldsymbol{\nu}^\top \mathbf{x}\widetilde{K}\left(\frac{1-y\mathbf{x}^\top\boldsymbol{\beta}^0}{h}\right)\right\} \\ &= \pi_+ \int_{\mathbb{R}^p} \boldsymbol{\nu}^\top \mathbf{x}\widetilde{K}\left(\frac{1-\mathbf{x}^\top\boldsymbol{\beta}^0}{h}\right) f(x)\mathrm{d}x - \pi_- \int_{\mathbb{R}^p} \boldsymbol{\nu}^\top \mathbf{x}\widetilde{K}\left(\frac{1+\mathbf{x}^\top\boldsymbol{\beta}^0}{h}\right) g(x)\mathrm{d}x.\end{aligned}$$

We focus on conditional expectation given $y = 1$, that is

$$\begin{aligned}I_1^* &= \int_{\mathbb{R}^p} \boldsymbol{\nu}^\top \mathbf{x}\widetilde{K}\left(\frac{1-\mathbf{x}^\top\boldsymbol{\beta}^0}{h}\right) f(x)\mathrm{d}x \\ &= \int_{\mathbb{R}^{p-1}}\int_{\mathbb{R}} \boldsymbol{\nu}^\top \mathbf{x}\widetilde{K}\left(\frac{1-\mathbf{x}^\top\boldsymbol{\beta}^0}{h}\right) f(x_d|\mathbf{x}_{-d}) f(\mathbf{x}_{-d})\mathrm{d}x_d\mathrm{d}\mathbf{x}_{-d}.\end{aligned}$$

Analogous derivations hold for $y = -1$. Let $u = (1 - \mathbf{x}^\top \boldsymbol{\beta}^0)/h$, so $x_d = (1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0 - uh)/\beta_d^0$. By a change of variables,

$$\begin{aligned} I_1^* =& -\frac{h}{\beta_d^0} \int \left(\nu_d \frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0 - uh}{\beta_d^0} \right) \widetilde{K}(u) f\left(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0 - uh}{\beta_d^0} \middle| \mathbf{x}_{-d} \right) f(\mathbf{x}_{-d}) \mathrm{d}u \mathrm{d}\mathbf{x}_{-d} \\ & - \frac{h}{\beta_d^0} \int_{\mathbb{R}^{p-1}} \int_{\mathbb{R}} \left(\boldsymbol{\nu}_{-d}^\top \mathbf{x}_{-d} \right) \widetilde{K}(u) f\left(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0 - uh}{\beta_d^0} \middle| \mathbf{x}_{-d} \right) f(\mathbf{x}_{-d}) \mathrm{d}u \mathrm{d}\mathbf{x}_{-d} \\ =:& I_d + I_{-d}. \end{aligned}$$

Next, we show that

$$I_d = -\nu_d \mathbb{E}_{\mathbf{x}|y=1} \left\{ y x_d \mathbb{I}(1 - y\mathbf{x}^\top \boldsymbol{\beta}^0 \geq 0) \right\} + \widetilde{I}_d, \qquad |\widetilde{I}_d| \leq \frac{C h^2 \kappa_2 |\nu_d|}{2(\beta_d^0)^2}, \tag{S1.1}$$

$$I_{-d} = -\boldsymbol{\nu}_{-d}^\top \mathbb{E}_{\mathbf{x}|y=1} \left\{ y \mathbf{x}_{-d} \mathbb{I}(1 - y\mathbf{x}^\top \boldsymbol{\beta}^0 \geq 0) \right\} + \widetilde{I}_{-d}, \qquad |\widetilde{I}_{-d}| \leq \frac{C h^2 \kappa_2 \mu_1}{2(\beta_d^0)^2}. \tag{S1.2}$$

Let $G(t) = \int_{-\infty}^t x f(x|\mathbf{x}_{-d}) \mathrm{d}x$. By the Leibniz rule,

$$\mathrm{d}G\left(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0 - uh}{\beta_d^0} \right) = -\frac{h}{\beta_d^0} \left(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0 - uh}{\beta_d^0} \right) f\left(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0 - uh}{\beta_d^0} \middle| \mathbf{x}_{-d} \right).$$

Substituting into the expression for I_d

$$I_d = \int_{\mathbb{R}^p} \widetilde{K}(u) \mathrm{d}G\left(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0 - uh}{\beta_d^0} \right) \mathrm{d}u = \int_{\mathbb{R}^p} K(u) G\left(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0 - uh}{\beta_d^0} \right) \mathrm{d}u$$

where the last equality was obtained using integration by parts and the assumption of finite moments of $\mathbf{x}$. By a first-order Taylor expansion around $(1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0)/\beta_d^0$,

$$G\left(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0 - uh}{\beta_d^0} \right) = G\left(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0} \right) + \mathrm{d}G\left(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0} \right) \left(\frac{-hu}{\beta_d^0} \right) + \mathcal{R}, \tag{S1.3}$$

where the remainder $\mathcal{R}$ has the following form

$$\mathcal{R} = -\frac{hu}{\beta_d^0}\int_0^1 \left\{ \mathrm{d}G\left(\frac{1-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0} + t\frac{-hu}{\beta_d^0}\right) - \mathrm{d}G\left(\frac{1-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0}\right)\right\}\mathrm{d}t.$$

Since $xf(x|\mathbf{x}_{-d})$ is bounded and continuous by Assumption 4, for any $a, b \in \mathbb{R}$, $|af(a|\mathbf{x}_{-d}) - bf(b|\mathbf{x}_{-d})| \leq C|a-b|$ for some $C > 0$. By a change of variables ($w = thu$), we have

$$\begin{aligned}
|\mathcal{R}| &= \left|\frac{1}{\beta_d^0}\int_0^{hu}\left\{\mathrm{d}G\left(\frac{1-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0} - \frac{w}{\beta_d^0}\right) - \mathrm{d}G\left(\frac{1-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0}\right)\right\}\mathrm{d}w\right| \\
&\leq \frac{1}{\beta_d^0}\int_0^{hu}\left|\mathrm{d}G\left(\frac{1-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0} - \frac{w}{\beta_d^0}\right) - \mathrm{d}G\left(\frac{1-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0}\right)\right|\mathrm{d}w \\
&\leq C\int_0^{hu}\frac{|w|}{(\beta_d^0)^2}\mathrm{d}w \leq C\frac{(hu)^2}{2(\beta_d^0)^2}.
\end{aligned} \tag{S1.4}$$

Using (S1.3), we have

$$\begin{aligned}
I_d =& -\nu_d\int_{\mathbb{R}^{p-1}}\int_{\mathbb{R}} G\left(\frac{1-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0}\right)K(u)f(\mathbf{x}_{-d})\mathrm{d}u\mathrm{d}\mathbf{x}_{-d} \\
&+\nu_d\int_{\mathbb{R}^{p-1}}\int_{\mathbb{R}}\mathrm{d}G\left(\frac{1-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0}\right)\frac{hu}{\beta_d^0}K(u)f(\mathbf{x}_{-d})\mathrm{d}u\mathrm{d}\mathbf{x}_{-d} \\
&-\nu_d\int_{\mathbb{R}^{p-1}}\int_{\mathbb{R}}\mathcal{R}K(u)f(\mathbf{x}_{-d})\mathrm{d}u\mathrm{d}\mathbf{x}_{-d} \\
=&: I_{d1} + I_{d2} + I_{d3}.
\end{aligned}$$

Note that $\int_{\mathbb{R}} K(u)\mathrm{d}u = 1$, thus

$$\begin{aligned} I_{d1} =& -\nu_d \int_{\mathbb{R}^{p-1}} G\left(\frac{1-\mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}\right) f(\mathbf{x}_{-d})\mathrm{d}\mathbf{x}_{-d} \\ =& -\nu_d \int_{\mathbb{R}^{p-1}} \int_{-\infty}^{\frac{1-\mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}} x_d f(x_d|\mathbf{x}_{-d}) f(\mathbf{x}_{-d})\mathrm{d}x_d\mathrm{d}\mathbf{x}_{-d} \\ =& -\nu_d \int_{\mathbb{R}^p} \mathbb{I}\left(\frac{1-\mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0} \geq x_d\right) x_d f(x_d|\mathbf{x}_{-d}) f(\mathbf{x}_{-d})\mathrm{d}x \\ =& -\nu_d \mathbb{E}_{\mathbf{x}|y=1}\left\{ y x_d \mathbb{I}(1 - y\mathbf{x}^\top \boldsymbol{\beta}^0 \geq 0)\right\}. \end{aligned} \tag{S1.5}$$

By the expression of $\mathrm{d}G$, we can write I_{d2} as

$$I_{d2} = \nu_d \int_{\mathbb{R}^{p-1}} \int_{\mathbb{R}} \frac{hu}{\beta_d^0} \left(\frac{1-\mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}\right) f\left(\frac{1-\mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}\middle|\mathbf{x}_{-d}\right) K(u) f(\mathbf{x}_{-d})\mathrm{d}u\mathrm{d}\mathbf{x}_{-d}. \tag{S1.6}$$

Note that $\int_{\mathbb{R}} uK(u)\mathrm{d}u = 0$ by the symmetry of the kernel density. Hence $I_{d2} = 0$.

Using modulus inequality, by (S1.4),

$$\begin{aligned} |I_{d3}| \leq& \int_{\mathbb{R}^{p-1}} \int_{\mathbb{R}} |\nu_d \mathcal{R} K(u) f(\mathbf{x}_{-d})|\mathrm{d}u\mathrm{d}\mathbf{x}_{-d} \\ \leq& |\nu_d| C \frac{h^2}{2(\beta_d^0)^2} \int_{\mathbb{R}^{p-1}} \left\{ \int_{\mathbb{R}} u^2 K(u)\mathrm{d}u \right\} f(\mathbf{x}_{-d})\mathrm{d}\mathbf{x}_{-d} \\ \leq& \kappa_2 C \frac{h^2}{2(\beta_d^0)^2}. \end{aligned} \tag{S1.7}$$

As a result, combining (S1.5), (S1.6) and (S1.7), we have demonstrate (S1.1).

To prove (S1.2), integrating by part yields that

$$I_{-d} = -\int_{\mathbb{R}} \boldsymbol{\nu}_{-d}^\top \mathbf{x}_{-d} K(u) F\left(\frac{1-\mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0 - uh}{\beta_d^0}\right) \mathrm{d}u.$$

Similarly, using Taylor expansion of $F\left\{(1-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0-uh)/\beta_d^0\right\}$ around $(1-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0)/\beta_d^0$, we have

$$\begin{aligned}
I_{-d} =& -\int_{\mathbb{R}^{p-1}}\boldsymbol{\nu}_{-d}^\top\mathbf{x}_{-d}\int_{\mathbb{R}}F\left(\frac{1-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0}\right)K(u)f(\mathbf{x}_{-d})\mathrm{d}u\mathrm{d}\mathbf{x}_{-d}\\
&+\int_{\mathbb{R}^{p-1}}\boldsymbol{\nu}_{-d}^\top\mathbf{x}_{-d}\int_{\mathbb{R}}f\left(\frac{1-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0}\right)\frac{hu}{\beta_d^0}K(u)f(\mathbf{x}_{-d})\mathrm{d}u\mathrm{d}\mathbf{x}_{-d}\\
&-\int_{\mathbb{R}^{p-1}}\boldsymbol{\nu}_{-d}^\top\mathbf{x}_{-d}\int_{\mathbb{R}}\mathcal{R}_-K(u)f(\mathbf{x}_{-d})\mathrm{d}u\mathrm{d}\mathbf{x}_{-d}\\
=&: I_{-d1}+I_{-d2}+I_{-d3}.
\end{aligned}$$

Likewise, we have

$$\begin{aligned}
&I_{-d1}=-\boldsymbol{\nu}_{-d}^\top\mathbb{E}_{\mathbf{x}|y=1}\left\{y\mathbf{x}_{-d}\mathbb{I}(1-y\mathbf{x}^\top\boldsymbol{\beta}^0\geq 0)\right\},\qquad I_{-d2}=0,\\
&|I_{-d3}|\leq C\frac{h^2}{2(\beta_d^0)^2}\int_{\mathbb{R}^{p-1}}|\boldsymbol{\nu}_{-d}^\top\mathbf{x}_{-d}|\left\{\int_{\mathbb{R}}u^2K(u)\mathrm{d}u\right\}f(\mathbf{x}_{-d})\mathrm{d}\mathbf{x}_{-d}\leq\frac{Ch^2\kappa_2\mu_1}{2(\beta_d^0)^2}.
\end{aligned}$$

Hence (S1.2) holds. Consequently, adding (S1.1) and (S1.2), by triangle inequality we have

$$\left|I_1^*+\boldsymbol{\nu}^\top\mathbb{E}_{\mathbf{x}|y=1}\left\{y\mathbf{x}\mathbb{I}(1-y\mathbf{x}^\top\boldsymbol{\beta}^0\geq 0)\right\}\right|\leq|\widetilde{I}_d|+|\widetilde{I}_{-d}|\leq\frac{Ch^2\kappa_2(1+\mu_1)}{2(\beta_d^0)^2}.$$

By analogous derivations for $y=-1$ and since $\mathbb{E}\left\{y\mathbf{x}\mathbb{I}(1-y\mathbf{x}^\top\boldsymbol{\beta}^0\geq 0)\right\}=0$, we obtain that for any unit vector $\boldsymbol{\nu}\in\mathbb{R}^p$,

$$\left|\boldsymbol{\nu}^\top\boldsymbol{S}_h(\boldsymbol{\beta}^0)\right|\leq\frac{Ch^2\kappa_2(1+\mu_1)}{2(\beta_d^0)^2}.$$

Letting $\boldsymbol{\nu}=\boldsymbol{S}_h(\boldsymbol{\beta}^0)/\|\boldsymbol{S}_h(\boldsymbol{\beta}^0)\|_2$ then completes the proof. □

Proof of Theorem 1. As the minimizer of the optimization (2.4), $\widehat{\boldsymbol{\beta}}_h$ satisfies the KKT condition $\widehat{S}_h(\widehat{\boldsymbol{\beta}}_h) + \lambda\boldsymbol{\phi} = 0$, where $\boldsymbol{\phi}$ is a subgradient of $\|\cdot\|_1$ at $\widehat{\boldsymbol{\beta}}_h$. Denote $\widehat{\boldsymbol{\xi}} = \widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^0$, we then have

$$\begin{aligned}
&\langle \widehat{\boldsymbol{S}}_h(\widehat{\boldsymbol{\beta}}_h) - \widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^0), -\widehat{\boldsymbol{\xi}}\rangle \\
=&\lambda\langle \boldsymbol{\phi}, -\widehat{\boldsymbol{\xi}}\rangle + \langle \widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^0;\mathcal{D}) - \boldsymbol{S}_h(\boldsymbol{\beta}^0), -\widehat{\boldsymbol{\xi}}\rangle + \langle \boldsymbol{S}_h(\boldsymbol{\beta}^0), -\widehat{\boldsymbol{\xi}}\rangle \\
=:&U_1 + U_2 + U_3. \qquad \text{(S1.8)}
\end{aligned}$$

By the definition of subgradient, using the triangle inequality we have

$$U_1 \le \lambda(\|\boldsymbol{\beta}^0\|_1 - \|\widehat{\boldsymbol{\beta}}_h\|_1) = \lambda(\|\boldsymbol{\beta}_{0,S_0}\|_1 - \|\widehat{\boldsymbol{\xi}} + \boldsymbol{\beta}^0\|_1) \le \lambda(\|\widehat{\boldsymbol{\xi}}_{S_0}\|_1 - \|\widehat{\boldsymbol{\xi}}_{S_0^c}\|_1), \qquad \text{(S1.9)}$$

where $\boldsymbol{\beta}_{0,S_0}$ is the sub-vector of $\boldsymbol{\beta}^0$ on the signal set S_0. Moreover, by Lemma 2, given $\lambda = O\big\{\sqrt{\log(p)/n}\big\}$, we know that $\|\widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^0;\mathcal{D}) - \boldsymbol{S}_h(\boldsymbol{\beta}^0)\|_\infty \le \lambda$ with probability approaching 1. Thus, by Hölder's inequality, with a high probability we have

$$|U_2| \le \|\widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^0;\mathcal{D}) - \boldsymbol{S}_h(\boldsymbol{\beta}^0)\|_\infty \|\widehat{\boldsymbol{\xi}}\|_1 \le \lambda\|\widehat{\boldsymbol{\xi}}\|_1. \qquad \text{(S1.10)}$$

By lemma 1 and Hölder's inequality, we have with high probability,

$$|U_3| \le \|\boldsymbol{S}_h(\boldsymbol{\beta}^0)\|_2 \|\widehat{\boldsymbol{\xi}}\|_2 \le Ch^2\|\widehat{\boldsymbol{\xi}}\|_2. \qquad \text{(S1.11)}$$

Combining (S1.9), (S1.10), (S1.11) and (S1.8), we have

$$\begin{aligned}
\langle \widehat{\boldsymbol{S}}_h(\widehat{\boldsymbol{\beta}}_h) - \widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^0), -\widehat{\boldsymbol{\xi}}\rangle \le&\lambda(\|\widehat{\boldsymbol{\xi}}_{S_0}\|_1 - \|\widehat{\boldsymbol{\xi}}_{S_0^c}\|_1) + \lambda\|\widehat{\boldsymbol{\xi}}\|_1 + Ch^2\|\widehat{\boldsymbol{\xi}}\|_2 \\
\le&(2\lambda s_0^{1/2} + Ch^2)\|\widehat{\boldsymbol{\xi}}\|_2.
\end{aligned}$$

Note that $h^2 \lesssim \{s_0 \log(p)/n\}^{1/2} \asymp s_0^{1/2}\lambda$, thus we have $\langle \widehat{\boldsymbol{S}}_h(\widehat{\boldsymbol{\beta}}_h) - \widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^0), -\widehat{\boldsymbol{\xi}}\rangle \leq C'\lambda s_0^{1/2}\|\widehat{\boldsymbol{\xi}}\|_2$. Moreover, using similar results of (S2.57) in Lemma 5, we get

$$\langle \widehat{\boldsymbol{S}}_h(\widehat{\boldsymbol{\beta}}_h) - \widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^0), -\widehat{\boldsymbol{\xi}}\rangle \geq C''\|\widehat{\boldsymbol{\xi}}\|_2^2,$$

provided the condition $s_0^2 \log(p)/h^2 = o(n)$. It then follows that

$$\|\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^0\|_2 = O_p\big[\{s_0 \log(p)/n\}^{1/2}\big].$$

Moreover, using the fact $\langle \nabla\widehat{\boldsymbol{S}}_h(\widehat{\boldsymbol{\beta}}_h) - \nabla\widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^0), -\widehat{\boldsymbol{\xi}}\rangle \geq 0$, we know that

$$0 \leq \lambda(\|\widehat{\boldsymbol{\xi}}_{S_0}\|_1 - \|\widehat{\boldsymbol{\xi}}_{S_0^c}\|_1) + \lambda\|\widehat{\boldsymbol{\xi}}\|_1 + Ch^2\|\widehat{\boldsymbol{\xi}}\|_2.$$

Applying Cauchy-Schwartz inequality, it then follows that

$$\lambda\|\widehat{\boldsymbol{\xi}}\|_1 \leq 3\lambda\|\widehat{\boldsymbol{\xi}}_{S_0}\|_1 + Ch^2\|\widehat{\boldsymbol{\xi}}\|_2 \leq \big(3\lambda s_0^{1/2} + Ch^2\big)\|\widehat{\boldsymbol{\xi}}\|_2.$$

Recall that $h^2 \lesssim s_0^{1/2}\lambda$. Thus we conclude that $\|\widehat{\boldsymbol{\xi}}\|_1 \lesssim C' s_0^{1/2}\|\widehat{\boldsymbol{\xi}}\|_2$ and

$$\|\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^0\|_1 = O_p\big[s_0\{\log(p)/n\}^{1/2}\big].$$

□

Proof of Lemma 2. By the union bound,

$$\begin{aligned}
&P\left\{\|\widehat{\boldsymbol{S}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \boldsymbol{S}_{h_l}(\boldsymbol{\beta}^0)\|_\infty > \theta_l^{-1}\lambda_l\right\} \\
\leq &\sum_{j=1}^{p} P\left\{c_0\sqrt{\frac{\log(p)}{n_l}} < \frac{1}{n_l}\left|\sum_{i=1}^{n_l} L'_{h_l}(y_{li}\mathbf{x}_{li}^\top\boldsymbol{\beta}^0)y_{li}x_{li,j} - \mathbb{E}\big\{L'_{h_l}(y_{li}\mathbf{x}_{li}^\top\boldsymbol{\beta}^0)y_{li}x_{li,j}\big\}\right|\right\}.
\end{aligned} \tag{S1.12}$$

Further, by Assumption 1 and $|L_h^{'}(\cdot)| \leq 1$, we know that for each $j \in \{1, \ldots, p\}$, $L_{h_l}^{'}(y_{li}\mathbf{x}_{li}^{\top}\boldsymbol{\beta}^0)y_{li}\mathbf{x}_{li,j}, i = 1, \ldots, n_l$ are *i.i.d* bounded random variables. By Hoeffding's inequality, we then have

$$P\left\{\frac{1}{n_l}\left|\sum_{i=1}^{n_l} L_{h_l}^{'}(y_{li}\mathbf{x}_{li}^{\top}\boldsymbol{\beta}^0)y_{li}x_{li,j} - \mathbb{E}\big\{L_{h_l}^{'}(y_{li}\mathbf{x}_{li}^{\top}\boldsymbol{\beta}^0)y_{li}x_{li,j}\big\}\right| > c_0\sqrt{\frac{\log(p)}{n_l}}\right\}$$

$$\leq 2\exp\big(-\lambda_l^2 n_l/u_{\mathbf{x}}^2\big) = 2p^{-c_0^2/(2u_{\mathbf{x}}^2)}, \tag{S1.13}$$

Combining (S1.12) and (S1.13) yields

$$P\left\{\|\widehat{\boldsymbol{S}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \boldsymbol{S}_{h_l}(\boldsymbol{\beta}^0)\|_\infty > \theta_l^{-1}\lambda_l\right\} \leq 2p^{1-c_0^2/(2u_{\mathbf{x}}^2)}.$$

□

Proof of Theorem 2. We prove this theorem through mathematical induction. To be specific, the proof is divided into the following 3 steps:

(i). $\|\widehat{\boldsymbol{\beta}}_{h_1}^1 - \boldsymbol{\beta}^0\|_2 \leq c_1^{(2)}\sqrt{s_0}\lambda_1$ and $\|\widehat{\boldsymbol{\beta}}_{h_1}^1 - \boldsymbol{\beta}^0\|_1 \leq c_1^{(1)}s_0\lambda_1$ for some constants $c_1^{(2)}, c_1^{(1)} > 0$.

(ii). Given (i), $\|\widehat{\boldsymbol{\beta}}_{h_2}^2 - \boldsymbol{\beta}^0\|_2 \leq c_2^{(2)}\sqrt{s_0}\lambda_2$ and $\|\widehat{\boldsymbol{\beta}}_{h_2}^2 - \boldsymbol{\beta}^0\|_1 \leq c_2^{(1)}s_0\lambda_2$ for some constants $c_2^{(2)}, c_2^{(1)} > 0$.

(iii). Given that $\widehat{\boldsymbol{\beta}}_{h_1}^1, \widehat{\boldsymbol{\beta}}_{h_2}^2, \ldots, \widehat{\boldsymbol{\beta}}_{h_{b-1}}^{b-1}$ satisfy the error bound, then $\|\widehat{\boldsymbol{\beta}}_{h_b}^b - \boldsymbol{\beta}^0\|_2 \leq c_b^{(2)}\sqrt{s_0}\lambda_b$ and $\|\widehat{\boldsymbol{\beta}}_{h_b}^b - \boldsymbol{\beta}^0\|_1 \leq c_b^{(1)}s_0\lambda_b$ for some constants $c_b^{(2)}, c_b^{(1)} > 0$.

For (i), the theoretical results in Theorem 1 can be readily applied under the offline setting. Assume that $\lambda_1 = O\big\{\sqrt{\log(p)/n_1}\big\}$ and $h_1 = O\big\{\{s_0 \log(p)/n_1\}^{1/4}\big\}$, we have

$$\|\widehat{\boldsymbol{\beta}}^1_{h_1} - \boldsymbol{\beta}^0\|_2 \leq c_1^{(2)}\sqrt{s_0}\lambda_1, \quad \|\widehat{\boldsymbol{\beta}}^1_{h_1} - \boldsymbol{\beta}^0\|_1 \leq c_1^{(1)} s_0 \lambda_1.$$

For (ii), recall that $\widehat{\boldsymbol{\beta}}^2_{h_2}$ is the minimizer of optimization problem as

$$\widehat{\boldsymbol{\beta}}^2_{h_2} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \big\{\mathcal{L}_2(\boldsymbol{\beta}) + \lambda_2\|\boldsymbol{\beta}\|_1\big\},$$

where

$$\mathcal{L}_2(\boldsymbol{\beta}) = \frac{1}{N_2}\left\{\frac{1}{2}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^1_{h_1})^\top \widehat{\mathbf{Q}}_1(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^1_{h_1}) + n_2 L_{h_2}(\boldsymbol{\beta}; \mathcal{D}_2)\right\}.$$

Let $\nabla\mathcal{L}_2(\boldsymbol{\beta}) = \{\widehat{\mathbf{Q}}_1(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^1_{h_1}) + n_2\widehat{\boldsymbol{S}}_{h_2}(\boldsymbol{\beta}; \mathcal{D}_2)\}/N_2$ be the first-order derivation of $\mathcal{L}_2(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. To demonstrate the ℓ_2-norm error bound for $\widehat{\boldsymbol{\beta}}^2_{h_2}$, we aim to prove the following inequality

$$C\|\widehat{\boldsymbol{\beta}}^2_{h_2} - \boldsymbol{\beta}^0\|_2^2 \leq \langle\nabla\mathcal{L}_2(\widehat{\boldsymbol{\beta}}^2_{h_2}) - \nabla\mathcal{L}_2(\boldsymbol{\beta}^0), \widehat{\boldsymbol{\beta}}^2_{h_2} - \boldsymbol{\beta}^0\rangle \leq C'\lambda_2\sqrt{s_0}\|\widehat{\boldsymbol{\beta}}^2_{h_2} - \boldsymbol{\beta}^0\|_2. \tag{S1.14}$$

We first derive the upper bound for the right side of (S1.14). By the KKT condition, there exists a subgradient $\boldsymbol{\kappa}_2 \in \partial\|\widehat{\boldsymbol{\beta}}^2_{h_2}\|_1$ such that $\nabla\mathcal{L}_2(\widehat{\boldsymbol{\beta}}^2_{h_2}) + \lambda_2\boldsymbol{\kappa}_2 = 0$.

Define $\widehat{\boldsymbol{\xi}}_j = \widehat{\boldsymbol{\beta}}^j_{h_j} - \boldsymbol{\beta}^0, j = 1, 2$. We then derive

$$\begin{aligned}
&\langle \nabla \mathcal{L}_2(\widehat{\boldsymbol{\beta}}^2_{h_2}) - \nabla \mathcal{L}_2(\boldsymbol{\beta}^0), \widehat{\boldsymbol{\beta}}^2_{h_2} - \boldsymbol{\beta}^0 \rangle \\
=&\lambda_2 \langle \boldsymbol{\kappa}_2, -\widehat{\boldsymbol{\xi}}_2 \rangle + \frac{n_2}{N_2} \langle \widehat{\boldsymbol{S}}_{h_2}(\boldsymbol{\beta}^0; \mathcal{D}_2) - \boldsymbol{S}_{h_2}(\boldsymbol{\beta}^0), -\widehat{\boldsymbol{\xi}}_2 \rangle + \frac{n_2}{N_2} \langle \boldsymbol{S}_{h_2}(\boldsymbol{\beta}^0), -\widehat{\boldsymbol{\xi}}_2 \rangle + \frac{1}{N_2} \langle \widehat{\mathbf{Q}}_1 \widehat{\boldsymbol{\xi}}_1, \widehat{\boldsymbol{\xi}}_2 \rangle \\
=:& A_1 + A_2 + A_3 + A_4. \qquad \text{(S1.15)}
\end{aligned}$$

For A_1, by the definition of subgradient, we have

$$A_1 \leq \lambda_2 (\|\boldsymbol{\beta}^0\|_1 - \|\widehat{\boldsymbol{\beta}}^2_{h_2}\|_1) = \lambda_2 (\|\boldsymbol{\beta}_{0,S_0}\|_1 - \|\widehat{\boldsymbol{\xi}}_2 + \boldsymbol{\beta}^0\|_1) \leq \lambda_2 (\|\widehat{\boldsymbol{\xi}}_{2,S_0}\|_1 - \|\widehat{\boldsymbol{\xi}}_{2,S_0^c}\|_1), \qquad \text{(S1.16)}$$

where the last inequality holds by the Triangle inequality, $S_0^c = \{1, \ldots, p\}/S_0$.

For A_2, by lemma 2, let $\lambda_2 = c_0 \sqrt{\log(p)/N_2}$, we know that $\theta_2 \|\widehat{\boldsymbol{S}}_{h_2}(\boldsymbol{\beta}^0; \mathcal{D}_2) - \boldsymbol{S}_{h_2}(\boldsymbol{\beta}^0)\|_\infty \leq \lambda_2/3$ with probability $1 - 2p^{-c_0^2/(36u_{\mathbf{x}}^2)}$. Thus, by Hölder's inequality, with a high probability we have

$$|A_2| \leq \frac{n_2}{N_2} \|\widehat{\boldsymbol{S}}_{h_2}(\boldsymbol{\beta}^0; \mathcal{D}_2) - \boldsymbol{S}_{h_2}(\boldsymbol{\beta}^0)\|_\infty \|\widehat{\boldsymbol{\xi}}_2\|_1 \leq \frac{n_2}{N_2} \theta_2^{-1} \lambda_2 \|\widehat{\boldsymbol{\xi}}_2\|_1/3 \leq \lambda_2 \|\widehat{\boldsymbol{\xi}}_2\|_1/3. \qquad \text{(S1.17)}$$

For A_3, by lemma 1 and Hölder's inequality, we have

$$|A_3| \leq \frac{n_2}{N_2} \|\boldsymbol{S}_{h_2}(\boldsymbol{\beta}^0)\|_2 \|\widehat{\boldsymbol{\xi}}_2\|_2 \leq C' h_2^2 \|\widehat{\boldsymbol{\xi}}_2\|_2. \qquad \text{(S1.18)}$$

For A_4, we have

$$\begin{aligned}
A_4 =& \frac{1}{N_2}\widehat{\boldsymbol{\xi}}_1^\top \widehat{\mathbf{Q}}_1 \widehat{\boldsymbol{\xi}}_2 \\
=& \frac{n_1}{N_2}\widehat{\boldsymbol{\xi}}_1^\top \{\widehat{\mathbf{H}}_{h_1}(\widehat{\boldsymbol{\beta}}_{h_1}^1; \mathcal{D}_1) - \widehat{\mathbf{H}}_{h_1}(\boldsymbol{\beta}^0; \mathcal{D}_1)\}\widehat{\boldsymbol{\xi}}_2 + \frac{n_1}{N_2}\widehat{\boldsymbol{\xi}}_1^\top \{\widehat{\mathbf{H}}_{h_1}(\boldsymbol{\beta}^0; \mathcal{D}_1) - \mathbf{H}_{h_1}(\boldsymbol{\beta}^0)\}\widehat{\boldsymbol{\xi}}_2 \\
&+ \frac{n_1}{N_2}\widehat{\boldsymbol{\xi}}_1^\top \{\mathbf{H}_{h_1}(\boldsymbol{\beta}^0) - \mathbf{H}(\boldsymbol{\beta}^0)\}\widehat{\boldsymbol{\xi}}_2 + \frac{n_1}{N_2}\widehat{\boldsymbol{\xi}}_1^\top \{\mathbf{H}(\boldsymbol{\beta}^0)\}\widehat{\boldsymbol{\xi}}_2 \\
=:& A_{41} + A_{42} + A_{43} + A_{44}. \qquad \text{(S1.19)}
\end{aligned}$$

Denote $\boldsymbol{J}_1 = \widehat{\mathbf{H}}_{h_1}(\widehat{\boldsymbol{\beta}}_{h_1}^1; \mathcal{D}_1) - \widehat{\mathbf{H}}_{h_1}(\boldsymbol{\beta}^0; \mathcal{D}_1)$, $\boldsymbol{J}_2 = \widehat{\mathbf{H}}_{h_1}(\boldsymbol{\beta}^0; \mathcal{D}_1) - \mathbf{H}_{h_1}(\boldsymbol{\beta}^0)$ and $\boldsymbol{J}_3 = \mathbf{H}_{h_1}(\boldsymbol{\beta}^0) - \mathbf{H}(\boldsymbol{\beta}^0)$, representing the estimation error, the approximation error and smoothing error respectively. Given that $\|\widehat{\boldsymbol{\xi}}_1\|_1 \leq c_1^{(1)} s_0 \sqrt{\log(p)/n_1}$, by Lemma 3, we have

$$\begin{aligned}
\|\boldsymbol{J}_1\|_{\max} =& \|\widehat{\mathbf{H}}_{h_1}(\widehat{\boldsymbol{\beta}}_{h_1}^1; \mathcal{D}_1) - \widehat{\mathbf{H}}_{h_1}(\boldsymbol{\beta}^0; \mathcal{D}_1)\|_{\max} \\
\leq& C_1 \left\{ s_0 \sqrt{\frac{\log^2(p)}{n_1^2 h_l^3}} + s_0 \sqrt{\frac{\log^3(p)}{n_1^3 h_l^4}} + s_0 \sqrt{\frac{\log(p)}{n_1}} + \frac{s_0^2 \log(p)}{n_1 h_l^3} \right\}, \qquad \text{(S1.20)} \\
\|\boldsymbol{J}_2\|_{\max} =& \|\widehat{\mathbf{H}}_{h_1}(\boldsymbol{\beta}^0; \mathcal{D}_1) - \mathbf{H}_h(\boldsymbol{\beta}^0)\|_{\max} \leq C_2 \sqrt{\frac{\log(p)}{n_1 h_1^2}}, \qquad \text{(S1.21)} \\
\|\boldsymbol{J}_3\|_{\max} =& \|\mathbf{H}_h(\boldsymbol{\beta}^0) - \mathbf{H}(\boldsymbol{\beta}^0)\|_{\max} \leq C_3 h_1. \qquad \text{(S1.22)}
\end{aligned}$$

Note that by our choice of tuning parameter, $\lambda_1 \sqrt{n_1} \asymp \lambda_2 \sqrt{N_2}$. Combining (S1.20), (S1.21), (S1.22) with (S1.19), by lemma 4, using Hölder's and Cauchy-Schwartz inequalities, we have:

$$
\begin{aligned}
|A_4| \leq & \frac{n_1}{N_2}\|\widehat{\boldsymbol{\xi}}_1\|_1 \left\{\|\boldsymbol{J}_1\|_{\max} + \|\boldsymbol{J}_2\|_{\max} + \|\boldsymbol{J}_3\|_{\max}\right\} \|\widehat{\boldsymbol{\xi}}_2\|_1 + \frac{n_1}{N_2}\|\mathbf{H}(\boldsymbol{\beta}^0)\widehat{\boldsymbol{\xi}}_1\|_2\|\widehat{\boldsymbol{\xi}}_2\|_2 \\
\leq & Cc_1^{(1)}s_0^2\left\{\sqrt{\frac{\log^2(p)}{n_1^2h_1^3}} + \sqrt{\frac{\log^3(p)}{n_1^3h_1^4}} + \sqrt{\frac{\log(p)}{n_1}} + \frac{s_0\log(p)}{n_1h_1^3}\right\}\lambda_2\|\widehat{\boldsymbol{\xi}}_2\|_1 \\
& + Cc_1^{(1)}s_0\left\{\sqrt{\frac{\log(p)}{n_1h_1^2}} + h_1\right\}\lambda_2\|\widehat{\boldsymbol{\xi}}_2\|_1 + \tau_{\max}\big\{\mathbf{H}(\boldsymbol{\beta}^0)\big\}c_1^{(2)}\sqrt{s_0}\lambda_2\|\widehat{\boldsymbol{\xi}}_2\|_2. \quad \text{(S1.23)}
\end{aligned}
$$

The second inequality holds because $\|\widehat{\boldsymbol{\xi}}_1\|_1 \leq c_1^{(1)}s_0\sqrt{\log(p)/n_1}$, which is ensured by step (i). Given $h_1 = O\big[\{s_0\log(p)/n_1\}^{1/4}\big]$ and Assumption 2, under the condition that $s_0^9\log(p) = o(n_1)$, we know that (S1.23) is dominated by $\lambda_2\|\widehat{\boldsymbol{\xi}}_2\|_1/3 + Mc_1^{(2)}\sqrt{s_0}\lambda_2\|\widehat{\boldsymbol{\xi}}_2\|_2$. Then, by triangle and Cauchy-Schwartz inequalities, combining (S1.15), (S1.16), (S1.17), (S1.18) and (S1.23), we have

$$
\begin{aligned}
& \langle\nabla\mathcal{L}_2(\widehat{\boldsymbol{\beta}}_{h_2}^2) - \nabla\mathcal{L}_2(\boldsymbol{\beta}^0), \widehat{\boldsymbol{\beta}}_{h_2}^2 - \boldsymbol{\beta}^0\rangle \\
\leq & \lambda_2(\|\widehat{\boldsymbol{\xi}}_{2,S_0}\|_1 - \|\widehat{\boldsymbol{\xi}}_{2,S_0^c}\|_1) + \frac{2}{3}\lambda_2\|\widehat{\boldsymbol{\xi}}_2\|_1 + (Mc_1^{(2)}\sqrt{s_0}\lambda_2 + C'h_2^2)\|\widehat{\boldsymbol{\xi}}_2\|_2 \\
= & \lambda_2\Big(\frac{5}{3}\|\widehat{\boldsymbol{\xi}}_{2,S_0}\|_1 - \frac{1}{3}\|\widehat{\boldsymbol{\xi}}_{2,S_0^c}\|_1\Big) + (Mc_1^{(2)}\sqrt{s_0}\lambda_2 + C'h_2^2)\|\widehat{\boldsymbol{\xi}}_2\|_2 \\
\leq & 2\lambda_2\sqrt{s_0}\|\widehat{\boldsymbol{\xi}}_2\|_2 + (Mc_1^{(2)}\sqrt{s_0}\lambda_2 + C'h_2^2)\|\widehat{\boldsymbol{\xi}}_2\|_2 \\
= & (C\lambda_2\sqrt{s_0} + C'h_2^2)\|\widehat{\boldsymbol{\xi}}_2\|_2, \quad \text{(S1.24)}
\end{aligned}
$$

where $C = 2 + Mc_1^{(1)}$. Under conditions $h_2 = O\big[\{s_0\log(p)/N_2\}^{1/4}\big]$, the inequality (S1.24) shows that the right side of (S1.14) holds.

Additionally, we demonstrate the restricted conditions of $\widehat{\boldsymbol{\xi}}_2$. Note that $\langle\nabla\mathcal{L}_2(\widehat{\boldsymbol{\beta}}_{h_2}^2) - \nabla\mathcal{L}_2(\boldsymbol{\beta}^0), \widehat{\boldsymbol{\beta}}_{h_2}^2 - \boldsymbol{\beta}^0\rangle \geq 0$, thus by the Cauchy-Schwartz inequality, (S1.24) indicates

the following cone-type constraint for $\widehat{\boldsymbol{\xi}}_2$:

$$\begin{aligned}\|\widehat{\boldsymbol{\xi}}_2\|_1 \leq & 6\sqrt{s_0}\|\widehat{\boldsymbol{\xi}}_2\|_2 + \big\{3(Mc_1^{(2)}\sqrt{s_0}\lambda_2)\sqrt{s_0} + C'h_2^2\lambda_2^{-1}\big\}\|\widehat{\boldsymbol{\xi}}_2\|_2 \\ = & \big\{3(2+Mc_1^{(2)})\sqrt{s_0} + h_2^2\lambda_2^{-1}\big\}\|\widehat{\boldsymbol{\xi}}_2\|_2. \end{aligned} \tag{S1.25}$$

Given $h_2 = O\big[\{s_0\log(p)/N_2\}^{1/4}\big]$, we know from (S1.25) that $\widehat{\boldsymbol{\xi}}_2$ is restricted in the cone $\mathbb{C}(l)$ with $l = 3(2+Mc_1^{(2)})$.

For the left side of (S1.14), suppose now $\|\widehat{\boldsymbol{\xi}}_2\|_2 > r$, then let $\gamma = r/\|\widehat{\boldsymbol{\xi}}_2\|_2 \in (0,1)$. Let $\widetilde{\boldsymbol{\beta}}_2 = \boldsymbol{\beta}^0 + \gamma\widehat{\boldsymbol{\xi}}_2$, we get $\|\widetilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}^0\|_2 = r$. Further note that $\widetilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}^0 = \gamma(\widehat{\boldsymbol{\beta}}^2_{h_2} - \boldsymbol{\beta}^0)$, thus $\widetilde{\boldsymbol{\beta}}_2 \in \boldsymbol{\beta}^0 + \mathbb{C}(l)$. By Lemma 5, we then have with probability approaching 1

$$\langle \nabla\mathcal{L}_2(\widetilde{\boldsymbol{\beta}}_2) - \nabla\mathcal{L}_2(\boldsymbol{\beta}^0), \widetilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}^0\rangle \geq M'\|\widetilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}^0\|_2^2,$$

where $M' = (M^{-1}+\rho)/2$. Similar to the (B1) in Tan et al. (2022), we have

$$\langle \nabla\mathcal{L}_2(\widetilde{\boldsymbol{\beta}}_2) - \nabla\mathcal{L}_2(\boldsymbol{\beta}^0), \widetilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}^0\rangle \leq \gamma\langle \nabla\mathcal{L}_2(\widehat{\boldsymbol{\beta}}^2_{h_2}) - \nabla\mathcal{L}_2(\boldsymbol{\beta}^0), \widetilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}^0\rangle.$$

Further from (S1.24) we have

$$\begin{aligned}\langle \nabla\mathcal{L}_2(\widetilde{\boldsymbol{\beta}}_2) - \nabla\mathcal{L}_2(\boldsymbol{\beta}^0), \widetilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}^0\rangle &\leq \gamma\langle \nabla\mathcal{L}_2(\widehat{\boldsymbol{\beta}}^2_{h_2}) - \nabla\mathcal{L}_2(\boldsymbol{\beta}^0), \widehat{\boldsymbol{\beta}}^2_{h_2} - \boldsymbol{\beta}^0\rangle \\ &\leq (2+Mc_1^{(2)})\lambda_2\sqrt{s_0}\|\widetilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}^0\|_2.\end{aligned}$$

We then conclude that

$$\|\widetilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}^0\|_2 \leq M'^{-1}(2+Mc_1^{(2)})\lambda_2\sqrt{s_0}.$$

Then as long as $2(2+Mc_1^{(2)})M_2^{-1}\lambda_2\sqrt{s_0} < r$, we get a contradiction. This implies that $\gamma = 1$ and further

$$\|\widehat{\boldsymbol{\xi}}_2\|_2 \leq 2M'^{-1}(2+Mc_1^{(2)})\lambda_2\sqrt{s_0}. \tag{S1.26}$$

The inequality (S1.26) yields the ℓ_2-norm error bound for $\widehat{\boldsymbol{\beta}}_{h_2}^2$. In addition, condition (S1.25) indicates that the ℓ_1-norm error bound is controlled by the ℓ_2-norm error bound. Therefore, in conclusion, by (S1.14) and (S1.25), we prove both the ℓ_1-norm and ℓ_2-norm error bound for $\widehat{\boldsymbol{\beta}}_{h_2}^2$, that is

$$\|\widehat{\boldsymbol{\beta}}_{h_2}^2 - \boldsymbol{\beta}^0\|_1 \leq c_2^{(1)} s_0\lambda_2, \qquad \|\widehat{\boldsymbol{\beta}}_{h_2}^2 - \boldsymbol{\beta}^0\|_2 \leq c_2^{(2)}\sqrt{s_0}\lambda_2.$$

Here $c_2^{(2)} = 2M'^{-1}(2+Mc_1^{(2)})$ and $c_2^{(1)} = 6M'^{-1}\big(2+Mc_1^{(2)}\big)^2$, where M' and M are constants.

Last but most important, the third step (iii) is demonstrated in the following context. Recall that $\widehat{\boldsymbol{\beta}}_h^b$ is the minimizer of the following objective function $\mathcal{L}_b(\boldsymbol{\beta})$ as

$$\mathcal{L}_b(\boldsymbol{\beta}) = \frac{1}{N_b}\left\{\frac{1}{2}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{h_{b-1}}^{b-1})^\top \widehat{\mathbf{Q}}_{b-1}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{h_{b-1}}^{b-1}) + n_b L_{h_b}(\boldsymbol{\beta}; \mathcal{D}_b)\right\}.$$

The gradient of $\mathcal{L}_b(\boldsymbol{\beta})$ is given by $\nabla\mathcal{L}_b(\boldsymbol{\beta}) = \{\widehat{\mathbf{Q}}_{b-1}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{h_{b-1}}^{b-1}) + n_b\boldsymbol{S}_{h_b}(\boldsymbol{\beta}; \mathcal{D}_b)\}/N_b$. It suffices to prove the subsequent conditions

$$C\|\widehat{\boldsymbol{\beta}}_{h_b}^b - \boldsymbol{\beta}^0\|_2^2 \leq \langle\nabla\mathcal{L}_b(\widehat{\boldsymbol{\beta}}_{h_b}^b) - \nabla\mathcal{L}_b(\boldsymbol{\beta}^0), \widehat{\boldsymbol{\beta}}_{h_b}^b - \boldsymbol{\beta}^0\rangle \leq C\lambda_b\sqrt{s_0}\|\widehat{\boldsymbol{\beta}}_{h_b}^b - \boldsymbol{\beta}^0\|_2, \tag{S1.27}$$

$$\|\widehat{\boldsymbol{\beta}}_{h_b}^b - \boldsymbol{\beta}^0\|_1 \leq C\sqrt{s_0}\|\widehat{\boldsymbol{\beta}}_{h_b}^b - \boldsymbol{\beta}^0\|_2. \tag{S1.28}$$

We first demonstrate the right side of (S1.27). Denote $\widehat{\boldsymbol{\xi}}_l = \widehat{\boldsymbol{\beta}}^l_{h_l} - \boldsymbol{\beta}^0, l = 1, \ldots, b$. By the KKT condition, there exists a subgradient $\boldsymbol{\kappa}_b \in \partial\|\widehat{\boldsymbol{\beta}}_{h_b}\|_1$ such that $\nabla\mathcal{L}_b(\widehat{\boldsymbol{\beta}}_{h_b}) + \lambda_b\boldsymbol{\kappa}_b = 0$. Then we derive

$$
\begin{aligned}
&\langle\nabla\mathcal{L}_b(\widehat{\boldsymbol{\beta}}_{h_b}) - \nabla\mathcal{L}_b(\boldsymbol{\beta}^0), \widehat{\boldsymbol{\beta}}_{h_b} - \boldsymbol{\beta}^0\rangle \\
=&\lambda_b\langle\boldsymbol{\kappa}_b, -\widehat{\boldsymbol{\xi}}_b\rangle + \frac{n_b}{N_b}\langle\widehat{\boldsymbol{S}}_{h_b}(\boldsymbol{\beta}^0; \mathcal{D}_b) - \boldsymbol{S}_{h_b}(\boldsymbol{\beta}^0), -\widehat{\boldsymbol{\xi}}_b\rangle + \frac{n_b}{N_b}\langle\boldsymbol{S}_{h_b}(\boldsymbol{\beta}^0), -\widehat{\boldsymbol{\xi}}_b\rangle + \frac{1}{N_b}\langle\widehat{\mathbf{Q}}_{b-1}\widehat{\boldsymbol{\xi}}_{b-1}, \widehat{\boldsymbol{\xi}}_{b-1}\rangle \\
:=&D_1 + D_2 + D_3 + D_4. \qquad \text{(S1.29)}
\end{aligned}
$$

By lemma 2, we know that the event $\chi_b = \{\theta_b\|\widehat{\boldsymbol{S}}_{h_b}(\boldsymbol{\beta}^0; \mathcal{D}_b) - \boldsymbol{S}_{h_b}(\boldsymbol{\beta}^0)\|_\infty \leq \lambda_b/3\}$ occurs with probability $1 - 2p^{-c_0^2/(9u_{\mathbf{x}}^2)}$ when $\lambda_b = c_0\sqrt{\log(p)/N_b}$. Similar to the previous proof of $b = 2$, we have that with high probability,

$$
D_1 \leq \lambda_b(\|\widehat{\boldsymbol{\xi}}_{b,S_0}\|_1 - \|\widehat{\boldsymbol{\xi}}_{b,S_0^c}\|_1), \quad |D_2| \leq \lambda_b\|\widehat{\boldsymbol{\xi}}_b\|_1/3, \quad |D_3| \leq C'h_b^2\|\widehat{\boldsymbol{\xi}}_b\|_2 \qquad \text{(S1.30)}
$$

For D_4, we can write

$$
\begin{aligned}
D_4 =&\frac{1}{N_b}\widehat{\boldsymbol{\xi}}_{b-1}^\top\widehat{\mathbf{Q}}_{b-1}\widehat{\boldsymbol{\xi}}_b \\
=&\frac{N_{b-1}}{N_b}\widehat{\boldsymbol{\xi}}_{b-1}^\top\left\{\frac{\widehat{\mathbf{Q}}_{b-1}}{N_{b-1}} - \frac{1}{N_{b-1}}\sum_{l=1}^{b-1} n_l\widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l)\right\}\widehat{\boldsymbol{\xi}}_b \\
&+ \frac{N_{b-1}}{N_b}\widehat{\boldsymbol{\xi}}_{b-1}^\top\left\{\frac{1}{N_{b-1}}\sum_{l=1}^{b-1} n_l\widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \widetilde{\mathbf{H}}_{b-1}(\boldsymbol{\beta}^0)\right\}\widehat{\boldsymbol{\xi}}_b \\
&+ \frac{N_{b-1}}{N_b}\widehat{\boldsymbol{\xi}}_{b-1}^\top\left\{\widetilde{\mathbf{H}}_{b-1}(\boldsymbol{\beta}^0) - \mathbf{H}(\boldsymbol{\beta}^0)\right\}\widehat{\boldsymbol{\xi}}_b + \frac{N_{b-1}}{N_b}\widehat{\boldsymbol{\xi}}_{b-1}^\top\left\{\mathbf{H}(\boldsymbol{\beta}^0)\right\}\widehat{\boldsymbol{\xi}}_b \\
:=&\frac{N_{b-1}}{N_b}\widehat{\boldsymbol{\xi}}_{b-1}^\top(\boldsymbol{K}_1 + \boldsymbol{K}_2 + \boldsymbol{K}_3)\widehat{\boldsymbol{\xi}}_b + \frac{N_{b-1}}{N_b}\widehat{\boldsymbol{\xi}}_{b-1}^\top\left\{\mathbf{H}(\boldsymbol{\beta}^0)\right\}\widehat{\boldsymbol{\xi}}_b. \qquad \text{(S1.31)}
\end{aligned}
$$

where $\widetilde{\mathbf{H}}_{b-1}(\boldsymbol{\beta}^0) = \frac{1}{N_{b-1}} \sum_{l=1}^{b-1} n_l \mathbf{H}_{h_l}(\boldsymbol{\beta}^0)$. For $\boldsymbol{K}_1$, by Lemma 1 we have

$$\begin{aligned}
\|\boldsymbol{K}_1\|_{\max} &= \left\| \frac{1}{N_{b-1}} \sum_{l=1}^{b-1} n_l \{\widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}^l_{h_l}; \mathcal{D}_l) - \widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l)\} \right\|_{\max} \\
&\leq \frac{1}{N_{b-1}} \sum_{l=1}^{b-1} n_l \left\| \widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}^l_{h_l}; \mathcal{D}_l) - \widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) \right\|_{\max} \\
&\leq \frac{C}{N_{b-1}} \sum_{l=1}^{b-1} n_l \left\{ \sqrt{\frac{\log^2(p)}{n_l h_l^3}} + \sqrt{\frac{\log^3(p)}{n_l^2 h_l^4}} \right\} s_0 \sqrt{\frac{\log(p)}{N_l}} \\
&\quad + \frac{C}{N_{b-1}} \sum_{l=1}^{b-1} n_l s_0 \sqrt{\frac{\log(p)}{N_l}} + \frac{C}{N_{b-1}} \sum_{l=1}^{b-1} n_l \frac{s_0^2 \log(p)}{N_l h_l^3} \\
&:= K_{11} + K_{12} + K_{13}. \qquad \text{(S1.32)}
\end{aligned}$$

For K_{11}, by lemma 1 we have:

$$\begin{aligned}
K_{11} &= \frac{C}{N_{b-1}} \sum_{l=1}^{b-1} n_l \left\{ \sqrt{\frac{\log(p)}{n_l h_l^3}} + \sqrt{\frac{\log^2(p)}{n_l^2 h_l^4}} \right\} s_0 \sqrt{\frac{\log(p)}{N_l}} \\
&\leq \frac{s_0 \log(p)}{h_{\min}^{3/2}} \frac{C}{N_{b-1}} \sum_{l=1}^{b-1} \sqrt{\frac{n_l}{N_l}} + \frac{s_0 (b-1) \log^{3/2}(p)}{N_{b-1} h_{\min}^2 \sqrt{n_1}} \\
&\leq C' s_0 \left(\frac{\log(p) \left[(b-1) \{1 + \log(N_{b-1}/n_1)\} \right]^{1/2}}{N_{b-1} h_{\min}^{3/2}} + \frac{(b-1) \log^{3/2}(p)}{N_{b-1} h_{\min}^2 \sqrt{n_1}} \right). \qquad \text{(S1.33)}
\end{aligned}$$

Likewise, for K_{12} and K_{13}, by lemma 1, we have

$$K_{12} = \frac{C}{N_{b-1}} \sum_{l=1}^{b-1} \frac{n_l}{\sqrt{N_l}} s_0 \log(p) \leq C' s_0 \sqrt{\frac{\log(p)}{N_{b-1}}}, \qquad \text{(S1.34)}$$

$$K_{13} = \frac{C}{N_{b-1}} \sum_{l=1}^{b-1} n_l \frac{s_0^2 \log(p)}{N_l h_l^3} \leq C'' \frac{s_0^2 \log(p) \{1 + \log(N_{b-1}/n_1)\}}{h_{\min}^3 N_{b-1}}, \qquad \text{(S1.35)}$$

where $h_{\min} = \min\{h_1, \ldots, h_{b-1}\}$.

Furthermore, similar to the proof of (S1.21) and (S1.22), we have:

$$\|\boldsymbol{K}_2\|_{\max} \leq C\sqrt{\frac{\log(p)}{h_{\min}^2 N_{b-1}}}, \tag{S1.36}$$

and

$$\|\boldsymbol{K}_3\|_{\max} \leq \frac{C}{N_{b-1}} \sum_{l=1}^{b-1} n_l h_l \leq C h_{\max}, \tag{S1.37}$$

where $h_{\max} = \max\{h_1, \ldots, h_{b-1}\}$. Combining (S1.31), (S1.32) and (S1.36) with the fact that $\lambda_b\sqrt{N_b} \asymp \lambda_{b-1}\sqrt{N_{b-1}}$, by lemma 4, using Hölder's and Cauchy-Schwartz inequalities, we have:

$$\begin{aligned}
|D_4| \leq & \frac{N_{b-1}}{N_b}\|\widehat{\boldsymbol{\xi}}_{b-1}\|_1 \left\{\|\boldsymbol{K}_1\|_{\max} + \|\boldsymbol{K}_2\|_{\max} + \|\boldsymbol{K}_3\|_{\max}\right\} \|\widehat{\boldsymbol{\xi}}_b\|_1 + \frac{N_{b-1}}{N_b}\|\mathbf{H}(\boldsymbol{\beta}^0)\widehat{\boldsymbol{\xi}}_{b-1}\|_2 \|\widehat{\boldsymbol{\xi}}_b\|_2 \\
\leq & C c_{b-1}^{(1)} s_0^2 \left(\frac{\log(p)\big[(b-1)\{1+\log(N_{b-1}/n_1)\}\big]^{1/2}}{N_{b-1} h_{\min}^{3/2}} + \frac{(b-1)\log^{3/2}(p)}{N_{b-1} h_{\min}^2 \sqrt{n_1}} \right) \lambda_b \|\widehat{\boldsymbol{\xi}}_b\|_1 \\
& + C c_{b-1}^{(1)} s_0^2 \left[\sqrt{\frac{\log(p)}{N_{b-1}}} + \frac{s_0 \log(p)\{1+\log(N_{b-1}/n_1)\}}{h_{\min}^3 N_{b-1}} \right] \lambda_b \|\widehat{\boldsymbol{\xi}}_b\|_1 \\
& + C c_{b-1}^{(1)} s_0 \left\{ \sqrt{\frac{\log(p)}{h_{\min}^2 N_{b-1}}} + h_{\max} \right\} + \big[C' h_b^2 + \tau_{\max}\{\mathbf{H}(\boldsymbol{\beta}^0)\} c_{b-1}^{(2)} \sqrt{s_0}\lambda_b\big] \|\widehat{\boldsymbol{\xi}}_b\|_2.
\end{aligned} \tag{S1.38}$$

The second inequality holds because $\widehat{\boldsymbol{\xi}}_1, \widehat{\boldsymbol{\xi}}_2, \ldots, \widehat{\boldsymbol{\xi}}_{b-1}$ all satisfy the ℓ_1-norm and ℓ_2-norm error bounds. Similar to (S1.23), given $h_l = O\big[\{s_0 \log(p)/N_l\}^{1/4}\big]$ for $l = 1, \ldots, b-1$, under the condition $s_0^9 \log(p)(b-1)^2\{1+\log(N_{b-1}/n_1)\}^2 = o(N_{b-1})$, by Assumption 2, we know that D_4 can be bounded by $\lambda_b\|\widehat{\boldsymbol{\xi}}_b\|_1/3 + M c_{b-1}^{(2)}\sqrt{s_0}\lambda_b\|\widehat{\boldsymbol{\xi}}_b\|_2$. Correspondingly, with (S1.29) and (S1.30), using Triangle and Cauchy-Schwartz in-

equalities we have

$$
\begin{aligned}
&\langle \nabla \mathcal{L}_b(\widehat{\boldsymbol{\beta}}^b_{h_b}) - \nabla \mathcal{L}_b(\boldsymbol{\beta}^0), \widehat{\boldsymbol{\beta}}^b_{h_b} - \boldsymbol{\beta}^0 \rangle \\
\leq & \lambda_b(\|\widehat{\boldsymbol{\xi}}_{b,S_0}\|_1 - \|\widehat{\boldsymbol{\xi}}_{b,S_0^c}\|_1) + \frac{2}{3}\lambda_b \|\widehat{\boldsymbol{\xi}}_1\|_1 + (C' h_b^2 + M c_{b-1}^{(2)} \sqrt{s_0}\lambda_b)\|\widehat{\boldsymbol{\xi}}_b\|_2 \\
= & \lambda_b\Big(\frac{5}{3}\|\widehat{\boldsymbol{\xi}}_{b,S_0}\|_1 - \frac{1}{3}\|\widehat{\boldsymbol{\xi}}_{b,S_0^c}\|_1\Big) + (C' h_b^2 + M c_{b-1}^{(2)} \sqrt{s_0}\lambda_b)\|\widehat{\boldsymbol{\xi}}_b\|_2 \\
\leq & 2\lambda_b \sqrt{s_0}\|\widehat{\boldsymbol{\xi}}_b\|_2 + (C' h_b^2 + M c_{b-1}^{(2)} \sqrt{s_0}\lambda_b)\|\widehat{\boldsymbol{\xi}}_b\|_2 = \big\{\lambda_b \sqrt{s_0}(2 + M c_{b-1}^{(2)}) + C h_b^2\big\}\|\widehat{\boldsymbol{\xi}}_b\|_2.
\end{aligned}
\tag{S1.39}
$$

Under the condition $h_b = O\big[\{s_0 \log(p)/N_b\}^{1/4}\big]$, the bound in (S1.39) above demonstrates the right side of (S1.27).

Then, similar to (S1.25) when $b = 2$, by $\langle \nabla \mathcal{L}_b(\widehat{\boldsymbol{\beta}}^b_{h_b}) - \nabla \mathcal{L}_b(\boldsymbol{\beta}^0), \widehat{\boldsymbol{\beta}}^b_{h_b} - \boldsymbol{\beta}^0 \rangle \geq 0$, we know that $\widehat{\boldsymbol{\xi}}_b$ falls in the cone $\mathbb{C}(l)$ with $l = 3(2 + M c_{b-1}^{(2)})$. Through the same argument for deriving (S1.26), we have

$$
\|\widehat{\boldsymbol{\xi}}_b\|_2 \leq 2M'^{-1}(2 + M c_{b-1}^{(2)})\lambda_b \sqrt{s_0}. \tag{S1.40}
$$

Meanwhile, the condition (S1.28) is confirmed with $C = 3(2 + M c_{b-1}^{(2)})$. As a result, we obtain

$$
\|\widehat{\boldsymbol{\beta}}^b_{h_b} - \boldsymbol{\beta}^0\|_1 \leq c_b^{(1)} s_0 \lambda_b, \qquad \|\widehat{\boldsymbol{\beta}}^b_{h_b} - \boldsymbol{\beta}^0\|_2 \leq c_b^{(2)} \sqrt{s_0}\lambda_b.
$$

Here $c_b^{(2)} = 2M'^{-1}(2 + M c_{b-1}^{(2)})$ and $c_b^{(1)} = 6M'^{-1}\big(2 + M c_{b-1}^{(2)}\big)^2$, M' and M are constants. Note that $c_b^{(2)}$ and $c_b^{(1)}$ may grow exponentially with b. Under the condition $b = o\big\{\log(N_b)\big\}$, the two constants is well-controlled. Then we have completed the proof of Theorem 2. □

Proof of Lemma 3. By Triangle's inequality, we have

$$\left\|\widehat{\mathbf{Q}}_b/N_b - \mathbf{H}(\boldsymbol{\beta}^0)\right\|_{\max} \leq \|\widetilde{\mathbf{H}}_b(\boldsymbol{\beta}^0) - \mathbf{H}(\boldsymbol{\beta}^0)\|_{\max} + \left\|\frac{1}{N_b}\sum_{l=1}^{b}\widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0;\mathcal{D}_l) - \widetilde{\mathbf{H}}_b(\boldsymbol{\beta}^0)\right\|_{\max} + \left\|\widehat{\mathbf{Q}}_b/N_b - \frac{1}{N_b}\sum_{l=1}^{b}\widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0;\mathcal{D}_l)\right\|_{\max}.$$

Then, by (S2.46) and (S2.47) in Lemma 3, we know that $\|\widehat{\mathbf{Q}}_b/N_b - \mathbf{H}(\boldsymbol{\beta}^0)\|_{\max} \leq C_1\omega_{N_b}$ with probability $1-p^{-C_1'}$. Further, due to the inequality $\|\mathbf{AB}\|_{\max} \leq \|\mathbf{A}\|_{\max}\|\mathbf{B}\|_{L_1}$ for matrices $\mathbf{A}, \mathbf{B}$, we have

$$\|\boldsymbol{\Theta}_0\widehat{\mathbf{Q}}_b/N_b - \mathbf{I}\|_{\max} \leq \|\widehat{\mathbf{Q}}_b/N_b - \mathbf{H}(\boldsymbol{\beta}^0)\|_{\max}\|\boldsymbol{\Theta}_0\|_{L_1} \leq C_0C_1\omega_{N_b}. \tag{S1.41}$$

The conclusion in (S1.41) implies that the optimization problem (2.17) has a non-empty set of feasible solutions. Hence there must exist a solution $\widehat{\boldsymbol{\Theta}}_b$ satisfying the condition $\|\widehat{\boldsymbol{\Theta}}_b\widehat{\mathbf{Q}}_b/N_b - \mathbf{I}\|_{\max} \leq C_0C_2\omega_{N_b}$. Moreover, according to lemma 1 in Cai et al. (2011), solving the optimization problem (2.17) is equivalent to solving the p column-wise optimization problems as follows:

$$\widehat{\boldsymbol{\theta}}_j = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p}\|\boldsymbol{\theta}\|_1 \quad \text{s.t.} \quad \|\widehat{\mathbf{Q}}_b/N_b\boldsymbol{\theta}_j - \boldsymbol{e}_j\|_\infty \leq C\omega_{N_b}, \quad j = 1, \ldots, p.$$

Note that (S1.41) implies that $\boldsymbol{\Theta}_0$ lies in the set of feasible solutions of the optimization problem (2.17). Therefore, $\boldsymbol{\Theta}_0$ also satisfies the p optimization problems above. Thus, we can know that $\|\widehat{\boldsymbol{\Theta}}_{b,j}\|_1 \leq \|\boldsymbol{\Theta}_{0,j}\|_1$ for $j = 1, \ldots, p$, which further indicates that $\|\widehat{\boldsymbol{\Theta}}_b\|_{L_1} \leq \|\boldsymbol{\Theta}_0\|_{L_1} \leq C_0$. Eventually, we have completed the proof of

this lemma. □

Proof of Lemma 4. Denote $\widehat{\boldsymbol{\Theta}}_{b,j} = (\widehat{\Theta}_{b,j1}, \ldots, \widehat{\Theta}_{b,jp})^\top$ and $\boldsymbol{\Theta}_{0,j} = (\Theta_{0,j1}, \ldots, \Theta_{0,jp})^\top$. For any $j = 1, \ldots, p$, we decompose $\widehat{\boldsymbol{\Theta}}_{b,j} - \boldsymbol{\Theta}_{0,j} = \widehat{\boldsymbol{\Theta}}^u_{b,j} + \widehat{\boldsymbol{\Theta}}^l_{b,j}$ with

$$\widehat{\boldsymbol{\Theta}}^u_{b,j} = \big(\widehat{\Theta}_{b,j1} I\{|\widehat{\Theta}_{b,j1}| \geq 2\omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1}\}, \ldots, \widehat{\Theta}_{b,jp} I\{|\widehat{\Theta}_{b,jp}| \geq 2\omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1}\}\big)^\top.$$

We then have

$$\|\widehat{\boldsymbol{\Theta}}_{b,j} - \boldsymbol{\Theta}_{0,j}\|_1 \leq \|\widehat{\boldsymbol{\Theta}}^u_{b,j}\|_1 + \|\widehat{\boldsymbol{\Theta}}^l_{b,j}\|_1. \tag{S1.42}$$

By the triangle's inequality, we have

$$\|\boldsymbol{\Theta}_{0,j}\|_1 - \|\widehat{\boldsymbol{\Theta}}^u_{b,j}\|_1 + \|\widehat{\boldsymbol{\Theta}}^l_{b,j}\|_1 \leq \|\boldsymbol{\Theta}_{0,j} + \widehat{\boldsymbol{\Theta}}^u_{b,j}\|_1 + \|\widehat{\boldsymbol{\Theta}}^l_{b,j}\|_1 = \|\widehat{\boldsymbol{\Theta}}_{b,j}\|_1 \leq \|\boldsymbol{\Theta}_{0,j}\|_1,$$

where the last inequality holds by Lemma 3. Therefore, we have $\|\widehat{\boldsymbol{\Theta}}^l_{b,j}\|_1 \leq \|\widehat{\boldsymbol{\Theta}}^u_{b,j}\|_1$, it then follows by (S1.42) that

$$\|\widehat{\boldsymbol{\Theta}}_{b,j} - \boldsymbol{\Theta}_{0,j}\|_1 \leq 2\|\widehat{\boldsymbol{\Theta}}^u_{b,j}\|_1. \tag{S1.43}$$

Recall that $\|\widehat{\mathbf{Q}}_b/N_b - \mathbf{H}(\boldsymbol{\beta}^0)\|_{\max} \leq C_1\omega_{N_b}$ with probability approaching 1. According to (13) in Theorem 6 of Cai et al. (2011), we know that

$$\|\widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0\|_{\max} \leq 4\omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1}.$$

Defined the event $\chi_{b,jk} := \{|\widehat{\Theta}_{b,jk} - \Theta_{0,jk}| \leq \omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1}\}$ for $k = 1, \ldots, p$. Further

note that

$$
\begin{aligned}
\|\widehat{\boldsymbol{\Theta}}^u_{b,j}\|_1 \leq & \sum_{k=1}^{p} |\widehat{\Theta}_{b,jk} - \Theta_{0,jk}| I\{|\widehat{\Theta}_{b,jk}| \geq 2\omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1}\} + \sum_{k=1}^{p} |\Theta_{0,jk}| I\{|\widehat{\Theta}_{b,jk}| < 2\omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1}\} \\
\leq & \sum_{k=1}^{p} \|\widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0\|_{\max} I\{|\widehat{\Theta}_{b,jk}| \geq 2\omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1},\ \chi_{b,jk}\} + \sum_{k=1}^{p} |\Theta_{0,jk}| I\{|\widehat{\Theta}_{b,jk}| < 2\omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1},\ \chi_{b,jk}\} \\
\leq & \sum_{k=1}^{p} \|\widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0\|_{\max} I\{|\Theta_{0,jk}| \geq 2\omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1}\} + \sum_{k=1}^{p} |\Theta_{0,jk}| I\{|\Theta_{0,jk}| < 3\omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1}\} \\
\leq & \|\widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0\|_{\max} \times \sum_{k=1}^{p} I\big\{|\Theta_{0,jk}|^q \geq (2\omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1})^q\big\} \\
& + \sum_{k=1}^{p} |\Theta_{0,jk}|^q \times \max_{1\leq k\leq p} |\Theta_{0,jk}|^{1-q} I\{|\Theta_{0,jk}| < 3\omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1}\} \\
\leq & 7C_r(\omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1})^{1-q}
\end{aligned}
$$

for $0 \leq q < 1$. Combining this result with (S1.43) yields that

$$
\|\widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0\|_{L_1} = \max_{1\leq j\leq p} \|\widehat{\boldsymbol{\Theta}}_{b,j} - \boldsymbol{\Theta}_{0,j}\|_1 \leq 14C_r(\omega_{N_b}\|\boldsymbol{\Theta}_0\|_{L_1})^{1-q}.
$$

This completes the proof. □

Proof of Theorem 3. Recall the formulation of the debiased lasso estimator in

(2.15), by Taylor's expansion, we have

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{h_b}^{b,de} - \boldsymbol{\beta}^0 =& \widehat{\boldsymbol{\beta}}_{h_b}^{b} - \boldsymbol{\beta}^0 - \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Theta}}_b n_l \widehat{\boldsymbol{S}}_{h_l}(\widehat{\boldsymbol{\beta}}_{h_l}^{l}; \mathcal{D}_l) + \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Theta}}_b n_l \widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}_{h_l}^{l}; \mathcal{D}_l)(\widehat{\boldsymbol{\beta}}_{h_l}^{l} - \widehat{\boldsymbol{\beta}}_{h_b}^{b}) \\
=& \widehat{\boldsymbol{\beta}}_{h_b}^{b} - \boldsymbol{\beta}^0 - \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Theta}}_b n_l \widehat{\boldsymbol{S}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) + \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Theta}}_b n_l \left\{ \widehat{\boldsymbol{S}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \widehat{\boldsymbol{S}}_{h_l}(\widehat{\boldsymbol{\beta}}_{h_l}^{l}; \mathcal{D}_l) \right\} \\
&+ \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Theta}}_b n_l \widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}_{h_l}^{l}; \mathcal{D}_l)\{(\widehat{\boldsymbol{\beta}}_{h_l}^{l} - \boldsymbol{\beta}^0) - (\widehat{\boldsymbol{\beta}}_{h_b}^{b} - \boldsymbol{\beta}^0)\} \\
=& -\frac{1}{N_b}\sum_{l=1}^{b} \boldsymbol{\Theta}_0 n_l \{\widehat{\boldsymbol{S}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \boldsymbol{S}_{h_l}(\boldsymbol{\beta}^0)\} \\
&- \frac{1}{N_b}\sum_{l=1}^{b} \big(\widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0\big) n_l \{\widehat{\boldsymbol{S}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \boldsymbol{S}_{h_l}(\boldsymbol{\beta}^0)\} - \frac{1}{N_b}\sum_{l=1}^{b} \big(\widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0\big) n_l \boldsymbol{S}_{h_l}(\boldsymbol{\beta}^0) \\
&- \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Theta}}_b n_l \widehat{\mathbf{H}}_{h_l}(\widetilde{\boldsymbol{\beta}}_l; \mathcal{D}_l)(\widehat{\boldsymbol{\beta}}_{h_l}^{l} - \boldsymbol{\beta}^0) + \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Theta}}_b n_l \widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}_{h_l}^{l}; \mathcal{D}_l)(\widehat{\boldsymbol{\beta}}_{h_l}^{l} - \boldsymbol{\beta}^0) \\
&- \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Theta}}_b n_l \widehat{\mathbf{H}}_{h}(\widehat{\boldsymbol{\beta}}_{h_l}^{l}; \mathcal{D}_l)(\widehat{\boldsymbol{\beta}}_{h_b}^{b} - \boldsymbol{\beta}^0) + (\widehat{\boldsymbol{\beta}}_{h_b}^{b} - \boldsymbol{\beta}^0) \\
=& -\frac{1}{N_b}\sum_{l=1}^{b} \boldsymbol{\Theta}_0 n_l \{\widehat{\boldsymbol{S}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \boldsymbol{S}_{h_l}(\boldsymbol{\beta}^0)\} + \boldsymbol{L}_1 + \boldsymbol{L}_2 + \boldsymbol{L}_3 + \boldsymbol{L}_4,
\end{aligned}$$

where $\widetilde{\boldsymbol{\beta}}_l$ is the intermediate vector between $\widehat{\boldsymbol{\beta}}_{h_l}^{l}$ and $\boldsymbol{\beta}^0$, and

$$\begin{aligned}
\boldsymbol{L}_1 &= -\frac{1}{N_b}\sum_{l=1}^{b} \big(\widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0\big) n_l \{\widehat{\boldsymbol{S}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \boldsymbol{S}_{h_l}(\boldsymbol{\beta}^0)\}, \\
\boldsymbol{L}_2 &= -\frac{1}{N_b}\sum_{l=1}^{b} \big(\widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0\big) n_l \boldsymbol{S}_{h_l}(\boldsymbol{\beta}^0), \\
\boldsymbol{L}_3 &= \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Theta}}_b n_l \left\{ \widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_{h_l}^{l}; \mathcal{D}_l) - \widehat{\mathbf{H}}_h(\widetilde{\boldsymbol{\beta}}_l; \mathcal{D}_l) \right\} (\widehat{\boldsymbol{\beta}}_{h_l}^{l} - \boldsymbol{\beta}^0), \\
\boldsymbol{L}_4 &= \left\{ \mathbf{I} - \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Theta}}_b n_l \widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_{h_l}^{l}; \mathcal{D}_l) \right\} (\widehat{\boldsymbol{\beta}}_{h_b}^{b} - \boldsymbol{\beta}^0).
\end{aligned}$$

It then remains to prove that $\|\boldsymbol{L}_m\|_\infty = o_p\big(N_b^{-1/2}\big)$, $m = 1, 2, 3, 4$. By Lemma 2 and Lemma 4, using Hölder's inequality, we have

$$\begin{aligned}
\|\boldsymbol{L}_1\|_\infty &\le \left\| \frac{1}{N_b} \sum_{l=1}^{b} n_l \{ \widehat{\boldsymbol{S}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \boldsymbol{S}_{h_l}(\boldsymbol{\beta}^0) \} \right\|_\infty \left\| \widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0 \right\|_{L_1} \\
&\le \frac{1}{N_b} \sum_{l=1}^{b} n_l \left\| \widehat{\boldsymbol{S}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \boldsymbol{S}_{h_l}(\boldsymbol{\beta}^0) \right\|_\infty \left\| \widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0 \right\|_{L_1} \\
&\le C C_r C_0^{1-q} \omega_{N_b}^{1-q} \sqrt{\log(p)} \frac{1}{N_b} \sum_{l=1}^{b} \frac{n_l}{\sqrt{N_l}} \lesssim C_r C_0^{1-q} \omega_{N_b}^{1-q} \sqrt{\frac{\log(p)}{N_b}}.
\end{aligned}$$

For $\boldsymbol{L}_2$, by Lemma 1 and Lemma 1, using Hölder's inequality, we have

$$\begin{aligned}
\|\boldsymbol{L}_2\|_\infty \le & \frac{1}{N_b} \sum_{l=1}^{b} n_l \left\| \boldsymbol{S}_{h_l}(\boldsymbol{\beta}^0) \right\|_\infty \left\| \widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0 \right\|_{L_1} \le \frac{1}{N_b} \sum_{l=1}^{b} n_l \left\| \boldsymbol{S}_{h_l}(\boldsymbol{\beta}^0) \right\|_2 \left\| \widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0 \right\|_{L_1} \\
\le & C C_r C_0^{1-q} \omega_{N_b}^{1-q} \frac{1}{N_b} \sum_{l=1}^{b} n_l h_l^2 \asymp C_r C_0^{1-q} \omega_{N_b}^{1-q} \frac{1}{N_b} \sum_{l=1}^{b} n_l \sqrt{\frac{s_0 \log(p)}{N_l}} \\
\lesssim & C_r C_0^{1-q} s_0^{1/2} \omega_{N_b}^{1-q} \sqrt{\frac{\log(p)}{N_b}}.
\end{aligned}$$

For $\boldsymbol{L}_3$, by Lemma 2, 3, 1, 3, 4 and Theorem 2, applying Hölder's inequality and

the Triangle's inequality, we have

$$\begin{aligned}
\|\boldsymbol{L}_3\|_\infty &\le \frac{1}{N_b}\sum_{l=1}^{b} n_l \big\|\widehat{\boldsymbol{\Theta}}_b\big\|_{L_1}\big\|\widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}^l_{h_l};\mathcal{D}_l)-\widehat{\mathbf{H}}_{h_l}(\widetilde{\boldsymbol{\beta}}_l;\mathcal{D}_l)\big\|_{\max}\big\|\widehat{\boldsymbol{\beta}}^l_{h_l}-\boldsymbol{\beta}^0\big\|_1\\
&\le \frac{1}{N_b}\sum_{l=1}^{b} n_l \big\|\widehat{\boldsymbol{\Theta}}_b\big\|_{L_1}\big\|\widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}^l_{h_l};\mathcal{D}_l)-\mathbf{H}(\boldsymbol{\beta}^0)-\widehat{\mathbf{H}}_{h_l}(\widetilde{\boldsymbol{\beta}}_l;\mathcal{D}_l)+\mathbf{H}(\boldsymbol{\beta}^0)\big\|_{\max}\big\|\widehat{\boldsymbol{\beta}}^l_{h_l}-\boldsymbol{\beta}^0\big\|_1\\
&\le \frac{2}{N_b}\sum_{l=1}^{b} n_l C_0\big\|\widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}^l_{h_l};\mathcal{D}_l)-\mathbf{H}(\boldsymbol{\beta}^0)\big\|_{\max}\big\|\widehat{\boldsymbol{\beta}}^l_{h_l}-\boldsymbol{\beta}^0\big\|_1\\
&\le C_0\omega_{N_b}s_0\sqrt{\log(p)}\frac{2}{N_b}\sum_{l=1}^{b}\frac{n_l}{\sqrt{N_l}} \lesssim s_0\omega_{N_b}\sqrt{\frac{\log(p)}{N_b}}.
\end{aligned}$$

Finally, for $\boldsymbol{L}_4$, by Theorem 2 and Lemma 3, using Hölder's inequality, we have

$$\begin{aligned}
\|\boldsymbol{L}_4\|_\infty &\le \big\|\widehat{\boldsymbol{\Theta}}_b\widehat{\mathbf{Q}}_b/N_b-\mathbf{I}\big\|_{\max}\big\|\widehat{\boldsymbol{\beta}}^b_{h_b}-\boldsymbol{\beta}^0\big\|_1\\
&\le c_b^{(1)}s_0\lambda_b C_2C_0\omega_{N_b} \lesssim s_0\omega_{N_b}\sqrt{\frac{\log(p)}{N_b}}.
\end{aligned}$$

According to the condition $C_r^4C_0^{4-4q}s_0^{13-9q}\log^{3-q}(p)[b\{1+\log(N_b/n_1)\}]^{1-q}=o(N_b^{1-q})$, we conclude that $\|\boldsymbol{L}_m\|_\infty=o\big(N_b^{-1/2}\big)$, $m=1,2,3,4$. Then we complete the proof. □

Proof of Proposition 1. By (S2.46) in Lemma 3, we have

$$\begin{aligned}
&\Big\|\widehat{\mathbf{H}}_{h_b}(\widehat{\boldsymbol{\beta}}^*_{h_b};\mathcal{D}_l)-\widehat{\mathbf{H}}_{h_b}(\boldsymbol{\beta}^0;\mathcal{D}_l)\Big\|_{\max}\\
\le& C_1\left\{\sqrt{\frac{\log(p)}{n_lh_b^3}}+\sqrt{\frac{\log^2(p)}{n_l^2h_b^4}}+\frac{1}{\beta_d^0}\right\}\|\widehat{\boldsymbol{\beta}}^*_{h_b}-\boldsymbol{\beta}^0\|_1+\frac{C_1'}{h_b^3}\|\widehat{\boldsymbol{\beta}}^*_{h_b}-\boldsymbol{\beta}^0\|_1^2.
\end{aligned}$$

By Theorem 1, the offline estimator $\widehat{\boldsymbol{\beta}}^*_{h_b}$ satisfies $\|\widehat{\boldsymbol{\beta}}^*_{h_b}-\boldsymbol{\beta}^0\|_1=O\big\{s_0\sqrt{\log(p)/N_b}\big\}$.

Moreover, we have

$$\begin{aligned}\big\|\widehat{\Omega}_b/N_b - \mathbf{H}(\boldsymbol{\beta}^0)\big\|_{\max} \leq & \big\|\mathbf{H}_{h_b}(\boldsymbol{\beta}^0) - \mathbf{H}(\boldsymbol{\beta}^0)\big\|_{\max} + \left\|\frac{1}{N_b}\sum_{l=1}^{b} n_l \widehat{\mathbf{H}}_{h_b}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \mathbf{H}_{h_b}(\boldsymbol{\beta}^0)\right\|_{\max} \\ & + \left\|\frac{1}{N_b}\sum_{l=1}^{b} n_l \Big\{\widehat{\mathbf{H}}_{h_b}(\widehat{\boldsymbol{\beta}}^*_{h_b}; \mathcal{D}_l) - \widehat{\mathbf{H}}_{h_b}(\boldsymbol{\beta}^0; \mathcal{D}_l)\Big\}\right\|_{\max}.\end{aligned}$$

Similar with the proof of Lemma 3, we have:

$$\begin{aligned}&\|\boldsymbol{\Theta}_0 \widehat{\boldsymbol{\Omega}}_b/N_b - \mathbf{I}\|_{\max} \leq C_0 C_1' \zeta_{N_b}, \\ &\|\boldsymbol{\Theta}_0 \widehat{\boldsymbol{\Omega}}_b/N_b - \mathbf{I}\|_{L_1} \leq C_0 C_1' s_1 \zeta_{N_b}, \\ &\|\widehat{\boldsymbol{\Lambda}}_b\|_{L_1} \leq C_0.\end{aligned}$$

where

$$\zeta_{N_b} = s_0 \left\{\sqrt{\frac{b\log^2(p)}{N_b^2 h_b^3}} + \sqrt{\frac{\log^3(p)}{N_b^3 h_b^4}} + \sqrt{\frac{\log(p)}{N_b}}\right\} + \frac{s_0^2 \log(p)}{N_b h_b^3} + \sqrt{\frac{\log(p)}{N_b h_b^2}} + h_b.$$

Then, following the proof of Theorem 3, we have:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{*,de}_{h_b} - \boldsymbol{\beta}^0 =& \widehat{\boldsymbol{\beta}}^{*}_{h_b} - \boldsymbol{\beta}^0 - \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Lambda}}_b n_l \widehat{\boldsymbol{S}}_{h_b}(\widehat{\boldsymbol{\beta}}^{*}_{h_b}; \mathcal{D}_l) \\
=& \widehat{\boldsymbol{\beta}}^{*}_{h_b} - \boldsymbol{\beta}^0 - \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Lambda}}_b n_l \widehat{\boldsymbol{S}}_{h_b}(\boldsymbol{\beta}^0; \mathcal{D}_l) + \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Lambda}}_b n_l \big\{ \widehat{\boldsymbol{S}}_{h_b}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \widehat{\boldsymbol{S}}_{h_b}(\widehat{\boldsymbol{\beta}}^{*}_{h_b}; \mathcal{D}_l) \big\} \\
=& -\frac{1}{N_b}\sum_{l=1}^{b} \boldsymbol{\Theta}_0 n_l \big\{ \widehat{\boldsymbol{S}}_{h_b}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \boldsymbol{S}_{h_b}(\boldsymbol{\beta}^0) \big\} \\
& -\frac{1}{N_b}\sum_{l=1}^{b} \Big(\widehat{\boldsymbol{\Lambda}}_b - \boldsymbol{\Theta}_0 \Big) n_l \big\{ \widehat{\boldsymbol{S}}_{h_b}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \boldsymbol{S}_{h_b}(\boldsymbol{\beta}^0) \big\} - \Big(\widehat{\boldsymbol{\Lambda}}_b - \boldsymbol{\Theta}_0 \Big) \boldsymbol{S}_{h_b}(\boldsymbol{\beta}^0) \\
& -\frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Lambda}}_b n_l \{ \widehat{\mathbf{H}}_h(\widetilde{\boldsymbol{\beta}}^{*}_{b}; \mathcal{D}_l) - \widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}^{*}_{h_b}; \mathcal{D}_l) \} (\widehat{\boldsymbol{\beta}}^{*}_{h_b} - \boldsymbol{\beta}^0) \\
& -\frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Lambda}}_b n_l \widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}^{*}_{h_b}; \mathcal{D}_l) (\widehat{\boldsymbol{\beta}}^{*}_{h_b} - \boldsymbol{\beta}^0) + (\widehat{\boldsymbol{\beta}}^{*}_{h_b} - \boldsymbol{\beta}^0) \\
=& -\frac{1}{N_b}\sum_{l=1}^{b} \boldsymbol{\Theta}_0 n_l \widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^0; \mathcal{D}_l) + \boldsymbol{F}_1 + \boldsymbol{F}_2 + \boldsymbol{F}_3 + \boldsymbol{F}_4,
\end{aligned}$$

where $\widetilde{\boldsymbol{\beta}}^{*}_{b}$ is the intermediate value between $\boldsymbol{\beta}^0$ and $\widehat{\boldsymbol{\beta}}^{b}_{h_b}$. $\boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3, \boldsymbol{F}_4$ corresponds to $\boldsymbol{L}_1, \boldsymbol{L}_2, \boldsymbol{L}_3, \boldsymbol{L}_4$ in the proof of Theorem 3. Similarly, we have for $\boldsymbol{F}_1$,

$$\begin{aligned}
\|\boldsymbol{F}_1\|_\infty \leq& \left\| \frac{1}{N_b}\sum_{l=1}^{b} n_l \widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^0; \mathcal{D}_l) \right\|_\infty \left\| \widehat{\boldsymbol{\Lambda}}_b - \boldsymbol{\Theta}_0 \right\|_{L_1} \\
\leq& \frac{1}{N_b}\sum_{l=1}^{b} n_l \left\| \widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^0; \mathcal{D}_l) \right\|_\infty \left\| \widehat{\boldsymbol{\Lambda}}_b - \boldsymbol{\Theta}_0 \right\|_{L_1} \\
\leq& C_1' C_r C_0^{1-q} \zeta_{N_b}^{1-q} \sqrt{\log(p)} \frac{1}{N_b}\sum_{l=1}^{b} \frac{n_l}{\sqrt{N_l}} \lesssim C_r C_0^{1-q} \zeta_{N_b}^{1-q} \sqrt{\frac{\log(p)}{N_b}}.
\end{aligned}$$

For $\boldsymbol{F}_2$,

$$\begin{aligned}\|\boldsymbol{F}_2\|_\infty \leq & \left\|\boldsymbol{S}_{h_b}(\boldsymbol{\beta}^0)\right\|_\infty \left\|\widehat{\boldsymbol{\Lambda}}_b - \boldsymbol{\Theta}_0\right\|_{L_1} \leq \left\|\boldsymbol{S}_{h_b}(\boldsymbol{\beta}^0)\right\|_2 \left\|\widehat{\boldsymbol{\Lambda}}_b - \boldsymbol{\Theta}_0\right\|_{L_1} \\ \leq & CC_rC_0^{1-q}\zeta_{N_b}^{1-q}h_b^2 \lesssim C_rC_0^{1-q}s_0^{1/2}\zeta_{N_b}^{1-q}\sqrt{\frac{\log(p)}{N_b}}.\end{aligned}$$

For $\boldsymbol{F}_3$,

$$\begin{aligned}\|\boldsymbol{F}_3\|_\infty &\leq \frac{1}{N_b}\sum_{l=1}^{b} n_l\|\widehat{\boldsymbol{\Lambda}}_b\|_{L_1}\left\|\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}^*_{h_b};\mathcal{D}_l) - \widehat{\mathbf{H}}_h(\widetilde{\boldsymbol{\beta}}_l;\mathcal{D}_l)\right\|_{\max}\|\widehat{\boldsymbol{\beta}}^*_{h_b} - \boldsymbol{\beta}^0\|_1 \\ &\leq \frac{1}{N_b}\sum_{l=1}^{b} n_l\|\widehat{\boldsymbol{\Lambda}}_b\|_{L_1}\left\|\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}^*_{h_b};\mathcal{D}_l) - \mathbf{H}_h(\boldsymbol{\beta}^0) - \widehat{\mathbf{H}}_h(\widetilde{\boldsymbol{\beta}}^*_b;\mathcal{D}_l) + \mathbf{H}_h(\boldsymbol{\beta}^0)\right\|_{\max}\|\widehat{\boldsymbol{\beta}}^*_{h_b} - \boldsymbol{\beta}^0\|_1 \\ &\leq \frac{2}{N_b}\sum_{l=1}^{b} n_lC_0\left\|\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}^*_{h_b};\mathcal{D}_l) - \mathbf{H}_h(\boldsymbol{\beta}^0)\right\|_{\max}\|\widehat{\boldsymbol{\beta}}^*_{h_b} - \boldsymbol{\beta}^0\|_1 \\ &\leq \frac{2}{N_b}\sum_{l=1}^{b} n_lC_0\zeta_{N_b}s_0\sqrt{\frac{\log(p)}{N_b}} \leq s_0\zeta_{N_b}\sqrt{\frac{\log(p)}{N_b}},\end{aligned}$$

For $\boldsymbol{F}_4$,

$$\begin{aligned}\|\boldsymbol{F}_4\|_\infty &\leq \|\widehat{\boldsymbol{\Lambda}}_b\widehat{\boldsymbol{\Omega}}_b/N_b - \mathbf{I}\|_{\max}\|\widehat{\boldsymbol{\beta}}^*_{h_b} - \boldsymbol{\beta}^0\|_1 \\ &\leq Cs_0\lambda_bC_2'C_0\zeta_{N_b} \leq s_0\zeta_{N_b}\sqrt{\frac{\log(p)}{N_b}}.\end{aligned}$$

Under the condition $C_r^4C_0^{4-4q}s_0^{11-9q}\log^{3-q}(p) = o(N_b^{1-q})$, we have that $\|\boldsymbol{F}_m\|_\infty = o\big(N_b^{-1/2}\big)$. Then the proof is completed. □

S2 Technical Lemmas

In this appendix, we provide some technical lemmas to support our proof.

Notations. For a sub-exponential random variable x, $\|x\|_{\psi_1} = \inf\{t > 0 : \mathbb{E}e^{|x|/t} \leq 2\}$.

Lemma 1. *Let n_l and N_l be the batch size and the cumulative batch size, respectively, when the l-th data batch arrives, $l = 1, \ldots, b$. Then we have*

$$\sum_{l=1}^{b} \frac{n_l}{N_l} \leq 1+\log(N_b/n_1), \qquad \sum_{l=1}^{b} \frac{n_l}{\sqrt{N_l}} \leq 2\sqrt{N_b}, \qquad \sum_{l=1}^{b} \sqrt{\frac{n_l}{N_l}} \leq \left\{b+b\log(N_b/n_1)\right\}^{1/2}.$$

This lemma is widely used in online learning and has been confirmed in the literature, see for example Han et al. (2021) and Luo et al. (2023).

Lemma 2. *Let h be the bandwidth used for the smoothed loss function computed on the l-th data batch. Define $r_{li}(\boldsymbol{\beta}) = 1 - y_{li}\mathbf{x}_{li}^\top\boldsymbol{\beta}$, $l = 1, 2, \ldots, b$. Under Assumption 3, 4 and 5, we have*

$$\left|\mathbb{E}\left[K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}/h_l^2\right]\right| \leq C(\beta_d^0)^{-1}, \tag{S2.44}$$

$$\mathbb{E}\left[\left|K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}/h_l^2\right|^m\right] \leq C'(\beta_d^0)^{-1}h_l^{-(2m-1)}, \quad \textit{for } m \geq 2, \tag{S2.45}$$

for some positive constants C and C'.

Proof. By the similar arguments used in the proof of Lemma 1, by a change of

variables ($u = r_{li}(\boldsymbol{\beta}^0)/h_l$), we have

$$
\begin{aligned}
&\left|\mathbb{E}[K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}/h_l^2 \mid y_{li}=1]\right| \\
=&\left|\int_{\mathbb{R}^{p-1}}\int_{\mathbb{R}} K'(u) f\left(\frac{1-uh_l-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d}\right)\frac{1}{h_l\beta_d^0}\mathrm{d}u\mathrm{d}\mathbf{x}_{-d}\right| \\
=&\left|\frac{1}{h_l\beta_d^0}\int_{\mathbb{R}^{p-1}}\int_{\mathbb{R}} f\left(\frac{1-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d}\right)K'(u)\mathrm{d}u\mathrm{d}\mathbf{x}_{-d}\right| \\
&+\left|\frac{1}{(\beta_d^0)^2}\int_{\mathbb{R}^{p-1}}\int_{\mathbb{R}} K'(u)uf'\left(\frac{1-*-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d}\right)\mathrm{d}u\mathrm{d}\mathbf{x}_{-d}\right| \\
\le&\left|\int_{\mathbb{R}} K'(u)u\mathrm{d}u\right|\left|\int_{\mathbb{R}^{p-1}} f'\left(\frac{1-*-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d}\right)\mathrm{d}\mathbf{x}_{-d}\right| \le \frac{C}{(\beta_d^0)^2},
\end{aligned}
$$

where $*$ is a value between 0 and uh_l. The first inequality is derived by the fact $\int K'(u)\mathrm{d}u = 0$ and the last inequality holds due to Assumption 4. An analogous statement holds when conditioned on ($y_{li} = -1$), thus (S2.44) is verified. Next, we focus on (S2.45). Likewise, by Assumption 4 and Assumption 5, we can get

$$
\begin{aligned}
&\mathbb{E}\Big(\left[K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}/h_l^2\right]^2\big|y_{li}=1\Big) \\
=&\int_{\mathbb{R}^{p-1}}\int_{\mathbb{R}}\{K'(u)\}^2 f\left(\frac{1-uh_l-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d}\right)\frac{1}{h_l\beta_d^0}\mathrm{d}u\mathrm{d}\mathbf{x}_{-d} \\
=&\int_{\mathbb{R}}\frac{1}{h_l\beta_d^0}\{K'(u)\}^2 f\left(\frac{1-uh_l-\mathbf{x}_{-d}^\top\boldsymbol{\beta}_{-d}^0}{\beta_d^0}\Big|\mathbf{x}_{-d}\right)\mathrm{d}u \le \frac{C}{h_l\beta_d^0}.
\end{aligned}
$$

An analogous result can be confirmed given ($y_{li} = -1$). Moreover, note that $|K'\{r_{li}(\boldsymbol{\beta}^0)/h\}/h^2| \le Ch^{-2}$ by Assumption 5, it then follows that for $m \ge 3$,

$$
\begin{aligned}
&\mathbb{E}\left[|K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}/h_l^2|^m\right] \\
\le&\mathbb{E}\Big(\left[K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}/h_l^2\right]^2\cdot|K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}/h^2|^{m-2}\Big) \le C'(\beta_d^0)^{-1}h_l^{-(2m-1)}.
\end{aligned}
$$

As a result, (S2.45) is confirmed. This finishes the proof. □

Lemma 3. *For $l = 1, \ldots, b-1$, let $\widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}; \mathcal{D}_l)$ be the same argument as (2.7) and $\widetilde{\mathbf{H}}_{b-1} = \frac{1}{N_{b-1}} \sum_{l=1}^{b-1} n_l \mathbf{H}_{h_l}(\boldsymbol{\beta}^0)$. Under Assumption 1 and 5, we have*

$$\begin{aligned}
&\left\| \widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}^l_{h_l}; \mathcal{D}_l) - \widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) \right\|_{\max} \\
\leq & C_1 \left\{ \sqrt{\frac{\log(p)}{n_l h_l^3}} + \sqrt{\frac{\log^2(p)}{n_l^2 h_l^4}} + \frac{1}{\beta_d^0} \right\} \|\widehat{\boldsymbol{\beta}}^l_{h_l} - \boldsymbol{\beta}^0\|_1 + \frac{C_1'}{h_l^3} \|\widehat{\boldsymbol{\beta}}^l_{h_l} - \boldsymbol{\beta}^0\|_1^2, \qquad \text{(S2.46)} \\
&\left\| \frac{1}{N_{l-1}} \sum_{l=1}^{b-1} n_l \widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \widetilde{\mathbf{H}}_{b-1}(\boldsymbol{\beta}^0) \right\|_{\max} \leq C_2 \sqrt{\frac{\log(p)}{h_{\min}^2 N_{b-1}}}, \qquad \text{(S2.47)} \\
&\left\| \mathbf{H}_h(\boldsymbol{\beta}^0) - \mathbf{H}(\boldsymbol{\beta}^0) \right\|_{\max} \leq C_3 h, \qquad \text{(S2.48)}
\end{aligned}$$

hold with probability $1 - p^{-C'}$ for some constant $C' > 0$.

Proof. We first derive the bound (S2.46). Recall that $r_{li}(\boldsymbol{\beta}) = (1 - y_{li} \mathbf{x}_{li}^\top \boldsymbol{\beta})$. By

the Taylor's expansion, Triangle inequality and Hölder's inequality,

$$
\begin{aligned}
&\left\|\widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}^l_{h_l};\mathcal{D}_l)-\widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0;\mathcal{D}_l)\right\|_{\max}\\
=&\max_{j,k}\left|\frac{1}{n_l h_l}\sum_{i=1}^{n_l h_l}x_{li,j}x_{li,k}\left[K\{r_{li}(\widehat{\boldsymbol{\beta}}^l_{h_l})/h_l\}-K\{r_{li}(\boldsymbol{\beta}^0)/h_l\}\right]\right|\\
\le&\max_{j,k}\left|\frac{1}{n_l h_l^2}\sum_{i=1}^{n_l}x_{li,j}x_{li,k}K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}\mathbf{x}_{li}^\top(\widehat{\boldsymbol{\beta}}^l_{h_l}-\boldsymbol{\beta}^0)\right|\\
&+\max_{j,k}\left|\frac{1}{n_l h_l^3}\sum_{i=1}^{n_l}x_{li,j}x_{li,k}K''\{r_{li}(\widetilde{\boldsymbol{\beta}}_l)/h_l\}\left\{\mathbf{x}_{li}^\top(\widehat{\boldsymbol{\beta}}^l_{h_l}-\boldsymbol{\beta}^0)\right\}^2\right|\\
\le&\max_{j,k,l}\left|\frac{1}{n_l h_l^2}\sum_{i=1}^{n_l}\left(x_{li,j}x_{li,k}x_{li,l}K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}-\mathbb{E}[x_{li,j}x_{li,k}x_{li,l}K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}]\right)\right|\left\|\widehat{\boldsymbol{\beta}}^l_{h_l}-\boldsymbol{\beta}^0\right\|_1\\
&+\max_{j,k,l}\left|\frac{1}{h_l^2}\mathbb{E}[x_{li,j}x_{li,k}x_{li,l}K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}]\right|\left\|\widehat{\boldsymbol{\beta}}^l_{h_l}-\boldsymbol{\beta}^0\right\|_1\\
&+\max_{j,k,l,m}\left|\frac{1}{n_l h_l^3}\sum_{i=1}^{n_l}x_{li,j}x_{li,k}x_{li,l}x_{li,m}K''\{r_{li}(\widetilde{\boldsymbol{\beta}}_l)/h_l\}\right|\left\|\widehat{\boldsymbol{\beta}}^l_{h_l}-\boldsymbol{\beta}^0\right\|_1^2:=G_1+G_2+G_3,
\end{aligned}
$$

where $\mathbf{x}_{li}^\top\widetilde{\boldsymbol{\beta}}_l$ is an intermediate point between $\mathbf{x}_{li}^\top\widehat{\boldsymbol{\beta}}^l_{h_l}$ and $\mathbf{x}_{li}^\top\boldsymbol{\beta}^0$. By Assumption 1 we know that $x_{li,j}x_{li,k}x_{li,l}$ is bounded. Thus by (S2.45) in Lemma 2 we obtain that

$$
\mathbb{E}\big[\big|x_{li,j}x_{li,k}x_{li,l}K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}/h_l^2\big|^m\big]\le\big(C'/h_l^2\big)^m\big(h_l/\beta_d^0\big).
$$

Then, by Bernstein's inequality (lemma 2.2.11 in van der Vaart and Wellner (1996)), we have

$$
\begin{aligned}
&\left|\frac{1}{n_l h_l^2}\sum_{i=1}^{n_l}\left(x_{li,j}x_{li,k}x_{li,l}K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}-\mathbb{E}[x_{li,j}x_{li,k}x_{li,l}K'\{r_{li}(\boldsymbol{\beta}^0)/h_l\}]\right)\right|\\
\le&\frac{C}{\beta_d^0}\left\{\sqrt{\frac{\log(p)}{n_l h_l^3}}+\frac{\log(p)}{n_l h_l^2}\right\}
\end{aligned}
\tag{S2.49}
$$

with probability $1-p^{-C}$ for some constant $C>0$. Consequently, by (S2.49) we have

$$G_1 \leq C\|\widehat{\boldsymbol{\beta}}^l_{h_l} - \boldsymbol{\beta}^0\|_1 \left\{ \sqrt{\frac{\log(p)}{n_l h_l^3}} + \frac{\log(p)}{n_l h_l^2} \right\}. \tag{S2.50}$$

Furthermore, by (S2.44) in lemma 2 and Assumption 1, we know that

$$G_2 \leq \frac{C}{\beta_d^0}\|\widehat{\boldsymbol{\beta}}^l_{h_l} - \boldsymbol{\beta}^0\|_1. \tag{S2.51}$$

Using the fact $K''(u) < \infty$ from Assumption 5 and Assumption 1, we have

$$G_3 \leq \frac{C'}{h_l^3}\|\widehat{\boldsymbol{\beta}}^l_{h_l} - \boldsymbol{\beta}^0\|_1^2. \tag{S2.52}$$

Finally, combining (S2.50), (S2.51) and (S2.52), we have established the bound (S2.46).

To demonstrate (S2.47), note that

$$\begin{aligned}
&\left\| \frac{1}{N_{l-1}} \sum_{l=1}^{b-1} n_l \widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \widetilde{\mathbf{H}}_{b-1}(\boldsymbol{\beta}^0) \right\|_{\max} \\
=&\max_{j,k} \left| \frac{1}{N_{b-1}} \sum_{l=1}^{b-1} \sum_{i=1}^{n_l} \frac{1}{h_l} \left[x_{li,j} x_{li,k} K\left(r_{li}(\boldsymbol{\beta}^0)/h_l\right) - \mathbb{E}\left\{ x_{li,j} x_{li,k} K\left(r_{li}(\boldsymbol{\beta}^0)/h_l\right)\right\} \right] \right|.
\end{aligned}$$

Define $\omega^{(si)}_{0,j,k} = x_{li,j}x_{li,k}K\left\{r_{li}(\boldsymbol{\beta}^0)/h_l\right\}$. By Assumption 1 and 5, we know that $\left|\omega^{(si)}_{0,j,k}\right| \leq \kappa_{\max} u_{\mathbf{x}}^2$. By Hoeffding's inequality and the union bound,

$$P\left[\max_{j,k} \left| \frac{1}{N_{b-1}} \sum_{l=1}^{b-1} \sum_{i=1}^{n_l} \left\{ \omega^{(si)}_{0,j,k} - \mathbb{E}\big(\omega^{(si)}_{0,j,k}\big) \right\} \right| > t \right] \leq 2p^2 \exp\left(- \frac{N_{b-1} t^2}{2\kappa^2_{\max} u^4_{\mathbf{x}}} \right).$$

Therefore, with probability $1-2p^{-6}$, we get

$$\left\| \frac{1}{N_{l-1}} \sum_{l=1}^{b-1} n_l \widehat{\mathbf{H}}_{h}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \widetilde{\mathbf{H}}_{h}(\boldsymbol{\beta}^0) \right\|_{\max} \leq 4\kappa_{\max} u_{\mathbf{x}}^2 \sqrt{\frac{\log(p)}{h^2_{\min} N_{b-1}}},$$

then (S2.47) is demonstrated. To prove (S2.48), we define a function class $\mathcal{Q} = \{q(x) = x_k x_l, 1 \le k, l \le p\}$ and show that for any $q(x) \in \mathcal{Q}$,

$$\left| \mathbb{E}\Big\{ \frac{1}{h} K\Big(\frac{1 - y\mathbf{x}^\top \boldsymbol{\beta}^0}{h} \Big) q(x) \Big\} - \mathbb{E}\{\delta(1 - y\mathbf{x}^\top \boldsymbol{\beta}^0) q(x)\} \right| \le \frac{Ch^2}{(\beta_d^0)^2}.$$

We consider only the conditional expectation given $y = 1$ and the counterpart conditioned on $y = -1$ is similar. Applying the same arguments in the proof of Lemma 1, letting $u = (1 - \mathbf{x}^\top \boldsymbol{\beta}^0)/h$, we have

$$\begin{aligned}
&\mathbb{E}\Big\{ \frac{1}{h} K\Big(\frac{1 - y\mathbf{x}^\top \boldsymbol{\beta}^0}{h} \Big) q(x) \mid y = 1 \Big\} \\
&= \int_{\mathbb{R}^p} \frac{1}{h} K\Big(\frac{1 - \mathbf{x}^\top \boldsymbol{\beta}^0}{h} \Big) q(x) f(x) \mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^{p-1}} \int_{\mathbb{R}} K(u) q\Big(\frac{1 - uh - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d} \Big) f\Big(\frac{1 - uh - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d} \Big) \frac{1}{\beta_d^0} \mathrm{d}u \mathrm{d}\mathbf{x}_{-d} \\
&= \int_{\mathbb{R}^{p-1}} \int_{\mathbb{R}} K(u) (qf)\Big(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d} \Big) \frac{1}{\beta_d^0} \mathrm{d}u \mathrm{d}\mathbf{x}_{-d} \\
&\quad - \int_{\mathbb{R}^{p-1}} \int_{\mathbb{R}} K(u) (qf)'\Big(\frac{1 - * - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d} \Big) \frac{1}{(\beta_d^0)^2} uh \mathrm{d}u \mathrm{d}\mathbf{x}_{-d},
\end{aligned}$$

where $(qf)'$ is the partial derivative of $qf(\cdot)$ with respect to its first argument and $*$ represents a value between 0 and uh. The last equality holds by Taylor's expansion.

Moreover, letting $z_0 = \mathbf{x}^\top \boldsymbol{\beta}^0$, we have

$$\begin{aligned}
&\mathbb{E}\{\delta(1 - y\mathbf{x}^\top \boldsymbol{\beta}^0)q(x) \mid y = 1\} \\
&= \int_{\mathbb{R}^p} \delta(1 - \mathbf{x}^\top \boldsymbol{\beta}^0)q(x)f(x)\mathrm{d}x \\
&= \int_{\mathbb{R}^{p-1}} \int_{\mathbb{R}} \delta(1 - z_0)q\Big(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d}\Big) f\Big(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d}\Big) \frac{1}{\beta_d^0} \mathrm{d}z_0 \mathrm{d}\mathbf{x}_{-d} \\
&= \int_{\mathbb{R}^{p-1}} (qf)\Big(\frac{1 - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d}\Big) \frac{1}{\beta_d^0} \mathrm{d}\mathbf{x}_{-d}.
\end{aligned}$$

By the fact $\int_{\mathbb{R}} K(u)\mathrm{d}u = 1$, we know that

$$\begin{aligned}
&\left|\mathbb{E}\Big\{\frac{1}{h}K\Big(\frac{1 - y\mathbf{x}^\top \boldsymbol{\beta}^0}{h}\Big)q(x) \mid y = 1\Big\} - \mathbb{E}\{\delta(1 - y\mathbf{x}^\top \boldsymbol{\beta}^0)q(x) \mid y = 1\}\right| \\
=& \left|\int_{\mathbb{R}^{p-1}} \int_{\mathbb{R}} K(u)(qf)'\Big(\frac{1 - * - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d}\Big) \frac{1}{(\beta_d^0)^2} uh \mathrm{d}u \mathrm{d}\mathbf{x}_{-d}\right| \\
\leq& \frac{h}{(\beta_d^0)^2} \int_{\mathbb{R}} |u|K(u)\mathrm{d}u \left|\int_{\mathbb{R}^{p-1}} (qf)'\Big(\frac{1 - * - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d}\Big) \mathrm{d}\mathbf{x}_{-d}\right| \\
=& \frac{\kappa_1 h}{(\beta_d^0)^2} \left|\int_{\mathbb{R}^{p-1}} (q'f)\Big(\frac{1 - * - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d}\Big) + (qf')\Big(\frac{1 - * - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0}{\beta_d^0}, \mathbf{x}_{-d}\Big) \mathrm{d}\mathbf{x}_{-d}\right| \\
\leq& \frac{\kappa_1 h}{(\beta_d^0)^2} \big\{2x_{d*}f'(x_{d*}|\mathbf{x}_{-d}) + x_{d*}^2 f(x_{d*}|\mathbf{x}_{-d})\big\} \leq C'\kappa_1 \frac{h}{(\beta_d^0)^2},
\end{aligned}$$

for some constant $C' > 0$. Here the second inequality holds by a change of variables ($x_{d*} = (1 - * - \mathbf{x}_{-d}^\top \boldsymbol{\beta}_{-d}^0)/\beta_d^0$) and last inequality holds by Assumption 4. Thus (S2.48) is demonstrated. The proof is finished.

□

Lemma 4. *For two vectors* $\mathbf{u} = (u_1, \ldots, u_p)^\top \in \mathbb{R}^p$, $\mathbf{v} = (v_1, \ldots, u_p)^\top \in \mathbb{R}^p$ *and a*

matrix $\mathbf{A} = (a_{ij})_{i,j}^{p} \in \mathbb{R}^{p\times p}$, *we have*

$$\|\mathbf{Au}\|_\infty \leq \|\mathbf{A}\|_{\max}\|\mathbf{u}\|_1,$$

$$\mathbf{u}^\top\mathbf{Av} \leq \|\mathbf{A}\|_{\max}\|\mathbf{u}\|_1\|\mathbf{v}\|_1.$$

Proof. Firstly, we have

$$\|\mathbf{Au}\|_\infty = \max_{1\leq i\leq p}\left|\sum_{j=1}^{p} a_{ij}u_j\right| \leq \max_{1\leq i\leq p}\max_{1\leq j\leq p}|a_{ij}|\sum_{j=1}^{p}|u_j| = \|\mathbf{A}\|_{\max}\|\mathbf{u}\|_1.$$

Then, by Hölder's inequality

$$\mathbf{u}^\top\mathbf{Av} \leq \|\mathbf{Au}\|_\infty\|\mathbf{v}\|_1 \leq \|\mathbf{A}\|_{\max}\|\mathbf{u}\|_1\|\mathbf{v}\|_1.$$

□

Lemma 5. *For a p-dimensional vector* $\boldsymbol{\delta}$, *define* $\mathbb{B}(r) = \{\boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}\|_2 \leq r\}$ *and* $\mathbb{C}(l) = \{\boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}\|_1 \leq l\sqrt{s_0}\|\boldsymbol{\delta}\|_2\}$. *Further define the event*

$$\mathcal{A}_b(r,l,\kappa) = \left\{\frac{\langle\nabla\mathcal{L}_b(\boldsymbol{\beta}) - \nabla\mathcal{L}_b(\boldsymbol{\beta}^0), \boldsymbol{\beta}-\boldsymbol{\beta}^0\rangle}{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|_2^2} \geq \kappa,\ \forall\boldsymbol{\beta} \in \boldsymbol{\beta}^0 + \mathbb{B}(r)\cap\mathbb{C}(l)\right\}$$

Suppose that $\|\widehat{\boldsymbol{\beta}}_{h_l}^{l} - \boldsymbol{\beta}^0\|_2^2 = O_P\{s_0\log(p)/N_l\}$ *and* $h_l = O\big[\{s_0\log(p)/N_l\}^{1/4}\big]$ *for* $l = 1,\ldots,b-1$, *we have*

$$P\left\{\mathcal{A}_b(r,l,\kappa)\right\} \to 1.$$

Proof. First, observe that

$$
\begin{aligned}
&\langle \nabla \mathcal{L}_b(\boldsymbol{\beta}) - \nabla \mathcal{L}_b(\boldsymbol{\beta}^0), \boldsymbol{\beta} - \boldsymbol{\beta}^0 \rangle \\
&= \frac{1}{N_b} \langle \widehat{\mathbf{Q}}_{b-1}(\boldsymbol{\beta} - \boldsymbol{\beta}^0), \boldsymbol{\beta} - \boldsymbol{\beta}^0 \rangle + \frac{n_b}{N_b} \langle \widehat{\boldsymbol{S}}_{h_b}(\boldsymbol{\beta}; \mathcal{D}_b) - \widehat{\boldsymbol{S}}_{h_b}(\boldsymbol{\beta}^0; \mathcal{D}_b), \boldsymbol{\beta} - \boldsymbol{\beta}^0 \rangle \\
&= \frac{1}{N_b} (\boldsymbol{\beta} - \boldsymbol{\beta}^0)^\top \sum_{l=1}^{b-1} n_l \widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}^l_{h_l}; \mathcal{D}_l)(\boldsymbol{\beta} - \boldsymbol{\beta}^0) \\
&\quad + \frac{1}{N_b} \sum_{i=1}^{n_b} \Big\{ L'_{h_b}(y_{bi} \mathbf{x}_{bi}^\top \boldsymbol{\beta}) - L'_{h_b}(y_{bi} \mathbf{x}_{bi}^\top \boldsymbol{\beta}^0) \Big\} y_{bi} \langle \mathbf{x}_{bi}, \boldsymbol{\beta} - \boldsymbol{\beta}^0 \rangle \\
&=: B_1 + B_2. \qquad \text{(S2.53)}
\end{aligned}
$$

For the term B_1, note that

$$
\begin{aligned}
\frac{1}{N_b} \sum_{l=1}^{b-1} n_l \widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}^l_{h_l}; \mathcal{D}_l) =& \frac{1}{N_b} \sum_{l=1}^{b-1} n_l \{ \widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}^l_{h_l}; \mathcal{D}_l) - \widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) \} \\
&+ \frac{1}{N_b} \sum_{l=1}^{b-1} n_l \{ \widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \mathbf{H}_{h_l}(\boldsymbol{\beta}^0) \} \\
&+ \frac{1}{N_b} \sum_{l=1}^{b-1} n_l \{ \mathbf{H}_{h_l}(\boldsymbol{\beta}^0) - \mathbf{H}(\boldsymbol{\beta}^0) \} + \mathbf{H}(\boldsymbol{\beta}^0).
\end{aligned}
$$

From (S1.31) in the proof of Theorem 2, recall that

$$
\begin{aligned}
\boldsymbol{K}_1 &= \frac{1}{N_b} \sum_{l=1}^{b-1} n_l \{ \widehat{\mathbf{H}}_{h_l}(\widehat{\boldsymbol{\beta}}^l_{h_l}; \mathcal{D}_l) - \widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) \}; \\
\boldsymbol{K}_2 &= \frac{1}{N_b} \sum_{l=1}^{b-1} n_l \{ \widehat{\mathbf{H}}_{h_l}(\boldsymbol{\beta}^0; \mathcal{D}_l) - \mathbf{H}_{h_l}(\boldsymbol{\beta}^0) \}; \\
\boldsymbol{K}_3 &= \frac{1}{N_b} \sum_{l=1}^{b-1} n_l \{ \mathbf{H}_{h_l}(\boldsymbol{\beta}^0) - \mathbf{H}(\boldsymbol{\beta}^0) \}.
\end{aligned}
$$

By Lemma 3 and the calculation in (S1.32)-(S1.37), under the conditions $\|\widehat{\boldsymbol{\beta}}^l_{h_l} - \boldsymbol{\beta}^0\|_2 = O\big\{\sqrt{s_0 \log(p)/N_l}\big\}$, $h_l = O\big[\{s_0 \log(p)/N_l\}^{1/4}\big]$ for $l = 1, \ldots, b-1$ and

$s_0^9 \log(p)(b-1)^2\{1+\log(N_{b-1}/n_1)\}^2 = o(N_{b-1})$, we get $s_0(\|\boldsymbol{K}_1\|_{\max} + \|\boldsymbol{K}_2\|_{\max} + \|\boldsymbol{K}_3\|_{\max}) \le N_{b-1}/(2N_b)$. As a result, let $l = M^{-1/2}$, by Hölder's inequality, for any $\boldsymbol{\beta} \in \boldsymbol{\beta}^0 + \mathbb{C}(l)$

$$\begin{aligned} B_1 &\ge \frac{N_{b-1}}{N_b}\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|_2^2 \tau_{\min}\left\{\mathbf{H}(\boldsymbol{\beta}^0)\right\} - \|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|_1^2(\|\boldsymbol{K}_1\|_{\max} + \|\boldsymbol{K}_2\|_{\max} + \|\boldsymbol{K}_3\|_{\max}) \\ &\ge \frac{N_{b-1}}{N_b}\frac{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|_2^2}{2M}. \end{aligned} \tag{S2.54}$$

While for the term B_2, by Taylor's expansion

$$\begin{aligned} B_2 &= \frac{1}{N_b}(\boldsymbol{\beta}-\boldsymbol{\beta}^0)^\top \frac{1}{n_b}\sum_{i=1}^{n_b} L''_{h_b}(y_{bi}\mathbf{x}_{bi}^\top\widetilde{\boldsymbol{\beta}})\mathbf{x}_{bi}\mathbf{x}_{bi}^\top(\boldsymbol{\beta}-\boldsymbol{\beta}^0) \\ &= \frac{1}{N_b}(\boldsymbol{\beta}-\boldsymbol{\beta}^0)^\top \frac{1}{n_b h_b}\sum_{i=1}^{n_b} K\big(r_{bi}(\widetilde{\boldsymbol{\beta}})/h_b\big)\mathbf{x}_{bi}\mathbf{x}_{bi}^\top(\boldsymbol{\beta}-\boldsymbol{\beta}^0) \\ &\ge \|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|_2^2 \frac{n_b}{N_b}\tau_{\min}\left[\mathbb{E}\left\{\frac{1}{n_b h}\sum_{i=1}^{n_b} K\big(r_{bi}(\widetilde{\boldsymbol{\beta}})/h_b\big)\mathbf{x}_{bi}\mathbf{x}_{bi}^\top\right\}\right] - \frac{n_b}{N_b}\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|_1^2 \\ &\quad \times \sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}^0+\mathbb{B}(r)\cap\mathbb{C}(l)} \left\|\frac{1}{n_b h_b}\sum_{i=1}^{n_b} K\big(r_{bi}(\boldsymbol{\beta})/h_b\big)\mathbf{x}_{bi}\mathbf{x}_{bi}^\top - \mathbb{E}\left\{\frac{1}{n_b h_b}\sum_{i=1}^{n_b} K\big(r_{bi}(\boldsymbol{\beta})/h_b\big)\mathbf{x}_{bi}\mathbf{x}_{bi}^\top\right\}\right\|_{\max}. \end{aligned} \tag{S2.55}$$

Here $\mathbf{x}_{bi}^\top\widetilde{\boldsymbol{\beta}}$ is among $\mathbf{x}_{bi}^\top\boldsymbol{\beta}^0$ and $\mathbf{x}_{bi}^\top\boldsymbol{\beta}$. Next, we evaluate the random fluctuation term

$$\begin{aligned} D &:= \sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}^0+\mathbb{B}(r)\cap\mathbb{C}(l)} \left\|\frac{1}{n_b h_b}\sum_{i=1}^{n_b} K\big(r_{bi}(\boldsymbol{\beta})/h_b\big)\mathbf{x}_{bi}\mathbf{x}_{bi}^\top - \mathbb{E}\left\{\frac{1}{n_b h_b}\sum_{i=1}^{n_b} K\big(r_{bi}(\boldsymbol{\beta})/h_b\big)\mathbf{x}_{bi}\mathbf{x}_{bi}^\top\right\}\right\|_{\max} \\ &= \sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}^0+\mathbb{B}(r)\cap\mathbb{C}(l)} \max_{j,k}\left|\frac{1}{n_b h_b}\sum_{i=1}^{n_b} K\big(r_{bi}(\boldsymbol{\beta})/h_b\big)x_{bi,j}x_{bi,k} - \mathbb{E}\left\{\frac{1}{n_b h_b}\sum_{i=1}^{n_b} K\big(r_{bi}(\boldsymbol{\beta})/h_b\big)x_{bi,j}x_{bi,k}\right\}\right|. \end{aligned}$$

Define $\omega_{j,k}^{(i)} = K(r_{bi}(\boldsymbol{\beta})/h_b)x_{bi,j}x_{bi,k}$. Note that $|K(\cdot)| \le \kappa_{\max}$ by Assumption 5 and $|x_{bi,j}x_{bi,k}|_1 \le u_{\mathbf{x}}^2$ by Assumption 1. Thus we know that $|\omega_{j,k}^{(i)}| \le \kappa_{\max}u_{\mathbf{x}}^2$. Then it

follows that

$$P\left[\left|\frac{1}{n_b}\sum_{i=1}^{n_b}\omega_{j,k}^{(i)}-\mathbb{E}\left\{\frac{1}{n_b}\sum_{i=1}^{n_b}\omega_{j,k}^{(i)}\right\}\right|\geq t\right]\leq 2\exp\left(-\frac{n_b t^2}{2\kappa_{\max}^2 u_{\mathbf{x}}^4}\right).$$

By union bound

$$P(Dh_b\geq t)\leq\sum_{j=1}^{p}\sum_{k=1}^{p}P\left[\left|\frac{1}{n_b}\sum_{i=1}^{n_b}\omega_{j,k}^{(i)}-\mathbb{E}\left\{\frac{1}{n_b}\sum_{i=1}^{n_b}\omega_{j,k}^{(i)}\right\}\right|\geq t\right]\leq 2p^2\exp\left(\frac{n_b t^2}{2\kappa_{\max}^2 u_{\mathbf{x}}^4}\right),$$

thus with probability $1-2p^{-6}$, we have

$$Dh_b\leq 4\kappa_{\max}u_{\mathbf{x}}^2\sqrt{\log(p)/n_b}. \tag{S2.56}$$

Combining (S2.55) and (S2.56), by Assumption 2, we conclude that for any $\boldsymbol{\beta}\in\boldsymbol{\beta}^0+\mathbb{B}(r)\cap\mathbb{C}(l)$ with $r=o(1)$,

$$\begin{aligned}
B_2 &\geq \frac{n_b}{N_b}\tau_{\min}\{\mathbf{H}_{h_b}(\widetilde{\boldsymbol{\beta}})\}\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|_2^2-\frac{n_b}{N_b}D\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|_1^2\\
&\geq \frac{n_b}{N_b}(\rho-l^2s_0D)\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|_2^2\\
&\geq \frac{n_b}{N_b}\left\{\rho-4\kappa_{\max}u_{\mathbf{x}}^2l^2s_0\sqrt{\frac{\log(p)}{n_bh_b^2}}\right\}\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|_2^2.
\end{aligned}$$

Note that $h_b=O\big[\{s_0\log(p)/N_b\}^{1/4}\big]$, we have

$$s_0\sqrt{\frac{\log(p)}{n_bh_b^2}}\asymp\frac{s_0^{3/4}\log^{1/4}(p)(N_b/n_b)^{1/2}}{N_b^{1/4}}\asymp\frac{s_0^{3/4}\log^{1/4}(p)b^{1/2}}{N_b^{1/4}}.$$

Given $s_0^9\log(p)b^2\{1+\log(N_{b-1}/n_b)\}^2=o(N_b)$, we know that $s_0\sqrt{\log(p)/(n_bh_b^2)}=o(1)$. It then follows that

$$B_2\geq\frac{n_b}{N_b}\frac{\rho}{2}\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|_2^2. \tag{S2.57}$$

As a result, combining (S2.53), (S2.54) and (S2.57), this lemma is demonstrated. □

S3 Additional theoretical results

S3.1 Classification accuracy

We discuss the classification accuracy in this section. Given parameter $\boldsymbol{\beta}$, the classification rule of SVM is $\mathbf{sign}(\mathbf{x}^\top\boldsymbol{\beta})$ and the classification accuracy can be defined as

$$\begin{aligned}\mathrm{Acc}(\boldsymbol{\beta}) =& P(\mathbf{x}^\top\boldsymbol{\beta} > 0, y = 1) + P(\mathbf{x}^\top\boldsymbol{\beta} < 0, y = -1)\\ =& \pi_+ P(\mathbf{x}^\top\boldsymbol{\beta} > 0 \mid y = 1) + \pi_- P(\mathbf{x}^\top\boldsymbol{\beta} < 0 \mid y = -1),\end{aligned}$$

where $\pi_+ = P(y = 1)$ and $\pi_- = P(y = -1)$. Recall that $f(\cdot)$ and $g(\cdot)$ are the density functions of $\mathbf{x}$ given $y = 1$ and $y = -1$, respectively. For a $\boldsymbol{\beta} \in \mathbb{R}^p$, we then have

$$\begin{aligned}&\big|\mathrm{Acc}(\boldsymbol{\beta}) - \mathrm{Acc}(\boldsymbol{\beta}^0)\big|\\ \le& \pi_+\big|P(\mathbf{x}^\top\boldsymbol{\beta} > 0 \mid y = 1) - P(\mathbf{x}^\top\boldsymbol{\beta}^0 > 0 \mid y = 1)\big|\\ &+ \pi_-\big|P(\mathbf{x}^\top\boldsymbol{\beta} < 0 \mid y = -1) - P(\mathbf{x}^\top\boldsymbol{\beta}^0 < 0 \mid y = -1)\big|\\ =& \pi_+\left|\int_{\mathbf{x}^\top\boldsymbol{\beta}>0} f(\mathbf{x})\mathrm{d}\mathbf{x} - \int_{\mathbf{x}^\top\boldsymbol{\beta}^0>0} f(\mathbf{x})\mathrm{d}\mathbf{x}\right| + \pi_-\left|\int_{\mathbf{x}^\top\boldsymbol{\beta}>0} g(\mathbf{x})\mathrm{d}\mathbf{x} - \int_{\mathbf{x}^\top\boldsymbol{\beta}^0>0} g(\mathbf{x})\mathrm{d}\mathbf{x}\right|.\end{aligned} \tag{S3.58}$$

Let $\boldsymbol{\alpha} = \boldsymbol{\beta} - \boldsymbol{\beta}^0$ and $\boldsymbol{\gamma}(t) = \boldsymbol{\beta}^0 + t\boldsymbol{\alpha}$, we know that $\boldsymbol{\gamma}(0) = \boldsymbol{\beta}^0$ and $\boldsymbol{\gamma}(1) = \boldsymbol{\beta}$. Define $I(t) = \int_{\mathbf{x}^\top \boldsymbol{\gamma}(t) > 0} f(\mathbf{x}) \mathrm{d}\mathbf{x}$, we then have

$$\int_{\mathbf{x}^\top \boldsymbol{\beta} > 0} f(\mathbf{x}) \mathrm{d}\mathbf{x} - \int_{\mathbf{x}^\top \boldsymbol{\beta}^0 > 0} f(\mathbf{x}) \mathrm{d}\mathbf{x} = I(1) - I(0) = \int_0^1 \frac{\mathrm{d}I(t)}{\mathrm{d}t} \mathrm{d}t.$$

We next compute the derivative $\mathrm{d}I(t)/\mathrm{d}t$. Define the integral area $\Omega(t) = \{\mathbf{x} \in \mathbb{R}^p \mid \mathbf{x}^\top \boldsymbol{\gamma}(t) > 0\}$ and its boundary hyperplane $\partial\Omega(t) = \{\mathbf{x} \in \mathbb{R}^p \mid \mathbf{x}^\top \boldsymbol{\gamma}(t) = 0\}$. Denote the normal vector $\boldsymbol{n}(t) = \boldsymbol{\gamma}(t)/\|\boldsymbol{\gamma}(t)\|$ and the normal velocity as $\boldsymbol{v}$. A boundary point $\mathbf{x} \in \partial\Omega(t)$ satisfies

$$0 = \frac{\mathrm{d}\{\mathbf{x}^\top \boldsymbol{\gamma}(t)\}}{\mathrm{d}t} = \boldsymbol{v}^\top \boldsymbol{\gamma}(t) + \mathbf{x}^\top \frac{\mathrm{d}\{\boldsymbol{\gamma}(t)\}}{\mathrm{d}t},$$

thus we have $\boldsymbol{v}^\top \boldsymbol{n}(t) = -\mathbf{x}^\top \boldsymbol{\alpha}/\|\boldsymbol{\gamma}(t)\|_2$. By Leibniz's law (Chapter 3 in Samko et al. (1993)), we then derive

$$\begin{aligned} \frac{\mathrm{d}I(t)}{\mathrm{d}t} &= \int_{\Omega(t)} \frac{\partial f(\mathbf{x})}{\partial t} \mathrm{d}\mathbf{x} + \int_{\partial\Omega(t)} \{\boldsymbol{v}^\top \boldsymbol{n}(t)\} f(\mathbf{x}) \mathrm{d}S \\ &= -\frac{1}{\|\boldsymbol{\gamma}(t)\|_2} \int_{\partial\Omega(t)} (\mathbf{x}^\top \boldsymbol{\alpha}) \cdot f(\mathbf{x}) \mathrm{d}S, \end{aligned}$$

where $\mathrm{d}S$ is the area element of $\partial\Omega(t)$. Recall that in Assumption 3 we imposed $(\beta_d^0)^{-1} \leq c$. Here we assume the marginal density of x_d is upper-bounded as $\sup_{x \in \mathbb{R}} f_d(x) \leq F_d$. Further note that $x_d = -\sum_{j \neq d} \gamma_j(t) x_j / \gamma_d(t)$ for $\mathbf{x} \in \partial\Omega(t)$, where $\gamma_j(t)$ is the j-th element of $\boldsymbol{\gamma}(t)$. Thus we have

$$\mathrm{d}S = \frac{\|\boldsymbol{\gamma}(t)\|_2}{|\gamma_d(t)|} \mathrm{d}x_1 \cdots \mathrm{d}x_{d-1} \mathrm{d}x_{d+1} \cdots \mathrm{d}x_p.$$

It then follows that

$$\begin{aligned}
&\left| \int_{\mathbf{x}^\top \boldsymbol{\beta}>0} f(\mathbf{x})\mathrm{d}\mathbf{x} - \int_{\mathbf{x}^\top \boldsymbol{\beta}^0>0} f(\mathbf{x})\mathrm{d}\mathbf{x} \right| \\
=&\left| \int_0^1 -\frac{1}{\|\boldsymbol{\gamma}(t)\|_2} \Big\{ \int_{\mathbf{x}^\top \boldsymbol{\gamma}(t)=0} \mathbf{x}^\top \boldsymbol{\alpha} f(\mathbf{x})\mathrm{d}S \Big\} \mathrm{d}t \right| \\
=&\left| \int_0^1 \frac{1}{|\gamma_d(t)|} \Big\{ \int_{\mathbb{R}^{p-1}} (\mathbf{x}^\top \boldsymbol{\alpha}) f(\mathbf{x})\mathrm{d}x_1 \cdots \mathrm{d}x_{d-1}\mathrm{d}x_{d+1} \cdots \mathrm{d}x_p \Big\} \mathrm{d}t \right| \\
\leq&\left| \int_0^1 \frac{1}{|\gamma_d(t)|} \int_{\mathbb{R}^{p-1}} |\mathbf{x}^\top \boldsymbol{\alpha}| \cdot f(x_d, \mathbf{x}_{-d})\mathrm{d}\mathbf{x}_{-d}\mathrm{d}t \right|. \qquad \text{(S3.59)}
\end{aligned}$$

By the triangle inequality, we have $|\gamma_d(t)| \geq |\beta_d^0| - |t\alpha_d| \geq |\beta_d^0| - |\alpha_d|$. Further, from Theorem 2 we know that $\|\widehat{\boldsymbol{\beta}}_{h_b}^b - \boldsymbol{\beta}^0\|_1 \leq c_b^{(1)} s_0 \sqrt{\log(p)/N_b}$ and $\|\widehat{\boldsymbol{\beta}}_{h_b}^b - \boldsymbol{\beta}^0\|_2 \leq c_b^{(2)} \sqrt{s_0 \log(p)/N_b}$. Substituting $\widehat{\boldsymbol{\beta}}_{h_b}^b$ into $\boldsymbol{\beta}$ in (S3.59), we then have with probability approaching 1,

$$\begin{aligned}
\left| \int_{\mathbf{x}^\top \widehat{\boldsymbol{\beta}}_{h_b}^b>0} f(\mathbf{x})\mathrm{d}\mathbf{x} - \int_{\mathbf{x}^\top \boldsymbol{\beta}^0>0} f(\mathbf{x})\mathrm{d}\mathbf{x} \right| \leq& \frac{\|\mathbf{x}\|_\infty \|\widehat{\boldsymbol{\beta}}_{h_b}^b - \boldsymbol{\beta}^0\|_1}{|\beta_d^0| - \|\widehat{\boldsymbol{\beta}}_{h_b}^b - \boldsymbol{\beta}^0\|_2} \int_0^1 \int_{\mathbb{R}^{p-1}} f(x_d, \mathbf{x}_{-d})\mathrm{d}\mathbf{x}_{-d}\mathrm{d}t \\
\leq& \frac{F_d u_{\mathbf{x}} c_b^{(1)} s_0 \sqrt{\log(p)/N_b}}{|\beta_d^0| - c_b^{(2)} \sqrt{s_0 \log(p)/N_b}}.
\end{aligned}$$

Given that $N_b \geq 2c\{c_b^{(2)}\}^2 s_0 \log(p)$, we then obatin

$$\left| \int_{\mathbf{x}^\top \widehat{\boldsymbol{\beta}}_{h_b}^b>0} f(\mathbf{x})\mathrm{d}\mathbf{x} - \int_{\mathbf{x}^\top \boldsymbol{\beta}^0>0} f(\mathbf{x})\mathrm{d}\mathbf{x} \right| \leq c F_d u_{\mathbf{x}} c_b^{(1)} s_0 \sqrt{\log(p)/N_b}.$$

Analogous result holds when replacing $f(\mathbf{x})$ with $g(\mathbf{x})$. Consequently, we conclude from (S3.58) that

$$\left|\mathrm{Acc}(\widehat{\boldsymbol{\beta}}_{h_b}^b) - \mathrm{Acc}(\boldsymbol{\beta}^0)\right| = O_p\Big\{ s_0 \sqrt{\log(p)/N_b} \Big\}.$$

S3.2 Theoretical results and proof when targeting the true parameter of DCSVM

This subsection presents the theoretical results and the corresponding proof when targeting the true parameter $\boldsymbol{\beta}^{0h}$ of DCSVM. Since all the theoretical results are quite similar to those for $\boldsymbol{\beta}^0$, we will briefly present the proof process and mainly focus on the aspects different from the existing proofs.

As discussed in Remark 2, the bandwidth for DCSVM is selected from constants. Therefore, we let $h_1 = \cdots = h_b = h$. To start with, we make some modifications and use the following assumptions:

Assumption 1. There exist positive constants M_1, M_2 such that $0 < \tau_{\max}(\mathbf{H}_h(\boldsymbol{\beta}^{0h})) \leq M_1$ and $\min_{\|\boldsymbol{\xi}\|_2^2 = o(1)} \tau_{\min}\{\mathbf{H}_h(\boldsymbol{\beta}^{0h} + \boldsymbol{\xi})\} \geq M_2$.

This assumption is modified from Assumption 2 and Assumption 2 in Wang et al. (2023).

Assumption 2. Denote $\boldsymbol{\Theta}_0^h = \{\mathbf{H}_h(\boldsymbol{\beta}^{0h})\}^{-1}$. Assume that $\|\boldsymbol{\Theta}_0^h\|_{L_1} \leq C_0$ and $\max_{1 \leq i \leq p} \sum_{j=1}^p |\Theta_{0,ij}^h|^q < C_r$ for $q \in [0, 1)$.

This assumption is modified from Assumption 6.

Lemma 6 (Choice of tuning parameter λ_l)**.** *Suppose that Assumption 1 holds. If* $\lambda_l = c_0\sqrt{\log(p)/N_l}$ *for some constants* $c_0 > 0$*, we have, for any* $l = 1, 2, ..., b$

$$P\left\{\theta_l \|\widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_l)\|_\infty \leq \lambda_l\right\} \geq 1 - 2p^{1 - c_0^2/(2u_{\mathbf{x}}^2)},$$

where $\theta_l = \sqrt{n_l/N_l} \in (0,1)$.

Proof of Lemma 6. Note that $\boldsymbol{S}_h(\boldsymbol{\beta}^{0h}) = 0$ by definition. This lemma is then proved using the same argument as the proof of Lemma 2 in the main text. □

Lemma 7. *Denote* $\widetilde{\mathbf{H}}_h(\boldsymbol{\beta}) = \frac{1}{N_{b-1}} \sum_{l=1}^{b-1} n_l \mathbf{H}_h(\boldsymbol{\beta})$. *Under Assumption 1 and 5, we have*

$$\left\| \widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_h^l; \mathcal{D}_l) - \widehat{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_l) \right\|_{\max} \leq \frac{C_1 u_{\mathbf{x}}^3}{h^2} \|\widehat{\boldsymbol{\beta}}_h^l - \boldsymbol{\beta}^{0h}\|_1, \tag{S3.60}$$

$$\left\| \frac{1}{N_{l-1}} \sum_{l=1}^{b-1} n_l \widehat{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_l) - \widetilde{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}) \right\|_{\max} \leq C_2 \sqrt{\frac{\log(p)}{h^2 N_{b-1}}}, \tag{S3.61}$$

hold with probability $1 - p^{-C'}$ *for some constant* $C' > 0$.

Proof of Lemma 7. To prove (S3.60), note that by Assumption 1 and 5,

$$\begin{aligned}
&\left\| \widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_h^l; \mathcal{D}_l) - \widehat{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_l) \right\|_{\max} \\
=& \max_{j,k} \left| \frac{1}{n_l h} \sum_{i=1}^{n_l} x_{li,j} x_{li,k} \left[K\left\{ \frac{r_{li}(\widehat{\boldsymbol{\beta}}_h^l)}{h} \right\} - K\left\{ \frac{r_{li}(\boldsymbol{\beta}^{0h})}{h} \right\} \right] \right| \\
=& \max_{j,k} \left| \frac{1}{n_l h^2} \sum_{i=1}^{n_l} x_{li,j} x_{li,k} K'\left\{ \frac{r_{li}(\widetilde{\boldsymbol{\beta}}_l)}{h} \right\} \mathbf{x}_{li}^\top (\widehat{\boldsymbol{\beta}}_h^l - \boldsymbol{\beta}^{0h}) \right| \\
\leq& \max_{j,k,l} \left| \frac{1}{n_l h^2} \sum_{i=1}^{n_l} x_{li,j} x_{li,k} x_{li,l} K'\left\{ \frac{r_{li}(\widetilde{\boldsymbol{\beta}}_l)}{h} \right\} \right| \|\widehat{\boldsymbol{\beta}}_h^l - \boldsymbol{\beta}^{0h}\|_1 \\
\leq& \frac{C_1}{n_l h^2} \sum_{i=1}^{n_l} \max_j |x_{li,j}| \max_k |x_{li,k}| \max_l |x_{li,l}| \|\widehat{\boldsymbol{\beta}}_h^l - \boldsymbol{\beta}^{0h}\|_1 \leq \frac{C_1 u_{\mathbf{x}}^3}{h^2} \|\widehat{\boldsymbol{\beta}}_h^l - \boldsymbol{\beta}^{0h}\|_1,
\end{aligned}$$

where $\widetilde{\boldsymbol{\beta}}_l$ is mediation value between $\widehat{\boldsymbol{\beta}}_h^l$ and $\boldsymbol{\beta}^{0h}$. The proof of (S3.61) is similar to the proof of (S2.47) in Lemma 3. □

Proposition 1. *Define the event*

$$\mathcal{Q}_b(r,l,\kappa)=\left\{\frac{\langle\nabla\mathcal{L}_b(\boldsymbol{\beta})-\nabla\mathcal{L}_b(\boldsymbol{\beta}^{0h}),\boldsymbol{\beta}-\boldsymbol{\beta}^{0h}\rangle}{\|\boldsymbol{\beta}-\boldsymbol{\beta}^{0h}\|_2^2}\geq\kappa,\ \forall\boldsymbol{\beta}\in\boldsymbol{\beta}^{0h}+\mathbb{B}(r)\cap\mathbb{C}(l)\right\}.$$

Suppose that $\|\widehat{\boldsymbol{\beta}}_h^l-\boldsymbol{\beta}^{0h}\|_2^2=O(s_0\log(p)/N_l)$ *for* $l=1,2,...,b-1$. *Under the conditions* $s_0\log(p)/h=o(n_l^{1/4})$, *we have* $P\left(\mathcal{Q}_b(r,l,\kappa)\right)\to 1$.

Proof of Proposition 1. Similar to the proof of Lemma 5, we have

$$\begin{aligned}
&\langle\nabla\mathcal{L}_b(\boldsymbol{\beta})-\nabla\mathcal{L}_b(\boldsymbol{\beta}^{0h}),\boldsymbol{\beta}-\boldsymbol{\beta}^{0h})\rangle\\
=&\frac{1}{N_b}(\boldsymbol{\beta}-\boldsymbol{\beta}^{0h})^\top\sum_{l=1}^{b-1}n_l\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_h^l;\mathcal{D}_l)(\boldsymbol{\beta}-\boldsymbol{\beta}^{0h})\\
&+\frac{1}{N_b}\sum_{i=1}^{n_b}\left\{L_h^{'}(y_{bi}\mathbf{x}_{bi}^\top\boldsymbol{\beta})-L_h^{'}(y_{bi}\mathbf{x}_{bi}^\top\boldsymbol{\beta}^{0h})\right\}y_{bi}\langle\mathbf{x}_{bi},\boldsymbol{\beta}-\boldsymbol{\beta}^{0h}\rangle\\
=:&\ B_1+B_2.
\end{aligned}$$

For the term B_1, note that

$$\begin{aligned}
\frac{1}{N_b}\sum_{l=1}^{b-1}n_l\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_h^l;\mathcal{D}_l)=&\frac{1}{N_b}\sum_{l=1}^{b-1}n_l\{\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_h^l;\mathcal{D}_l)-\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_0;\mathcal{D}_l)\}\\
+\frac{1}{N_b}\sum_{l=1}^{b-1}n_l\{\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_0;\mathcal{D}_l)-\mathbf{H}_h(\boldsymbol{\beta}^{0h})\}&+\frac{1}{N_b}\sum_{l=1}^{b-1}n_l\mathbf{H}_h(\boldsymbol{\beta}^{0h}).
\end{aligned}$$

Given that $\|\widehat{\boldsymbol{\beta}}_h^l-\boldsymbol{\beta}^{0h}\|_1=O\{s_0\sqrt{\log(p)/N_l}\}$ for $s=1,2,...,b-1$, by Lemma 7,

$$\begin{aligned}
\|\boldsymbol{\Delta}_1\|_{\max}&\leq\frac{1}{N_b}\sum_{l=1}^{b-1}n_l\left\|\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_h^l;\mathcal{D}_l)-\widehat{\mathbf{H}}_h(\boldsymbol{\beta}^{0h};\mathcal{D}_l)\right\|_{\max}=O_p\left\{s_0\log^2(p)/N_b^{\frac{1}{2}}\right\};\\
\|\boldsymbol{\Delta}_2\|_{\max}&=\left\|\frac{1}{N_b}\sum_{l=1}^{b-1}n_l\{\widehat{\mathbf{H}}_h(\boldsymbol{\beta}^{0h};\mathcal{D}_l)-\mathbf{H}_{h_l}(\boldsymbol{\beta}^{0h})\}\right\|_{\max}=O_p\left\{\log^{\frac{1}{2}}(p)/N_b^{\frac{1}{2}2}\right\}.
\end{aligned}$$

Under the condition $s_0 \log(p) = O(N_b^{1/4})$, we get $s_0(\|\mathbf{\Delta}_1\|_{\max}+\|\mathbf{\Delta}_2\|_{\max}) \leq N_{b-1}/(2N_b)$. By Hölder's inequality, for any $\boldsymbol{\beta} \in \boldsymbol{\beta}^{0h} + \mathbb{C}(\sqrt{M_2})$,

$$\begin{aligned} B_1 &\geq \frac{N_{b-1}}{N_b}\|\boldsymbol{\beta} - \boldsymbol{\beta}^{0h}\|_2^2 \tau_{\min}\left\{\mathbf{H}_h(\boldsymbol{\beta}^{0h})\right\} - \|\boldsymbol{\beta} - \boldsymbol{\beta}^{0h}\|_1^2(\|\mathbf{\Delta}_1\|_{\max} + \|\mathbf{\Delta}_2\|_{\max}) \\ &\geq \frac{N_{b-1}}{N_b}\frac{M_2}{2}\|\boldsymbol{\beta} - \boldsymbol{\beta}^{0h}\|_2^2. \end{aligned}$$

For the term B_2, using the same argument as the proof of Lemma 5, we can derive

$$B_2 \geq \frac{n_b}{N_b}\left\{M_2 - 4\kappa_{\max}u_{\mathbf{x}}^2 l^2 s_0\sqrt{\frac{\log(p)}{n_b h^2}}\right\}\|\boldsymbol{\beta} - \boldsymbol{\beta}^{0h}\|_2^2.$$

Given that $s_0^2\log(p)/h^2 = o(n_b)$, we have $B_2 \geq (n_b M_2)/(2N_b)\|\boldsymbol{\beta} - \boldsymbol{\beta}^{0h}\|_2^2$. Thus the lemma is demonstrated. □

Theorem 1. *Under Assumption 1-5, for $l = 1, 2, ..., b$, let the tuning parameter $\lambda_l = O\left\{\sqrt{\log(p)/N_l}\right\}$ and the sparsity s_0 satisfy $s_0\log(p)/h = o(n_l^{1/4})$. Then, the renewable estimator $\widehat{\boldsymbol{\beta}}_h^b$ in (2.12) satisfies the following error bounds:*

$$\|\widehat{\boldsymbol{\beta}}_h^b - \boldsymbol{\beta}^{0h}\|_1 \leq c_b^{(1)} s_0\lambda_b, \quad \|\widehat{\boldsymbol{\beta}}_h^b - \boldsymbol{\beta}^{0h}\|_2 \leq c_b^{(2)}\sqrt{s_0}\lambda_b,$$

with probability $1 - p^{-C}$, for some positive constant C. Here the constants $c_b^{(1)}$ and $c_b^{(2)}$ do not rely on the sample size N_b.

Proof of Theorem 1. Likewise, this theorem is proved through mathematical induction with the following steps:

(i). $\|\widehat{\boldsymbol{\beta}}_h^1 - \boldsymbol{\beta}^{0h}\|_1 \leq c_1^{(1)} s_0\lambda_1$ and $\|\widehat{\boldsymbol{\beta}}_h^1 - \boldsymbol{\beta}^{0h}\|_2 \leq c_1^{(2)}\sqrt{s_0}\lambda_1$.

(ii). Given (i), $\|\widehat{\boldsymbol{\beta}}_h^2 - \boldsymbol{\beta}^{0h}\|_1 \leq c_2^{(1)} s_0 \lambda_2$ and $\|\widehat{\boldsymbol{\beta}}_h^2 - \boldsymbol{\beta}^{0h}\|_2 \leq c_2^{(2)} \sqrt{s_0}\lambda_2$.

(iii). Given that $\widehat{\boldsymbol{\beta}}_h^{b-1}$ satisfies the error bound, then $\|\widehat{\boldsymbol{\beta}}_h^b - \boldsymbol{\beta}^{0h}\|_1 \leq c_b^{(1)} s_0 \lambda_b$ and $\|\widehat{\boldsymbol{\beta}}_h^b - \boldsymbol{\beta}^{0h}\|_2 \leq c_b^{(2)} \sqrt{s_0}\lambda_b$.

(i) has been demonstrated by Wang et al. (2023) provided that $\lambda_1 = O\{\sqrt{\log(p)/n_1}\}$. Define $\widehat{\boldsymbol{\xi}}_l = \widehat{\boldsymbol{\beta}}_h^l - \boldsymbol{\beta}^{0h}, l = 1, \ldots, b$. First note the decomposition

$$\begin{aligned}
&\langle \nabla \mathcal{L}_2(\widehat{\boldsymbol{\beta}}_h^2) - \nabla \mathcal{L}_2(\boldsymbol{\beta}^{0h}), \widehat{\boldsymbol{\beta}}_h^2 - \boldsymbol{\beta}^{0h} \rangle \\
=&\lambda_2 \langle \boldsymbol{\kappa}_2, -\widehat{\boldsymbol{\xi}}_2 \rangle + \frac{n_2}{N_2} \langle \widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_2), -\widehat{\boldsymbol{\xi}}_2 \rangle + \frac{1}{N_2} \langle \widehat{\mathbf{Q}}_1 \widehat{\boldsymbol{\xi}}_1, \widehat{\boldsymbol{\xi}}_2 \rangle := A_1 + A_2 + A_3.
\end{aligned}$$

Using the same argument as the proof of Theorem 2, by Lemma 6 we have

$$A_1 \leq \lambda_2 (\|\widehat{\boldsymbol{\xi}}_{2,S_0}\|_1 - \|\widehat{\boldsymbol{\xi}}_{2,S_0^c}\|_1), \quad |A_2| \leq \lambda_2 \|\widehat{\boldsymbol{\xi}}_2\|_1 / 3,$$

with probability approaching 1. For A_3, we have

$$\begin{aligned}
A_3 =& \frac{n_1}{N_2} \widehat{\boldsymbol{\xi}}_1^\top \{\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_h^1; \mathcal{D}_1) - \widehat{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_1)\} \widehat{\boldsymbol{\xi}}_2 + \frac{n_1}{N_2} \widehat{\boldsymbol{\xi}}_1^\top \{\widehat{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_1) - \mathbf{H}_h(\boldsymbol{\beta}^{0h})\} \widehat{\boldsymbol{\xi}}_2 \\
&+ \frac{n_1}{N_2} \widehat{\boldsymbol{\xi}}_1^\top \{\mathbf{H}_h(\boldsymbol{\beta}^{0h})\} \widehat{\boldsymbol{\xi}}_2 \\
=:& A_{31} + A_{32} + A_{33}.
\end{aligned}$$

Given that $\|\widehat{\boldsymbol{\xi}}_1\|_1 \leq c_1^{(1)} s_0 \sqrt{\log(p)/n_1}$, by Lemma 7, we have

$$\begin{aligned}
\|\boldsymbol{J}_1\|_{\max} :=& \|\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_h^1; \mathcal{D}_1) - \widehat{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_1)\|_{\max} \leq \frac{s_0 \log^2(p)}{\sqrt{n_1} h^2}, \\
\|\boldsymbol{J}_2\|_{\max} :=& \|\widehat{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_1) - \mathbf{H}_h(\boldsymbol{\beta}^{0h})\|_{\max} \leq C \sqrt{\frac{\log(p)}{n_1 h^2}}.
\end{aligned}$$

Note that $\lambda_1\sqrt{n_1} \asymp \lambda_2\sqrt{N_2}$ and $\|\widehat{\boldsymbol{\xi}}_1\|_1 \leq c_1^{(1)}\lambda_1\sqrt{\log(p)/n_1}$, we then derive

$$|A_3| \leq Cc_1^{(1)}\left\{s_0^2\sqrt{\frac{\log^4(p)}{n_1h^4}} + s_0\sqrt{\frac{\log(p)}{n_1h^2}}\right\}\lambda_2\|\widehat{\boldsymbol{\xi}}_2\|_1 + \tau_{\max}\{\mathbf{H}_h(\boldsymbol{\beta}^{0h})\}c_1^{(2)}\sqrt{s_0}\lambda_2\|\widehat{\boldsymbol{\xi}}_2\|_2.s$$

Provided that $s_0\log(p)/h = o(n_1^{1/4})$, $|A_3|$ can be controlled by $\lambda_2\|\widehat{\boldsymbol{\xi}}_2\|_1/3 + M_1c_1^{(2)}\sqrt{s_0}\lambda_2\|\widehat{\boldsymbol{\xi}}_2\|_2$.

By triangle and Cauchy-Schwartz inequalities, we have

$$\begin{aligned}
&\langle\nabla\mathcal{L}_2(\widehat{\boldsymbol{\beta}}_h^2) - \nabla\mathcal{L}_2(\boldsymbol{\beta}^{0h}), \widehat{\boldsymbol{\beta}}_h^2 - \boldsymbol{\beta}^{0h})\rangle \\
\leq&\lambda_2(\|\widehat{\boldsymbol{\xi}}_{2,S_0}\|_1 - \|\widehat{\boldsymbol{\xi}}_{2,S_0^c}\|_1) + \frac{1}{3}\lambda_2\|\widehat{\boldsymbol{\xi}}_2\|_1 + M_1c_1^{(2)}\sqrt{s_0}\lambda_2\|\widehat{\boldsymbol{\xi}}_2\|_2 \\
\leq&\lambda_2\sqrt{s_0}\{2 + M_1c_1^{(2)}\}\|\widehat{\boldsymbol{\xi}}_2\|_2.
\end{aligned}$$

Moreover, using the fact $\langle\nabla\mathcal{L}_2(\widehat{\boldsymbol{\beta}}_h^2) - \nabla\mathcal{L}_2(\boldsymbol{\beta}^{0h}), \widehat{\boldsymbol{\beta}}_h^2 - \boldsymbol{\beta}^{0h})\rangle \geq 0$, we could derive the cone-type constraint $\|\widehat{\boldsymbol{\xi}}_2\|_1 \leq 3(2 + M_1c_1^{(2)})\sqrt{s_0}\|\widehat{\boldsymbol{\xi}}_2\|_2$. Combining this condition with Proposition 1, using the same apagoge as in proof of Theorem 2, we show that $\|\widehat{\boldsymbol{\xi}}_2\|_2 \leq 2(2 + M_1c_1^{(2)})\lambda_2\sqrt{s_0}/M_2$. By the cone-type condition, we further establish the ℓ_1-norm error bound for $\widehat{\boldsymbol{\beta}}_h^2$. Thus (ii) is confirmed.

Turn to (iii). We similarly decompose

$$\begin{aligned}
&\langle\nabla\mathcal{L}_b(\widehat{\boldsymbol{\beta}}_h^b) - \nabla\mathcal{L}_b(\boldsymbol{\beta}^{0h}), \widehat{\boldsymbol{\beta}}_h^b - \boldsymbol{\beta}^{0h}\rangle \\
=&\lambda_b\langle\boldsymbol{\kappa}_b, -\widehat{\boldsymbol{\xi}}_b\rangle + \frac{n_b}{N_b}\langle\widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_b), -\widehat{\boldsymbol{\xi}}_b\rangle + \frac{1}{N_b}\langle\widehat{\mathbf{Q}}_{b-1}\widehat{\boldsymbol{\xi}}_{b-1}, \widehat{\boldsymbol{\xi}}_{b-1}\rangle \\
:=&D_1 + D_2 + D_3.
\end{aligned}$$

Similar to the proof of $b = 2$, by lemma 6 we hav with probability approaching 1,

$$D_1 \leq \lambda_b(\|\widehat{\boldsymbol{\xi}}_{b,S_0}\|_1 - \|\widehat{\boldsymbol{\xi}}_{b,S_0^c}\|_1), \quad |D_2| \leq \lambda_b\|\widehat{\boldsymbol{\xi}}_b\|_1/3.$$

For D_3, denote $\widetilde{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}) = (N_{b-1})^{-1}\sum_{l=1}^{b-1} n_l \mathbf{H}_h(\boldsymbol{\beta}^{0h})$, we can write

$$\begin{aligned}
D_3 =& \frac{1}{N_b}\widehat{\boldsymbol{\xi}}_{b-1}^\top \widehat{\mathbf{Q}}_{b-1}\widehat{\boldsymbol{\xi}}_b \\
=& \frac{N_{b-1}}{N_b}\widehat{\boldsymbol{\xi}}_{b-1}^\top \left\{ \frac{\widehat{\mathbf{Q}}_{b-1}}{N_{b-1}} - \frac{1}{N_{b-1}}\sum_{l=1}^{b-1} n_l \widehat{\mathbf{H}}_h(\boldsymbol{\beta}^{0h};\mathcal{D}_l) \right\} \widehat{\boldsymbol{\xi}}_b \\
&+ \frac{N_{b-1}}{N_b}\widehat{\boldsymbol{\xi}}_{b-1}^\top \left\{ \frac{1}{N_{b-1}}\sum_{l=1}^{b-1} n_l \widehat{\mathbf{H}}_h(\boldsymbol{\beta}^{0h};\mathcal{D}_l) - \widetilde{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}) \right\} \widehat{\boldsymbol{\xi}}_b \\
&+ \frac{N_{b-1}}{N_b}\widehat{\boldsymbol{\xi}}_{b-1}^\top \left\{ \widetilde{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}) \right\} \widehat{\boldsymbol{\xi}}_b \\
:=& \frac{N_{b-1}}{N_b}\widehat{\boldsymbol{\xi}}_{b-1}^\top (\boldsymbol{K}_1 + \boldsymbol{K}_2)\widehat{\boldsymbol{\xi}}_b + \frac{N_{b-1}}{N_b}\widehat{\boldsymbol{\xi}}_{b-1}^\top \left\{ \widetilde{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}) \right\} \widehat{\boldsymbol{\xi}}_b.
\end{aligned}$$

By Lemma 7 and 1 we know that

$$\begin{aligned}
\|\boldsymbol{K}_1\|_{\max} \leq& \frac{C}{N_{b-1}}\sum_{l=1}^{b-1}\frac{n_l}{\sqrt{N_l}}\frac{s_0\log^2(p)}{h^2} \leq C'\frac{s_0\log^2(p)}{\sqrt{N_{b-1}}h}, \\
\|\boldsymbol{K}_2\|_{\max} \leq& C\sqrt{\frac{\log(p)}{h^2 N_{b-1}}}.
\end{aligned}$$

By the fact that $\widehat{\boldsymbol{\xi}}_1, \widehat{\boldsymbol{\xi}}_2, ..., \widehat{\boldsymbol{\xi}}_{b-1}$ all satisfy the ℓ_1-norm and ℓ_2-norm error bounds, given $\lambda_b\sqrt{N_b} \asymp \lambda_{b-1}\sqrt{N_{b-1}}$, we have

$$\begin{aligned}
|D_3| \leq& C c_{b-1}^{(1)} \left\{ \frac{s_0^2\log^2(p)}{\sqrt{N_{b-1}}h} + \sqrt{\frac{s_0^2\log(p)}{h^2 N_{b-1}}} \right\} \lambda_b \|\widehat{\boldsymbol{\xi}}_b\|_1 \\
&+ \frac{1}{N_b}\sum_{l=1}^{b-1} n_l \tau_{\max}\{\mathbf{H}_h(\boldsymbol{\beta}^{0h})\} c_{b-1}^{(2)}\sqrt{s_0}\lambda_b \|\widehat{\boldsymbol{\xi}}_b\|_2.
\end{aligned}$$

Under the condition $s_0\log(p)/h = o(N_{b-1}^{1/4})$, we know that $|D_3|$ is bounded by $\lambda_b\|\widehat{\boldsymbol{\xi}}_b\|_1/3 + M_1 c_{b-1}^{(2)}\sqrt{s_0}\lambda_b\|\widehat{\boldsymbol{\xi}}_b\|_2$ and correspondingly

$$\langle \nabla\mathcal{L}_b(\widehat{\boldsymbol{\beta}}_h^b) - \nabla\mathcal{L}_b(\boldsymbol{\beta}^{0h}), \widehat{\boldsymbol{\beta}}_h^b - \boldsymbol{\beta}^{0h})\rangle \leq \lambda_b\sqrt{s_0}\{2 + M_1 c_{b-1}^{(2)}\}\|\widehat{\boldsymbol{\xi}}_b\|_2.$$

Likewise, we could derive similar cone-type condition for $\widehat{\boldsymbol{\xi}}_b$ and prove by contradiction that $\|\widehat{\boldsymbol{\xi}}_b\|_2 \leq 2(2 + M_1 c_{b-1}^{(2)})\lambda_b\sqrt{s_0}/M_2$. Then both the ℓ_1-norm and ℓ_2-norm error bounds for $\widehat{\boldsymbol{\xi}}_b$ are demonstrated. This completes the proof.

□

Lemma 8. *Suppose the conditions in Theorem 1 and Assumption 2 hold. Then we have with probability at least* $1 - p^{-C_3}$,

$$\|\boldsymbol{\Theta}_0^h \widehat{\mathbf{Q}}_b/N_b - \mathbf{I}\|_{\max} \leq C_0 C_1 \omega_{N_b}, \quad \|\widehat{\boldsymbol{\Theta}}_b \widehat{\mathbf{Q}}_b/N_b - \mathbf{I}\|_{\max} \leq C_0 C_2 \omega_{N_b}, \quad \|\widehat{\boldsymbol{\Theta}}_b\|_{L_1} \leq C_0 \tag{S3.62}$$

for some positive constants C_1, C_2, C_3. *Here* $\omega_{N_b} = \{\sqrt{\log(p)/(N_b h^2)} + s_0 \log^2(p)/(\sqrt{N_{b-1}} h^2)\}$. *Moreover, we have*

$$\|\widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0^h\|_{L_1} \leq 14 C_r (\omega_{N_b} \|\boldsymbol{\Theta}_0^h\|_{L_1})^{1-q} \leq 14 C_r C_0^{1-q} \omega_{N_b}^{1-q}. \tag{S3.63}$$

Proof of Lemma 8. By the Triangle inequality and Lemma 6, we have

$$\begin{aligned}
&\|\widehat{\mathbf{Q}}_b/N_b - \mathbf{H}_h(\boldsymbol{\beta}^{0h})\|_{\max} \\
\leq &\left\| \widehat{\mathbf{Q}}_b/N_b - \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_l) \right\|_{\max} + \left\| \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\mathbf{H}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_l) - \mathbf{H}_h(\boldsymbol{\beta}^{0h}) \right\|_{\max} \\
\leq &C_1 \omega_{N_b},
\end{aligned}$$

with probability $1 - p^{-C'}$. Then, following the same logic as the proof of Lemma 3, (S3.62) can be demonstrated. Moreover, replacing $\boldsymbol{\Theta}_0$ with $\boldsymbol{\Theta}_0^h$ in the proof of Lemma 4, we directly prove (S3.63) under Assumption 2. □

Theorem 2. *Suppose that conditions in Theorem 1 and Lemma 8 hold. Then, for sufficiently large N_b, under the condition $C_r C_0^{1-q} s_0^{2-q} \log^{5/2-2q}(p)/h^{2-2q} = o(N_b^{1/2-q/2})$, we have*

$$\sqrt{N_b}\big(\widehat{\boldsymbol{\beta}}_h^{b,de} - \boldsymbol{\beta}^{0h}\big) = \boldsymbol{M}_b + \boldsymbol{R}_b,$$
$$\boldsymbol{M}_b = -\frac{1}{\sqrt{N_b}}\boldsymbol{\Theta}_0^h \sum_{l=1}^{b} n_l \widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_l), \quad \|\boldsymbol{R}_b\|_\infty = o_p(1).$$

Proof of Theorem 2. Using the same decomposition as in the proof of Theorem 3, we derive

$$\widehat{\boldsymbol{\beta}}_h^{b,de} - \boldsymbol{\beta}^{0h} = -\frac{1}{N_b}\sum_{l=1}^{b} \boldsymbol{\Theta}_0^h n_l \widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_l) + \boldsymbol{L}_1 + \boldsymbol{L}_2 + \boldsymbol{L}_3,$$

with

$$\boldsymbol{L}_1 = \frac{1}{N_b}\sum_{l=1}^{b} \left(\widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0^h\right) n_l \widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_l),$$
$$\boldsymbol{L}_2 = \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Theta}}_b n_l \left\{\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_h^l; \mathcal{D}_l) - \widehat{\mathbf{H}}_h(\widetilde{\boldsymbol{\beta}}_l; \mathcal{D}_l)\right\}(\widehat{\boldsymbol{\beta}}_h^l - \boldsymbol{\beta}^{0h}),$$
$$\boldsymbol{L}_3 = \left\{\mathbf{I} - \frac{1}{N_b}\sum_{l=1}^{b} \widehat{\boldsymbol{\Theta}}_b n_l \widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_h^l; \mathcal{D}_l)\right\}(\widehat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}^{0h}).$$

Here $\widetilde{\boldsymbol{\beta}}_l$ is an intermediate vector between $\widehat{\boldsymbol{\beta}}_h^l$ and $\boldsymbol{\beta}^{0h}$. For $\boldsymbol{L}_1$, we have

$$\begin{aligned}\|\boldsymbol{L}_1\|_\infty &\le \frac{1}{N_b}\sum_{l=1}^{b} n_l \left\|\widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^{0h}; \mathcal{D}_l)\right\|_\infty \left\|\widehat{\boldsymbol{\Theta}}_b - \boldsymbol{\Theta}_0^h\right\|_{L_1} \\ &\le C C_r C_0^{1-q} \omega_{N_b}^{1-q} \sqrt{\log(p)} \frac{1}{N_b}\sum_{l=1}^{b} \frac{n_l}{\sqrt{N_l}} \lesssim C_r C_0^{1-q} \omega_{N_b}^{1-q} \sqrt{\frac{\log(p)}{N_b}}.\end{aligned}$$

For $\boldsymbol{L}_2$, we have

$$\begin{aligned}\|\boldsymbol{L}_2\|_\infty &\le \frac{1}{N_b}\sum_{l=1}^{b} n_l \|\widehat{\boldsymbol{\Theta}}_b\|_{L_1} \left\|\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_h^l;\mathcal{D}_l) - \widehat{\mathbf{H}}_h(\widetilde{\boldsymbol{\beta}}_l;\mathcal{D}_l)\right\|_{\max} \|\widehat{\boldsymbol{\beta}}_h^l - \boldsymbol{\beta}^{0h}\|_1 \\ &\le \frac{2}{N_b}\sum_{l=1}^{b} n_l C_0 \left\|\widehat{\mathbf{H}}_h(\widehat{\boldsymbol{\beta}}_h^l;\mathcal{D}_l) - \mathbf{H}_h(\boldsymbol{\beta}^{0h})\right\|_{\max} \|\widehat{\boldsymbol{\beta}}_h^l - \boldsymbol{\beta}^{0h}\|_1 \\ &\le C_0\omega_{N_b} s_0\sqrt{\log(p)}\frac{2}{N_b}\sum_{l=1}^{b}\frac{n_l}{\sqrt{N_l}} \lesssim \frac{s_0^2\log^{5/2}(p)}{N_b h^2}.\end{aligned}$$

For $\boldsymbol{L}_3$, we have

$$\|\boldsymbol{L}_3\|_\infty \le \|\widehat{\boldsymbol{\Theta}}_b\widehat{\mathbf{Q}}_b/N_b - \mathbf{I}\|_{\max}\|\widehat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}^{0h}\|_1 \le c_b^{(1)} s_0\lambda_b C_2 C_0\omega_{N_b} \lesssim \frac{s_0^2\log^{5/2}(p)}{N_b h^2}.$$

Under the condition $C_r C_0^{1-q} s_0^{2-q}\log^{5/2-2q}(p)/h^{2-2q} = o(N_b^{1/2-q/2})$, we have $\|\boldsymbol{L}_m\|_\infty = o_p(1/\sqrt{N_b})$, $m = 1, 2, 3$. This completes the proof. □

Proposition 2. *Under the same conditions in Theorem 2, we have*

$$\begin{aligned}&\sqrt{N_b}(\widehat{\boldsymbol{\beta}}_h^{*,de} - \boldsymbol{\beta}_0) = \boldsymbol{M}_b^* + \boldsymbol{R}_b^*, \\ &\boldsymbol{M}_b^* = -\frac{1}{\sqrt{N_b}}\boldsymbol{\Theta}_0^h\sum_{l=1}^{b} n_l\widehat{\boldsymbol{S}}_h(\boldsymbol{\beta}^{0h};\mathcal{D}_l), \quad \|\boldsymbol{R}_b^*\|_\infty = o_p(1).\end{aligned}$$

The proof of proposition 2 follows directly from the proof of Theorem 2 and Proposition 1.

S4 Additional simulation results

S4.1 Figure for selection of c_h

We perform the procedure using the bandwidth $h_s = c_h * \{\log(p)/N_s\}^{1/4}$. Given different values of c_h, the ℓ_2-norm estimation errors are computed. The results are summarized in Figure 1. The left three panels exhibit results with $(N_b, b, p, n_l) = (1000, 10, 400, 100)$ and the right three panels correspond to $(N_b, b, p, n_l) = (2500, 10, 800, 250)$. In each panel, the values taken for c_h contain $\{0.25, 0.5, 1, 2, 2.5, 3, 4, 5\}$.

S4.2 Figures of estimation errors

We record the ℓ_2-norm estimation error of $\{\widehat{\boldsymbol{\beta}}_{h_l}^{(l)}, l = 1, \ldots, b\}$ and summarize the results in Figures 2-4. In each figure, the left panel shows results of $(N_b, b, p, n_l) = (2000, 20, 400, 100)$ and the right panel displays results under $(N_b, b, p, n_l) = (6000, 30, 1000, 200)$. The dot line, the square line, and the triangle line represent the error of online ℓ_1-DCSVM, the offline ℓ_1-DCSVM and the offline ℓ_1-SVM, respectively.

S4.3 High-dimensional cases

In addition, we examine high-dimensional cases with $N_b < p$. Specifically, consider $(N_b, b, p, n_l) = (200, 5, 400, 40)$ and $(N_b, b, p, n_l) = (800, 10, 1000, 80)$. We implement our online ℓ_1-DCSVM and compare it with the offline ℓ_1-DCSVM. The results under **Models 1-3** are summarized in Table 1. In each scenario, the test error of the

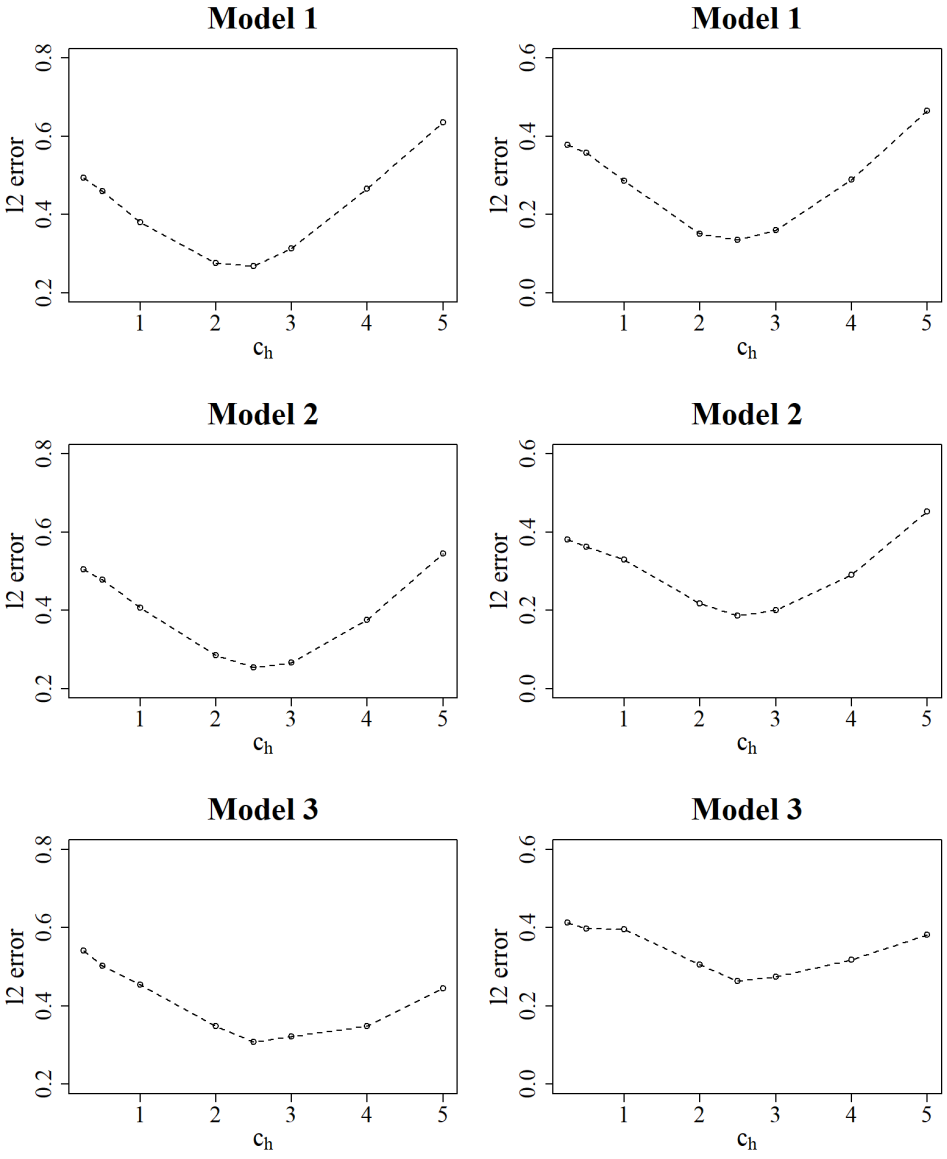

Figure 1: Estimation errors across different values of c_h.

online procedure is close to that of the offline method, implying robust prediction accuracy when p exceeds N_b. We also note that the online procedure retains more

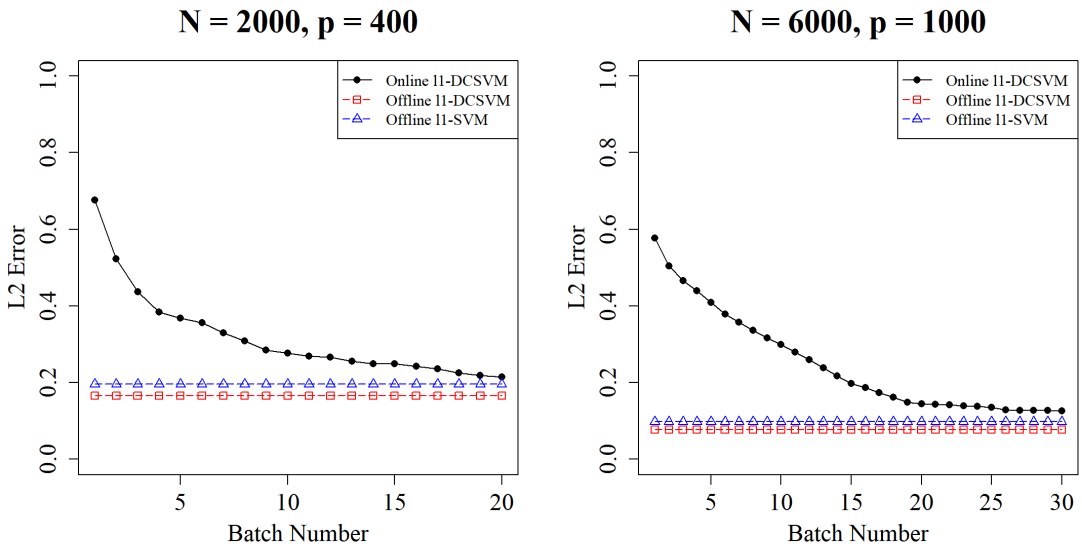

Figure 2: Estimation Error under **Model 1**.

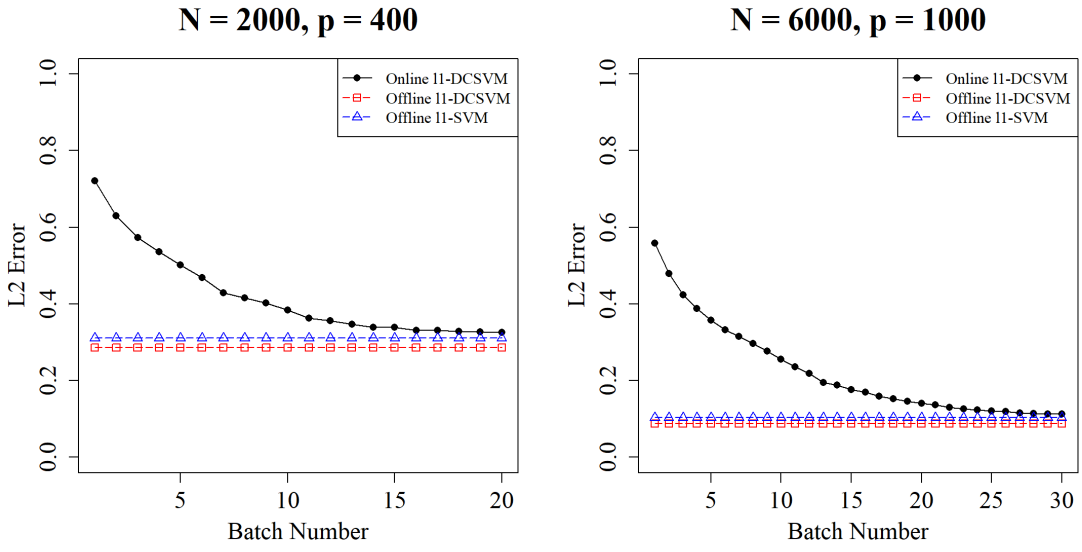

Figure 3: Estimation Error under **Model 2**.

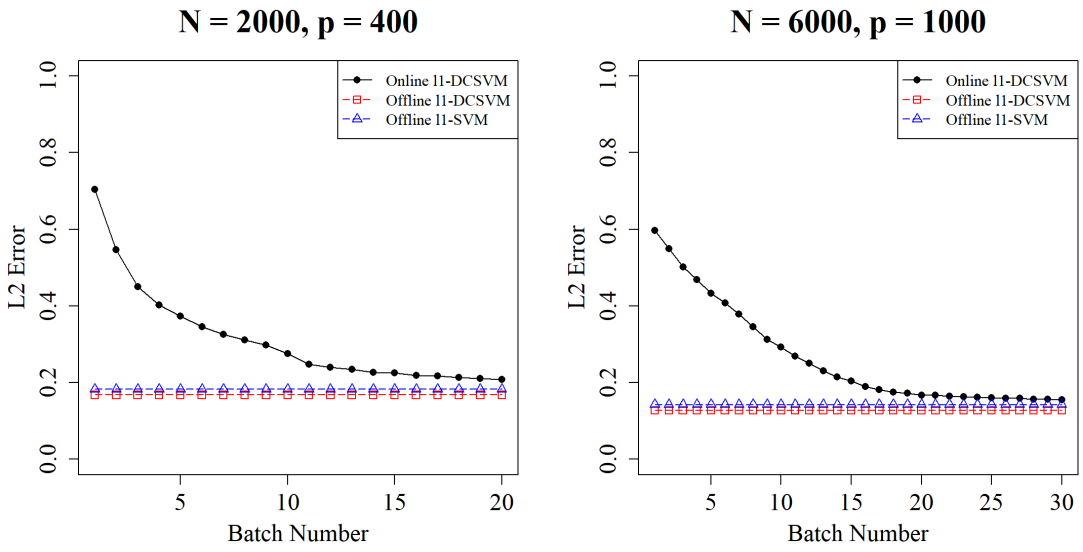

Figure 4: Estimation Error under **Model 3**.

noise variables and exhibits larger estimation errors than the offline benchmark. Compared with the offline procedure, the online procedure suffers from a lack of batch samples, making it difficult to maintain good estimation properties on a single batch.

S4.4 Sensitivity analysis on minimum batch size

According to referees' suggestion, we implement a sensitivity analysis to the minimum batch size. Fixing the total sample size N_b, we perform our online procedure under multiple combinations of (n_l, b). From Tables 2-3, we see that the noise and the run time are quite large when n_l is small. As n_l grows and b reduces, both the

Table 1: Estimation performance when $N_b < p$.

Model 1						
(N_b, b, p, n_l)	$(200, 5, 400, 40)$			$(800, 10, 1000, 80)$		
Method	l2-Err	Sig-Noi	TE(%)	l2-Err	Sig-Noi	TE(%)
on-l1-DCSVM	0.6445	$(5.00, 15.28)$	16.48	0.3851	$(5.00, 26.83)$	9.35
off-l1-DCSVM	0.4752	$(5.00, 4.96)$	14.26	0.3676	$(5.00, 5.50)$	9.27
Model 2						
(N_b, b, p, n_l)	$(200, 5, 400, 40)$			$(800, 10, 1000, 80)$		
Method	l2-Err	Sig-Noi	TE(%)	l2-Err	Sig-Noi	TE(%)
on-l1-DCSVM	0.6747	$(5.00, 10.44)$	24.37	0.4885	$(5.00, 25.34)$	15.96
off-l1-DCSVM	0.4450	$(5.00, 4.58)$	21.01	0.3563	$(4.94, 5.22)$	15.53
Model 3						
(N_b, b, p, n_l)	$(200, 5, 400, 40)$			$(800, 10, 1000, 80)$		
Method	l2-Err	Sig-Noi	TE(%)	l2-Err	Sig-Noi	TE(%)
on-l1-DCSVM	0.6920	$(5.00, 14.28)$	25.95	0.4843	$(5.00, 21.10)$	15.89
off-l1-DCSVM	0.5270	$(5.00, 5.24)$	24.01	0.3461	$(4.98, 5.06)$	15.70

estimation error and the run time decrease. This phenomenon suggests that practitioners can combine batches to improve the performance when the batch size is too small. Appropriate choices of minimum batch size are $n_l = 50$ for $N_b = 2000$ and $n_l = 80$ for $N_b = 6000$.

Table 2: Estimation results under multiple batch sizes ($N_b = 2000$).

n_l	b	l2-error	signal	noise	TE(%)	time
10	200	0.6025	5.00	169.14	13.12	152.31
16	125	0.8604	5.00	147.93	15.34	116.08
20	100	0.6076	5.00	81.48	11.33	69.91
25	80	0.5111	5.00	65.86	12.19	56.24
40	50	0.8035	5.00	32.96	14.84	43.96
50	40	0.3437	5.00	7.15	10.37	32.24
80	25	0.2984	5.00	8.54	9.41	23.97
100	20	0.3088	5.00	5.75	9.18	20.38
125	16	0.2516	5.00	6.28	8.85	19.12
200	10	0.1996	5.00	1.97	9.33	16.02
400	5	0.2385	5.00	1.32	10.38	16.49

Table 3: Estimation results under multiple batch sizes ($N_b = 6000$).

n_l	b	l2-error	signal	noise	TE(%)	time
20	300	0.6873	5	241.65	10.411	2004.858
30	200	0.5797	5	189.6	12.835	2030.822
40	150	0.4699	5	126.05	10.89	1260.226
50	120	0.3646	5	42.3	9.338	878.503
60	100	0.4056	5	34.65	9.775	674.2
80	75	0.3579	5	25.2	9.623	508.791
100	60	0.3982	5	13.7	13.162	468.822
120	50	0.2957	5	10.45	9.544	376.286
150	40	0.2761	5	6.97	9.719	299.465
200	30	0.2487	5	2.45	9.348	260.998
300	20	0.1814	5	2.5	8.338	203.336
600	10	0.1191	5	1.35	8.665	176.236

Bibliography

Cai, T., W. Liu, and X. Luo (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association 106*(494), 594–607.

Han, R., L. Luo, Y. Lin, and J. Huang (2021). Online debiased lasso for streaming data. *arXiv preprint arXiv:2106.05925*.

Luo, L., R. Han, Y. Lin, and J. Huang (2023). Statistical inference in high-dimensional generalized linear models with streaming data. *Electronic Journal of Statistics 17*, 3443–3471.

Samko, S. G., A. A. Kilbas, and O. I. Marichev (1993). *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach Science.

Tan, K. M., L. Wang, and W.-X. Zhou (2022). High-dimensional quantile regression: Convolution smoothing and concave regularization. *Journal of the Royal Statistical Society Series B: Statistical Methodology 84*(1), 205–233.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Process: With Application to Statistics*. Springer-Verlag.

Wang, B., L. Zhou, Y. Gu, and H. Zou (2023). Density-convoluted support vector

machines for high-dimensional classification. *IEEE transactions on Information Theory 69*(4), 2523–2536.