where $I\left\{1-\beta_0^*-\beta_1^*\frac{1-\alpha_0-\boldsymbol{x}_{-1}^{\mathrm{T}}\boldsymbol{\alpha}_{-1}-hy}{\alpha_1}-\boldsymbol{x}_{-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}^*\geq 0\right\}=I\{y\geq\frac{\alpha_1}{h}\quad(\boldsymbol{\alpha},\boldsymbol{\beta}^*,\boldsymbol{x}_{-1},0)\}$ and $\boldsymbol{x}^*$ denotes $\boldsymbol{x}$ with x_1 being replaced by $\frac{1-\alpha_0-\boldsymbol{x}_{-1}^{\mathrm{T}}\boldsymbol{\alpha}_{-1}-hy}{\alpha_1}$. According to Condition (C2), (C3) and (C4), note that

$$\begin{aligned}
&\int_{\mathbb{R}}(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}^*)^2\exp(t|\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}^*|K(\boldsymbol{x}^*,\widetilde{\boldsymbol{\theta}}))\left[H(y)-I\{y\geq\frac{\alpha_1}{h}\quad(\boldsymbol{\alpha},\boldsymbol{\beta}^*,\boldsymbol{x}_{-1},0)\}\right]^2\\
&\quad\times f\left(\frac{1-\alpha_0-\boldsymbol{x}_{-1}^{\mathrm{T}}\boldsymbol{\alpha}_{-1}-hy}{\alpha_1}|\boldsymbol{x}_{-1}\right)dy\\
\leq& C\left(\boldsymbol{v}_{-1}^{\mathrm{T}}\boldsymbol{x}_{-1}\right)^2\exp(t'|\boldsymbol{v}_{-1}^{\mathrm{T}}\boldsymbol{x}_{-1}|^2+t'(1+|\boldsymbol{\theta}_{-1}^{\mathrm{T}}\boldsymbol{x}_{-1}|+|\boldsymbol{\alpha}_{-1}^{\mathrm{T}}\boldsymbol{x}_{-1}|+|\boldsymbol{\beta}_{-1}^{*\mathrm{T}}\boldsymbol{x}_{-1}|)^2)\\
&\quad\times\left(1+\left|\frac{\alpha_1}{h}\quad(\boldsymbol{\alpha},\boldsymbol{\beta}^*,\boldsymbol{x}_{-1},0)\right|\right),
\end{aligned}$$

for some $t'>0$. Therefore, by (C3)

$$\int_{\mathbb{R}^p}(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x})^2\exp(t|\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}|K(\boldsymbol{x},\widetilde{\boldsymbol{\theta}}))\left[H\left(\frac{1-\boldsymbol{x}^{\mathrm{T}}\widetilde{\boldsymbol{\alpha}}}{h}\right)-I\{\epsilon\geq 0\}\right]^2 f(\boldsymbol{x})d\boldsymbol{x}\leq O(h+\|\widetilde{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\beta}}^*\|_2).$$

On the other hand, we can easily prove that

$$\begin{aligned}
&\int_{\mathbb{R}^p}(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x})^2\exp(t|\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}|K(\boldsymbol{x},\widetilde{\boldsymbol{\theta}}))\left[\frac{\epsilon}{h}H'\left(\frac{1-\boldsymbol{x}^{\mathrm{T}}\widetilde{\boldsymbol{\alpha}}}{h}\right)\right]^2 f(\boldsymbol{x})d\boldsymbol{x}\\
=&-\frac{h\beta_1^{*2}}{\alpha_1}\int_{\mathbb{R}^{p-1}}f_{-1}(\boldsymbol{x}_{-1})\int_{-1}^{1}(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}^*)^2\exp(t|\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}^*|K(\boldsymbol{x}^*,\widetilde{\boldsymbol{\theta}}))\left[\frac{(\boldsymbol{\alpha},\boldsymbol{\beta}^*,\boldsymbol{x}_{-1},y)}{h}H'(y)\right]^2\\
&\qquad\times f\left(\frac{1-\alpha_0-\boldsymbol{x}_{-1}^{\mathrm{T}}\boldsymbol{\alpha}_{-1}-hy}{\alpha_1}|\boldsymbol{x}_{-1}\right)dyd\boldsymbol{x}_{-1}\\
\leq& O(h+\|\widetilde{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\beta}}^*\|_2^2/h).
\end{aligned}$$

Now we complete the proof of the lemma. ■

A.2. Proofs of the Main Results

After introducing and proving the above three lemmas, we begin to prove Proposition 3 and 4, Theorem 1 and 2.

Proof of Proposition 3. Recall that $\epsilon_i=1-y_i\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}^*$. De ne $(\widetilde{\boldsymbol{\alpha}})=\widetilde{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\beta}}^*$, and

$$\begin{aligned}
\boldsymbol{C}_{n,h}(\widetilde{\boldsymbol{\alpha}})&=\boldsymbol{A}_{n,h}(\widetilde{\boldsymbol{\alpha}})-\frac{1}{n}\sum_{i=1}^{n}y_i\widetilde{\boldsymbol{X}}_iI\{\epsilon_i\geq 0\}\\
&=\frac{1}{n}\sum_{i=1}^{n}y_i\widetilde{\boldsymbol{X}}_i\left[H\left(\frac{1-y_i\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\widetilde{\boldsymbol{\alpha}}}{h}\right)-I\{\epsilon_i\geq 0\}+\frac{1-y_i\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}^*}{h}H'\left(\frac{1-y_i\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\widetilde{\boldsymbol{\alpha}}}{h}\right)\right].
\end{aligned}$$

Let $\boldsymbol{C}_{n,h}=\boldsymbol{C}_{n,h}(\widetilde{\boldsymbol{\beta}}^*)$. Note that $\|\boldsymbol{C}_{n,h}\|_2=\sup_{\tilde{\boldsymbol{v}}\in\mathbb{R}^{(p+1)},\|\tilde{\boldsymbol{v}}\|_2=1}|\boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}_{n,h}|$.

Let $S_{1/2}^p$ be a 1/2 net of the unit sphere S^p in the Euclidean distance in $\mathbb{R}^{p+1}$. According to the proof of Lemma 3 in Cai et al. (2010), we have d_{p+1}:=Card$(S_{1/2}^p)\leq 5^{p+1}$. Let

$\boldsymbol{v}_1, ..., \boldsymbol{v}_{d_{p+1}}$ be the centers of the d_{p+1} elements in the net. Therefore for any $\boldsymbol{v}$ in S^p, we have $\|\boldsymbol{v}-\boldsymbol{v}_j\|_2 \leq 1/2$ for some j. Therefore, $\|\boldsymbol{C}_{n,h}\|_2 \leq 2\sup_{j\leq d_{p+1}} |\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{C}_{n,h}|$. It is easy to see that for any $M > 0$, there exists a set of points in $\mathbb{R}^{p+1}$, $\{\widetilde{\boldsymbol{\alpha}}_k, 1 \leq k \leq s_{p+1}\}$ with $s_{p+1} \leq n^{M(p+1)}$, such that for any $\widetilde{\boldsymbol{\alpha}}$ in the ball $\|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*\|_2 \leq a_n$, we have $\|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\alpha}}_k\|_2 \leq 2\sqrt{p+1}a_n/n^M$ for some $1 \leq k \leq s_{p+1}$ and $\|\widetilde{\boldsymbol{\alpha}}_k - \widetilde{\boldsymbol{\beta}}^*\|_2 \leq a_n$.

Define

$$C_{n,h,j}(\widetilde{\boldsymbol{\alpha}}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{v}_j^{\mathrm{T}} y_i \widetilde{\boldsymbol{X}}_i \left[H\left(\frac{1 - y_i\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\widetilde{\boldsymbol{\alpha}}}{h}\right) - I\{\epsilon_i \geq 0\} + \frac{\epsilon_i}{h} H'\left(\frac{1 - y_i\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\widetilde{\boldsymbol{\alpha}}}{h}\right)\right]$$
$$\triangleq \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{v}_j^{\mathrm{T}} y_i \widetilde{\boldsymbol{X}}_i H_{h,i}(\widetilde{\boldsymbol{\alpha}}).$$

According to the proof of Proposition 4.1 in Chen et al. (2018), it is enough to show that

$$\sup_j \sup_k |C_{n,h,j}(\widetilde{\boldsymbol{\alpha}}_k)| = O_{\mathbb{P}}\left(\sqrt{\frac{ph\log n}{n}} + a_n^2 + h^2\right).$$

Since H and $xH'(x)$ are bounded, it is easy to see that

$$|H_{h,i}(\widetilde{\boldsymbol{\alpha}})| \leq C(1 + |\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}(\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*)|/\|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*\|_2) =: K(\widetilde{\boldsymbol{X}}_i, \widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}^*).$$

By Lemma 3, we have for some $t > 0$,

$$\mathbb{E}(\boldsymbol{v}_j^{\mathrm{T}} y_i \widetilde{\boldsymbol{X}}_i H_{h,i}(\widetilde{\boldsymbol{\alpha}}))^2 \exp(t|\boldsymbol{v}_j^{\mathrm{T}}\widetilde{\boldsymbol{X}}_i| K(\widetilde{\boldsymbol{X}}_i, \widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}^*)) \leq Ch(1 + \|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*\|_2/h + \|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*\|_2^2/h^2).$$

By $\sqrt{p\log n} = o(\sqrt{nh})$ and Lemma 1 in Cai and Liu (2011), we can get for any $\gamma > 0$, there exists a constant C such that

$$\sup_j \sup_k \mathbb{P}\left(|C_{n,h,j}(\widetilde{\boldsymbol{\alpha}}_k) - \mathbb{E}C_{n,h,j}(\widetilde{\boldsymbol{\alpha}}_k)| \geq C\sqrt{\frac{ph\log n}{n}}\right) = O(n^{-\gamma p}).$$

The remaining work is to give a bound for $\mathbb{E}C_{n,h,j}(\widetilde{\boldsymbol{\alpha}}_k)$. According to Lemma 1, we know that $\mathbb{E}C_{n,h,j}(\widetilde{\boldsymbol{\alpha}}_k) = O(h^2 + \|\widetilde{\boldsymbol{\alpha}}_k - \widetilde{\boldsymbol{\beta}}^*\|_2^2)$. Hence, $\sup_j \sup_k |\mathbb{E}C_{n,h,j}(\widetilde{\boldsymbol{\alpha}}_k)| = O(h^2 + \|\widetilde{\boldsymbol{\alpha}}_k - \widetilde{\boldsymbol{\beta}}^*\|_2^2)$. Combining with the above analysis, the proof is completed. ■

Proof of Proposition 4. For simplicity, denote $\boldsymbol{D}(\widetilde{\boldsymbol{\beta}}^*)$ by $\boldsymbol{D}$. According to the proof of Lemma 3 in Cai et al. (2010), for $\boldsymbol{D}_{n,h}$ we have

$$\|\boldsymbol{D}_{n,h} - \boldsymbol{D}\| \leq 10 \sup_{j\leq b_{p+1}} |\boldsymbol{v}_j^{\mathrm{T}}(\boldsymbol{D}_{n,h} - \boldsymbol{D})\boldsymbol{v}_j|.$$

where $\boldsymbol{v}_j$, $1 \leq j \leq b_{p+1}$, are some non-random vectors with $\|\boldsymbol{v}_j\|_2 = 1$ and $b_{p+1} \leq 5^{p+1}$. Define

$$D_{n,h,j}(\widetilde{\boldsymbol{\alpha}}) = \frac{1}{nh}\sum_{i=1}^{n} (\boldsymbol{v}_j^{\mathrm{T}}\widetilde{\boldsymbol{X}}_i)^2 H'\left(\frac{1 - y_i\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\widetilde{\boldsymbol{\alpha}}}{h}\right).$$

When $\|\widetilde{\boldsymbol{\beta}}_0 - \widetilde{\boldsymbol{\beta}}^*\|_2 \le a_n$, then

$$\sup_{j \le b_{p+1}} |\boldsymbol{v}_j^{\mathrm{T}}(\boldsymbol{D}_{n,h} - \boldsymbol{D})\boldsymbol{v}_j| \le \sup_{j \le b_{p+1}} \sup_{\|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*\|_2 \le a_n} |D_{n,h,j}(\widetilde{\boldsymbol{\alpha}}) - \boldsymbol{v}_j^{\mathrm{T}} \boldsymbol{D} \boldsymbol{v}_j|.$$

As the proof of Lemma 3, we obtain that

$$\mathbb{E}\left[\boldsymbol{v}^{\mathrm{T}} \widetilde{\boldsymbol{x}}_i H'\left(\frac{\epsilon_i - y_i \widetilde{\boldsymbol{x}}_i^{\mathrm{T}} \quad (\widetilde{\boldsymbol{\alpha}})}{h}\right)\right]^2 = O(h).$$

According to the proof of Proposition 4.2 in Chen et al. (2018), $D_{n,h,j}$ satisfies

$$\sup_j \sup_k |D_{n,h,j}(\widetilde{\boldsymbol{\alpha}}_k) - \mathbb{E} D_{n,h,j}(\widetilde{\boldsymbol{\alpha}}_k)| = O_{\mathbb{P}}\left(\sqrt{\frac{p \log n}{nh}}\right).$$

The remaining work is to give a bound for $\mathbb{E} D_{n,h,j}(\widetilde{\boldsymbol{\alpha}}_k) - \boldsymbol{v}_j^{\mathrm{T}} \boldsymbol{D} \boldsymbol{v}_j$. From Lemma 2, we obtain that

$$\mathbb{E} D_{n,h,j}(\widetilde{\boldsymbol{\alpha}}) - \boldsymbol{v}_j^{\mathrm{T}} \boldsymbol{D} \boldsymbol{v}_j = O(h + \|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*\|_2).$$

Hence, $\sup_j \sup_k |\mathbb{E} D_{n,h,j}(\widetilde{\boldsymbol{\alpha}}_k) - \boldsymbol{v}_j^{\mathrm{T}} \boldsymbol{D} \boldsymbol{v}_j| = O(h + \|\widetilde{\boldsymbol{\alpha}}_k - \widetilde{\boldsymbol{\beta}}^*\|_2)$. Combining with the above analysis, the proof of the proposition is completed. ■

Proof of Theorem 1 and 2. We first assume that $\|\widetilde{\boldsymbol{\beta}}_0 - \widetilde{\boldsymbol{\beta}}^*\| = O_{\mathbb{P}}(a_n)$ with $a_n = o(1)$ and $a_n = O(h)$. For independent random vectors $\{(y_i, \widetilde{\boldsymbol{X}}_i), i = 1, ..., n\}$ with $\sup_j \mathbb{E}|X_j|^3 = O(1)$, we can see that

$$\|\frac{1}{n}\sum_{i=1}^n y_i \widetilde{\boldsymbol{X}}_i I(\epsilon_i \ge 0)\|_2 = O_{\mathbb{P}}(\sqrt{p/n}).$$

Note that $\|\widetilde{\boldsymbol{\beta}}_0 - \widetilde{\boldsymbol{\beta}}^*\| = o_{\mathbb{P}}(1)$ and $\|\widetilde{\boldsymbol{\beta}}^*\|_2 \le C$. From the above result and Proposition 3 and 4 we know that for the estimator $\widetilde{\boldsymbol{\beta}}$ and the true parameter $\widetilde{\boldsymbol{\beta}}^*$,

$$\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}^* = \boldsymbol{D}(\widetilde{\boldsymbol{\beta}}^*)^{-1}\left(\frac{1}{n}\sum_{i=1}^n y_i \widetilde{\boldsymbol{X}}_i I\{\epsilon_i \ge 0\} - \lambda\begin{pmatrix}0\\ \boldsymbol{\beta}^*\end{pmatrix} + \lambda\begin{pmatrix}0\\ \boldsymbol{\beta}^* - \boldsymbol{\beta}_0\end{pmatrix}\right) + \boldsymbol{r}_n.$$

with

$$\|\boldsymbol{r}_n\|_2 = O_{\mathbb{P}}\left(\sqrt{\frac{p^2 \log n}{n^2 h}} + \sqrt{\frac{p \log n}{nh}}\lambda + \sqrt{\frac{ph \log n}{n}} + \lambda h + h^2\right).$$

Since $\lambda \le h$ and $\|\widetilde{\boldsymbol{\beta}}_0 - \widetilde{\boldsymbol{\beta}}^*\| = O_{\mathbb{P}}(a_n)$, we have

$$\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}^* = \boldsymbol{D}(\widetilde{\boldsymbol{\beta}}^*)^{-1}\left(\frac{1}{n}\sum_{i=1}^n y_i \widetilde{\boldsymbol{X}}_i I\{\epsilon_i \ge 0\} - \lambda\begin{pmatrix}0\\ \boldsymbol{\beta}^*\end{pmatrix}\right) + \boldsymbol{r}_n.$$

with

$$\|\boldsymbol{r}_n\|_2 = O_{\mathbb{P}}\left(\sqrt{\frac{p^2 \log n}{n^2 h}} + \sqrt{\frac{ph \log n}{n}} + h^2\right). \tag{28}$$

Note that $h_g \geq \sqrt{p/n}$, then $\sqrt{\frac{p^2 \log n}{n^2 h_g}} \leq \sqrt{\frac{p h_g \log n}{n}}$.

For $q = 1$, it is easy to see that Theorem 1 holds. Suppose the theorem holds for $q = g-1$ with $g \geq 2$. Note that $p = O(m/(\log n)^2)$ and $\lambda = O(1/\log n)$, then $h_{g-1} = O(1/\log n)$ and we have $\sqrt{p h_{(g-1)}(\log n)/n} = O(\sqrt{p/n})$. Then we have $a_n = \max\{\lambda^2, \sqrt{p/n}, (p/m)^{2^{g-2}}\} = O(h_g)$ for $q = g$ with initial estimator $\widehat{\beta}_0 = \widehat{\beta}^{(g-1)}$. Now we complete the proof of Theorem 1 by (28). Theorem 2 follows directly from Theorem 1 and the Lindeberg-Feller central limit theorem.

■

Proof of Theorem 3. To prove Theorem 3, we first introduce the following lemma, which shows that $\widehat{\boldsymbol{G}}(\widetilde{\boldsymbol{\beta}}^{(q)})$ is a consistent estimator of $\boldsymbol{G}(\widetilde{\boldsymbol{\beta}}^*)$.

Lemma 4 *Under the conditions of Theorem 2 and $a_n \to 0$, we have*

$$\sup_{\|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*\|_2 \leq a_n} \|\widehat{\boldsymbol{G}}(\widetilde{\boldsymbol{\alpha}}) - \boldsymbol{G}(\widetilde{\boldsymbol{\beta}}^*)\| = o_{\mathbb{P}}(1).$$

Proof of Lemma 4. Let $\boldsymbol{v}_j$, $1 \leq j \leq b_{p+1}$ be defined as in the proof Proposition 4. Define

$$G_{n,j}(\widetilde{\boldsymbol{\alpha}}) = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{v}_j^{\mathrm{T}} \widetilde{\boldsymbol{X}}_i)^2 I\{1 - Y \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} \widetilde{\boldsymbol{\alpha}} \geq 0\}.$$

By (C3) and Lemma 1 in Cai and Liu (2011), we can show that, for any $\gamma > 0$, there exists a constant $C > 0$ such that

$$\max_j \sup_{\|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*\|_2 \leq a_n} \mathbb{P}\left(\left| G_{n,j}(\widetilde{\boldsymbol{\alpha}}) - \mathbb{E} G_{n,j}(\widetilde{\boldsymbol{\alpha}}) \right| \geq C \sqrt{\frac{p \log n}{n}} \right) = O(n^{-\gamma p}).$$

Let $\widetilde{\boldsymbol{\alpha}}_k$, $1 \leq k \leq s_p$, be defined as in the proof of Proposition 3. Therefore

$$\max_{1 \leq j \leq b_{p+1}} \max_{1 \leq k \leq s_p} \left| G_{n,j}(\widetilde{\boldsymbol{\alpha}}_k) - \mathbb{E} G_{n,j}(\widetilde{\boldsymbol{\alpha}}_k) \right| = O_{\mathbb{P}}\Big(\sqrt{\frac{p \log n}{n}} \Big). \tag{29}$$

Put $t_M = 2\sqrt{p+1} a_n / n^M$. In the following, we show that

$$\mathbb{P}\left(\max_j \max_k \sup_{\|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\alpha}}_k\|_2 \leq t_M} \left| G_{n,j}(\widetilde{\boldsymbol{\alpha}}) - G_{n,j}(\widetilde{\boldsymbol{\alpha}}_k) \right| \geq C \sqrt{\frac{p \log n}{n}} \right) = o(1) \tag{30}$$

and

$$\max_j \max_k \sup_{\|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\alpha}}_k\|_2 \leq t_M} \left| \mathbb{E} G_{n,j}(\widetilde{\boldsymbol{\alpha}}) - \mathbb{E} G_{n,j}(\widetilde{\boldsymbol{\alpha}}_k) \right| \leq C \sqrt{\frac{p \log n}{n}}. \tag{31}$$

Note that for $\|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\alpha}}_k\|_2 \leq t_M$,

$$\left| I\{1 - y_i \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} \widetilde{\boldsymbol{\alpha}} \geq 0\} - I\{1 - y_i \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} \widetilde{\boldsymbol{\alpha}}_k \geq 0\} \right| \leq I\{-t_M \|\widetilde{\boldsymbol{X}}_i\|_2 \leq 1 - y_i \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} \widetilde{\boldsymbol{\alpha}}_k \leq t_M \|\widetilde{\boldsymbol{X}}_i\|_2\}$$

$$\leq I\{-t_M n \leq 1 - y_i \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} \widetilde{\boldsymbol{\alpha}}_k \leq t_M n\} + I\{\|\widetilde{\boldsymbol{X}}_i\|_2 \geq n\}.$$

By (C3), we have

$$\mathbb{P}(\max_{1\leq i\leq n} \|\widetilde{\boldsymbol{X}}_i\|_2 \geq n) = o(1). \tag{32}$$

Define

$$Z_{i,j,k}(\widetilde{\boldsymbol{\alpha}}_k) = (\boldsymbol{v}_j^{\mathrm{T}} \widetilde{\boldsymbol{X}}_i)^2 I\{-t_M n \leq 1 - y_i \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} \widetilde{\boldsymbol{\alpha}}_k \leq t_M n\}.$$

By (C2) and (C3), we have

$$\mathbb{E} Z_{i,j,k}(\widetilde{\boldsymbol{\alpha}}_k) \leq \sqrt{\mathbb{E}(\boldsymbol{v}_j^{\mathrm{T}} \widetilde{\boldsymbol{X}}_i)^4} \sqrt{\mathbb{P}(-t_M n \leq 1 - y_i \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} \widetilde{\boldsymbol{\alpha}}_k \leq t_M n)} = O(\sqrt{t_M n}) \tag{33}$$

and

$$\mathbb{E}(Z_{i,j,k}(\widetilde{\boldsymbol{\alpha}}_k))^2 \exp(t_0 Z_{i,j,k}(\widetilde{\boldsymbol{\alpha}}_k)) = O(\sqrt{t_M n}).$$

Note that

$$\sup_{\|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\alpha}}_k\|_2 \leq t_M} \left| G_{n,j}(\widetilde{\boldsymbol{\alpha}}) - G_{n,j}(\widetilde{\boldsymbol{\alpha}}_k) \right| I\{\max_{1\leq i\leq n} \|\widetilde{\boldsymbol{X}}_i\|_2 < n\} \leq \frac{1}{n} \sum_{i=1}^n Z_{i,j,k}(\widetilde{\boldsymbol{\alpha}}_k). \tag{34}$$

By (C3) and Lemma 1 in Cai and Liu (2011), for any $\gamma > 0$, there exists a constant $C > 0$ such that

$$\max_{j,k} \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n (Z_{i,j,k}(\widetilde{\boldsymbol{\alpha}}_k) - \mathbb{E} Z_{i,j,k}(\widetilde{\boldsymbol{\alpha}}_k)) \right| \geq C \sqrt{\frac{p \log n}{n}} \right) = O(n^{-\gamma p}).$$

By (32)-(34), we can see that (30) and (31) hold.

By (29), (30) and the definition of $\boldsymbol{v}_j$, we have

$$\sup_{\|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*\|_2 \leq a_n} \|\widehat{\boldsymbol{G}}(\widetilde{\boldsymbol{\alpha}}) - \boldsymbol{G}(\widetilde{\boldsymbol{\alpha}})\| = O_{\mathbb{P}}\Big(\sqrt{\frac{p \log n}{n}} \Big).$$

Moreover, we have

$$\left| I\{1 - y_i \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} \widetilde{\boldsymbol{\alpha}} \geq 0\} - I\{1 - y_i \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} \widetilde{\boldsymbol{\beta}}^* \geq 0\} \right| \leq I\{-\sqrt{a_n} \leq 1 - y_i \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} \widetilde{\boldsymbol{\alpha}}_k \leq \sqrt{a_n}\} + I\{|\widetilde{\boldsymbol{X}}_i^{\mathrm{T}} (\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*)| / \|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*\|_2 \geq a_n^{-1/2}\}.$$

By this inequality, (C2) and (C3), it is easy to show that

$$\mathbb{E}(\boldsymbol{v}^{\mathrm{T}} \widetilde{\boldsymbol{X}}_i)^2 I\{1 - Y \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} \widetilde{\boldsymbol{\alpha}} \geq 0\} - \mathbb{E}(\boldsymbol{v}^{\mathrm{T}} \widetilde{\boldsymbol{X}}_i)^2 I\{1 - Y \widetilde{\boldsymbol{X}}_i^{\mathrm{T}} \widetilde{\boldsymbol{\beta}}^* \geq 0\} = o(1)$$

uniformly in $\|\boldsymbol{v}\|_2 = 1$ and $\|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*\|_2 \leq a_n$. This implies that $\sup_{\|\widetilde{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\beta}}^*\|_2 \leq a_n} \|\boldsymbol{G}(\widetilde{\boldsymbol{\alpha}}) - \boldsymbol{G}(\widetilde{\boldsymbol{\beta}}^*)\| = o_{\mathbb{P}}(1)$. ■

Now we prove Theorem 3. Without loss of generality, we can assume that $\|\boldsymbol{v}\|_2 = 1$. By the consistency of $\widehat{\boldsymbol{D}}(\widetilde{\boldsymbol{\beta}}^{(q-1)})$ and $\widehat{\boldsymbol{G}}(\widetilde{\boldsymbol{\beta}}^{(q)})$ (see Lemma 4), we have

$$\widehat{\sigma}_{n,q} \to \sqrt{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{D}(\widetilde{\boldsymbol{\beta}}^*)^{-1}\boldsymbol{G}(\widetilde{\boldsymbol{\beta}}^*)\boldsymbol{D}(\widetilde{\boldsymbol{\beta}}^*)^{-1}\boldsymbol{v}}$$

as $n, p \to \infty$. This implies the theorem. ■

A.3. Proof of Auxiliary Results

Proof of Proposition 2. By Proposition 4, in the g-th iteration, we have

$$\left\| N^{-1}\sum_{k=1}^{N}\boldsymbol{V}_k - \boldsymbol{D}(\widetilde{\boldsymbol{\beta}}^*)\right\| = O_{\mathbb{P}}\left(\sqrt{\frac{p\log n}{nh_g}} + h_g\right).$$

Therefore, it suffices to show that

$$\left\|\widehat{\boldsymbol{V}}_1 - \boldsymbol{D}(\widetilde{\boldsymbol{\beta}}^*)\right\| = O_{\mathbb{P}}\left(m^{-\delta}\right)$$

for some $\delta > 0$. With the notation in the proof of Proposition 4, we have

$$\sup_j \sup_k |D_{m,h,j}(\widetilde{\boldsymbol{\alpha}}_k) - \mathbb{E}D_{m,h,j}(\widetilde{\boldsymbol{\alpha}}_k)| = O_{\mathbb{P}}\left(\sqrt{\frac{p\log n}{mh}}\right)$$

with $h = \sqrt{p/m}$. Also, $|\mathbb{E}D_{m,h,j}(\widetilde{\boldsymbol{\alpha}}_k) - \boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{D}\boldsymbol{v}_j| = O(h + \|\widetilde{\boldsymbol{\alpha}}_k - \widetilde{\boldsymbol{\beta}}^*\|_2)$ uniformly in j, k. This completes the proof as $p = O(m^\gamma)$ for some $0 < \gamma < 1$. ■

Proof of Claim 1. Let us define $\boldsymbol{\epsilon} = Y\boldsymbol{\epsilon} = (\epsilon^1, \ldots, \epsilon^p)$. It is easy to show it follows normal distribution $\mathcal{N}(0, \sigma^2\mathbf{I})$. By the construction of $\boldsymbol{X}$ (i.e., $\boldsymbol{X} = Y\mathbf{1} + \boldsymbol{\epsilon}$), we have $Y\boldsymbol{X} = Y^2\mathbf{1} + Y\boldsymbol{\epsilon} = \mathbf{1} + \boldsymbol{\epsilon}$ and $Y\widetilde{\boldsymbol{X}} = \binom{Y}{\mathbf{1}+\boldsymbol{\epsilon}}$. Recall that $S(\widetilde{\boldsymbol{\beta}}^*) = -\mathbb{E}[I\{1 - Y\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}^* \geq 0\}Y\widetilde{\boldsymbol{X}}]$. Therefore we have

$$S(\widetilde{\boldsymbol{\beta}}^*) = -\mathbb{E}\left[I\left\{1 - \frac{1}{a}(\mathbf{1}+\boldsymbol{\epsilon})^{\mathrm{T}}\mathbf{1} \geq 0\right\}\binom{Y}{\mathbf{1}+\boldsymbol{\epsilon}}\right] = -\mathbb{E}\left[I\{a \geq p + \mathbf{1}^{\mathrm{T}}\boldsymbol{\epsilon}\}\binom{Y}{\mathbf{1}+\boldsymbol{\epsilon}}\right].$$

In order to show that $S(\widetilde{\boldsymbol{\beta}}^*) = 0$, we only need to show that $\mathbb{E}\left[I\{a \geq p + \mathbf{1}^{\mathrm{T}}\boldsymbol{\epsilon}\}Y\right] = 0$ and $\mathbb{E}\left[I\{a \geq p + \mathbf{1}^{\mathrm{T}}\boldsymbol{\epsilon}\}(1+\epsilon^i)\right] = 0$ for $i = 1, \ldots, p$. The first equation holds because Y is independent of $\boldsymbol{\epsilon}$ and $\mathbb{E}[Y] = p_+ - p_- = 0$. To show the second equation, we note that for any $i, j \in \{1, \ldots, p\}$, we have $\mathbb{E}[I\{a \geq p + \mathbf{1}^{\mathrm{T}}\boldsymbol{\epsilon}\}(1+\epsilon^j)] = \mathbb{E}[I\{a \geq p + \mathbf{1}^{\mathrm{T}}\boldsymbol{\epsilon}\}(1+\epsilon^i)]$ by distributional symmetry of ϵ^i and ϵ^j. Therefore it is enough to show that

$$\sum_{j=1}^{p}\mathbb{E}[I\{a \geq p + \mathbf{1}^{\mathrm{T}}\boldsymbol{\epsilon}\}(1+\epsilon^j)] = \mathbb{E}[I\{\nu \leq a\}\nu] = 0,$$

where $\nu = p + \mathbf{1}^{\mathrm{T}}\boldsymbol{\epsilon}$. Recall that a satisfies $\int_{-\infty}^{a}\phi_1(x)x dx = 0$ where $\phi_1(x)$ is the p.d.f. of the distribution $\mathcal{N}(p, \sigma^2 p)$. Since ν follows the normal distribution $\mathcal{N}(p, \sigma^2 p)$, we have $E[I\{\nu \leq a\}\nu] = 0$. Therefore we have shown that $S(\widetilde{\boldsymbol{\beta}}^*) = 0$. By the convexity of the loss function and uniqueness of the minimizer, we have proved that $\widetilde{\boldsymbol{\beta}}^*$ is the true coefficient under the given setting. ■

References

R Raj Bahadur. A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37(3):577{580, 1966.

Moulinath Banerjee, Cecile Durot, and Bodhisattva Sen. Divide and conquer in non-standard problems and the super-e ciency phenomenon. *Ann. Statist.*, (To appear), 2018.

Peter L Bartlett, Michael I Jordan, and Jon D McAuli e. Convexity, classi cation, and risk bounds. *J. Amer. Statist. Assoc.*, 101(473):138{156, 2006.

Heather Battey, Jianqing Fan, Han Liu, Junwei Lu, and Ziwei Zhu. Distributed estimation and inference with statistical guarantees. *Ann. Statist.*, (To appear), 2018.

Gilles Blanchard, Olivier Bousquet, and Pascal Massart. Statistical performance of support vector machines. *Ann. Statist.*, 36(2):489{531, 2008.

T Tony Cai, Cun-Hui Zhang, and Harrison H Zhou. Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.*, 38(4):2118{2144, 2010.

Tony Cai and Weidong Liu. Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.*, 106(494):672{684, 2011.

Probal Chaudhuri. Nonparametric estimates of regression quantiles and their local bahadur representation. *Ann. Statist.*, 19(2):760{777, 1991.

Xi Chen, Weidong Liu, and Yichen Zhang. Quantile regression under memory constraint. *arXiv preprint arXiv:1810.08264*, 2018.

Xueying Chen and Min-ge Xie. A split-and-conquer approach for analysis of extraordinarily large data. *Statistica Sinica*, 24(4):1655{1684, October 2014.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3): 273{297, 1995.

Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge university press, 2000.

Jianqing Fan, Dong Wang, Kaizheng Wang, and Ziwei Zhu. Distributed estimation of principal eigenspaces. *arXiv preprint arXiv:1702.06488*, 2017.

Pedro A Forero, Alfonso Cano, and Georgios B Giannakis. Consensus-based distributed support vector machines. *Journal of Machine Learning Research*, 11:1663{1707, 2010.

Hans P Graf, Eric Cosatto, Leon Bottou, Igor Dourdanovic, and Vladimir Vapnik. Parallel support vector machines: the cascade SVM. In *Proceedings of the Advances in Neural Information Processing Systems*, 2005.

Magnus Rudolph Hestenes and Eduard Stiefel. *Methods of Conjugate Gradients for Solving Linear Systems*, volume 49. NBS Washington, DC, 1952.

Joel L Horowitz. Bootstrap methods for median regression models. *Econometrica*, pages 1327{1351, 1998.

Cho-Jui Hsieh, Si Si, and Inderjit Dhillon. A divide-and-conquer solver for kernel support vector machines. In *Proceedings of the International Conference on Machine Learning*, 2014.

Cheng Huang and Xiaoming Huo. A distributed one-step estimator. *arXiv preprint arXiv:1511.01443v2*, 2015.

Hanwen Huang. Asymptotic behavior of support vector machine for spiked population model. *Journal of Machine Learning Research*, 18(45):1{21, 2017.

Bo Jiang, Xuegong Zhang, and Tianxi Cai. Estimating the con dence interval for prediction errors of support vector machine classi ers. *Journal of Machine Learning Research*, 9: 521{540, 2008.

Michael I Jordan, Jason D Lee, and Yun Yang. Communication-e cient distributed statistical inference. *J. Amer. Statist. Assoc.*, (To appear), 2018.

Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: Journal of the Econometric Society*, pages 33{50, 1978.

Ja-Yong Koo, Yoonkyung Lee, Yuwon Kim, and Changyi Park. A bahadur representation of the linear support vector machine. *Journal of Machine Learning Research*, 9:1343{1368, 2008.

Jason D Lee, Qiang Liu, Yuekai Sun, and Jonathan E Taylor. Communication-e cient sparse regression. *Journal of Machine Learning Research*, 18(5):1{30, 2017.

Sangkyun Lee and Stephen J Wright. Approximate stochastic subgradient estimation training for support vector machines. *arXiv preprint arXiv:1111.0432*, 2011.

Runze Li, Dennis KJ Lin, and Bing Li. Statistical inference in massive data sets. *Appl. Stoch. Model Bus.*, 29(5):399{409, 2013.

Heng Lian and Zengyan Fan. Divide-and-conquer for debiased ℓ_1-norm support vector machine in ultra-high dimensions. *Journal of Machine Learning Research*, 18(1):6691{ 6716, 2017.

Yi Lin. Some asymptotic properties of the support vector machine. Technical report, University of Wisconsin-Madison, 1999.

Yufeng Liu, Hao Helen Zhang, Cheolwoo Park, and Jeongyoun Ahn. Support vector machines with adaptive lq penalty. *Computational Statistics & Data Analysis*, 51(12):6380{ 6394, 2007.

Lei Pang, Wenbin Lu, and Huixia Judy Wang. Variance estimation in censored quantile regression via induced smoothing. *Computational Statistics & Data Analysis*, 56(4):785{ 796, 2012.

Bo Peng, Lan Wang, and Yichao Wu. An error bound for ℓ_1-norm support vector machine coe cients in ultra-high dimension. *Journal of Machine Learning Research*, 17(236):1{26, 2016.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems*, 2008.

Bernhard Scholkopf and Alexander J Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT press, 2002.

Ohad Shamir, Nati Srebro, and Tong Zhang. Communication-e cient distributed optimization using an approximate newton-type method. In *Proceedings of the International Conference on Machine Learning*, 2014.

Chengchun Shi, Wenbin Lu, and Rui Song. A massive data framework for m-estimators with cubic-rate. *J. Amer. Statist. Assoc.*, (To appear), 2017.

Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Springer Science & Business Media, 2008.

Andrea Vedaldi and Andrew Zisserman. E cient additive kernels via explicit feature maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):480{492, 2012.

Stanislav Volgushev, Shih-Kang Chao, and Guang Cheng. Distributed inference for quantile regression processes. *arXiv preprint arXiv:1701.06088*, 2017.

Dongli Wang and Yan Zhou. Distributed support vector machines: an overview. In *IEEE Control and Decision Conference (CCDC)*, 2012.

Jialei Wang, Mladen Kolar, Nathan Srebro, and Tong Zhang. E cient distributed learning with sparsity. In *Proceedings of the International Conference on Machine Learning*, 2017.

Tong Zhang. Statistical behavior and consistency of classi cation methods based on convex risk minimization. *Ann. Statist.*, 32(1):56{85, 2004.

Xiang Zhang, Yichao Wu, Lan Wang, and Runze Li. A consistent information criterion for support vector machines in diverging model spaces. *Journal of Machine Learning Research*, 17(1):466{491, 2016a.

Xiang Zhang, Yichao Wu, Lan Wang, and Runze Li. Variable selection for support vector machines in moderately high dimensions. *J. Roy. Statist. Soc. Ser. B*, 78(1):53{76, 2016b.

Yuchen Zhang, John Duchi, and Martin Wainwright. Divide and conquer kernel ridge regression: a distributed algorithm with minimax optimal rates. *Journal of Machine Learning Research*, 16:3299{3340, 2015.

Tianqi Zhao, Guang Cheng, and Han Liu. A partially linear framework for massive heterogeneous data. *Ann. Statist.*, 44(4):1400{1437, 2016.

Kaihua Zhu, Hao Wang, Hongjie Bai, Jian Li, Zhihuan Qiu, Hang Cui, and Edward Y Chang. Parallelizing support vector machines on distributed computers. In *Proceedings of the Advances in Neural Information Processing Systems*, 2008.