Supplement to “High-Dimensional Quantile Regression: Convolution Smoothing and Concave Regularization”

Kean Ming Tan*, Lan Wang† and Wen-Xin Zhou‡

Abstract

This online supplementary material contains the proofs of all the theoretical results in the main text.

1. Appendix A provides the statistical theory when the error and covariates are independent.
2. Appendix B contains the proof of Proposition 4.1 regarding the smoothing bias.
3. Appendix C contains the proofs of Proposition 4.2 on local restricted strong convexity property, and Theorem 4.1 on ℓ_1-penalized SQR estimator.
4. Appendix D provides the proofs of Theorems 4.2 and 4.3, regarding the rate of convergence of the iteratively reweighted ℓ_1-penalized SQR estimator.
5. Appendix E contains the proofs of Theorems 4.4 and 4.5, regarding the strong oracle property of the iteratively reweighted ℓ_1-penalized SQR estimator.
6. Appendix F contains the proofs of the results in Section A.

All of the probabilistic bounds are non-asymptotic with explicit errors. The values of the constants involved are obtained with the goal of making the proof transparent, and may be improved by more careful calculations or under less general assumptions on the random covariates and error distribution.

A Regularized Smoothed Quantile Regression under Independence

In Section 2.2, we discussed the bias induced by smoothing. Recall that $\beta_h^* \in \mathrm{argmin}_{\beta\in\mathbb{R}^p} Q_h(\beta)$ is the population minimizer under the smoothed quantile objective, where $Q_h(\beta) = \mathbb{E}\{\widehat{Q}_h(\beta)\}$. Proposition 4.1 shows that under a Lipschitz condition on the conditional density $f_{\varepsilon|\boldsymbol{x}}(\cdot)$, the smoothing bias $\|\beta_h^* - \beta^*\|_2$ is of the order h^2. The assumed sparsity of $\beta^* \in \mathbb{R}^p$, however, is not necessarily inherited by β_h^*. Therefore, there is a statistical price to be paid by not having a sparse β_h^* after

*Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, USA. E-mail: keanming@umich.edu.

†Miami Herbert Business School, University of Miami, Coral Gables, FL 33146, USA. E-mail: lanwang@mbs.miami.edu.

‡Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, USA. E-mail: wez243@ucsd.edu.

smoothing. This results in stronger growth conditions on (s, p) in Theorem 4.3 and Theorem 4.5. Specifically, we have shown that with suitably chosen penalty level λ and bandwidth h,

$$\|\widehat{\beta}^{(\ell)} - \beta^*\|_2 \lesssim \sqrt{\frac{s + \log(p)}{n}} \quad \text{(weak oracle property)}$$

with probability at least $1 - C_1 p^{-1}$ as long as $\ell \gtrsim \log(\log p)$ and $n \gtrsim s^2 \log(p)$. In addition,

$$\widehat{\beta}^{(\ell)} = \widehat{\beta}^{\mathrm{ora}} \quad \text{(strong oracle property)}$$

with probability at least $1 - C_2(p^{-1} + n^{-1})$ as long as $\ell \gtrsim \log(s)$ and $n \gtrsim \max\{s^{8/3}, s^2 \log(p)\}$.

In the following, we show that under a stronger independence assumption between the random feature vector $\boldsymbol{x}$ and error variable ε, smoothing only introduces bias on the intercept, and therefore β_h^* preserves the sparsity of the true parameter β^*. This observation guarantees that we pay almost no price for estimating $\beta_-^* := (\beta_2^*, \ldots, \beta_p^*)^{\mathrm{T}}$ with the use of convolution smoothing. It is worth noticing that such an independence assumption is typically stringent in the context of quantile regression, whose main feature is the ability to capture heterogeneity in the set of important predictors at different quantile levels of the response distribution. The results of this section complement the existing theory for composite quantile regression in high dimensions (Zou and Yuan, 2008; Bradic, Fan and Wang, 2011).

(B1*) $\varepsilon \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^p$ are independent, and the function $m(\alpha) := \mathbb{E}\{\ell_h(\varepsilon - \alpha)\}$ has a unique minimizer b_h, where $\ell_h(\cdot)$ is defined in (2.5). The density function of ε, denoted by $f_\varepsilon(\cdot)$, satisfies $f_\varepsilon(0) > 0$ and $|f_\varepsilon(u) - f_\varepsilon(v)| \le l_0|u - v|$ for all $u, v \in \mathbb{R}$ and some $l_0 > 0$.

Proposition A.1. Assume Conditions (B1*) and (B2) hold. Then, β_h^* is the unique minimizer of $Q_h(\cdot)$ and satisfies $\beta_h^* = (\beta_1^* + b_h, \beta_-^{*\mathrm{T}})^{\mathrm{T}}$. Provided that $0 < h < f_0/(c_1 l_0)$, we have

$$|b_h| \le l_0 c_1 \kappa_2^{1/2} f_0^{-1} h^2, \tag{A.1}$$

where $c_1 = \kappa_1 + \kappa_2^{1/2}$ and $f_0 = f_\varepsilon(0)$.

Remark A.1. According to (4.1), the first and second derivatives of $\alpha \mapsto m(\alpha)$ are

$$m'(\alpha) = \int_{-\infty}^{\infty} K(u) F_\varepsilon(\alpha - hu)\,\mathrm{d}u - \tau \quad \text{and} \quad m''(\alpha) = \mathbb{E}\{K_h(\alpha - \varepsilon)\} = \int_{-\infty}^{\infty} K(u) f_\varepsilon(\alpha - hu)\,\mathrm{d}u,$$

where F_ε and f_ε denote, respectively, the distribution and density functions of ε. Moreover, note that $\lim_{\alpha\to\infty} m'(\alpha) = 1 - \tau$ and $\lim_{\alpha\to-\infty} m'(\alpha) = -\tau$. Provided that $F_\varepsilon(\cdot)$ is strictly increasing, there exits a unique b_h for which $m'(b_h) = 0$. In other words, b_h is the unique minimizer of $m(\cdot)$.

Under independence, our first result is on the weak oracle property, which is in parallel with Theorem 4.3.

Theorem A.1. Assume that Conditions (A1) and (B1*), (B2) and (B3) hold, and there exist $\alpha_1 > \alpha_0 > 0$ satisfying (4.16) with f_l replaced by $f_0 = f_\varepsilon(0)$. Let the penalty level λ and bandwidth h satisfy $\lambda \asymp \sigma_{\boldsymbol{x}} \sqrt{\log(p)/n}$ and $\sigma_{\boldsymbol{x}} f_0^{-1} \sqrt{s \log(p)/n} \lesssim h \lesssim f_0$. Under the beta-min condition $\|\beta_{\mathcal{S}}^*\|_{\min} \ge (\alpha_0 + \alpha_1)\lambda$ and sample size requirement $n \gtrsim s\log(p) + t$, the multi-step estimator $\widehat{\beta}^{(\ell)}$ with $\ell \gtrsim \lceil \log(\log p)/\log(1/\delta) \rceil$ satisfies, for any $t \ge 0$,

$$\|\widehat{\beta}^{(\ell)} - \beta_h^*\|_2 \lesssim f_0^{-1} \sqrt{\frac{s+t}{n}} \quad \text{and} \quad \|\widehat{\beta}^{(\ell)} - \beta_h^*\|_1 \lesssim f_0^{-1} s^{1/2} \sqrt{\frac{s+t}{n}}$$

with probability at least $1 - p^{-1} - e^{-t}$, where $\delta = \sqrt{4 + \{q'(\alpha_0)\}^2}/(\alpha_0 \kappa_l f_0 \gamma_p)$.

To establish the strong oracle property, we first refine Proposition 4.3 on the oracle estimator.

Proposition A.2. Assume Conditions (B1*) and (B1$'$)–(B3$'$) hold. For any $t \geq 0$, et the sample size n and the bandwidth $h = h_n$ be such that $n \gtrsim s + t$ and $\sqrt{(s+t)/n} \lesssim h \lesssim 1$. Then, the oracle estimator $\widehat{\beta}^{\rm ora}$ satisfies

$$\|\mathbf{S}^{1/2}(\widehat{\beta}^{\rm ora} - \beta^*_h)_{\mathcal{S}}\|_2 \lesssim f_0^{-1}\sqrt{\frac{s+t}{n}} \tag{A.2}$$

with probability at least $1 - 2e^{-t}$, where $\mathbf{S} := \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}} = \mathbb{E}(\boldsymbol{x}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}}^{\rm T})$. Moreover,

$$\left\|\mathbf{S}^{-1/2}\mathbf{D}_h(\widehat{\beta}^{\rm ora} - \beta^*_h)_{\mathcal{S}} + \mathbf{S}^{-1/2}\frac{1}{n}\sum_{i=1}^n\{\bar{K}((b_h - \varepsilon_i)/h) - \tau\}\boldsymbol{x}_{i,\mathcal{S}}\right\|_2 \lesssim \frac{s+t}{h^{1/2}n} \tag{A.3}$$

with probability at least $1-3e^{-t}$, where $\mathbf{D}_h = m''(b_h)\cdot\mathbf{S}$ with $m(\cdot)$ and b_h defined in Condition (B1*).

Finally, Theorem A.2 below relaxes the sample size scaling required for the strong oracle property given in Theorem 4.5.

Theorem A.2. Assume that Conditions (A1), (B1*) and (B1$'$)–(B3$'$) hold, and

$$\max_{j\in\mathcal{S}^{\rm c}}\|\mathbf{\Sigma}_{j\mathcal{S}}(\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}})^{-1}\|_1 \leq A_1 \tag{A.4}$$

for some $A_1 \geq 1$. For a prespecified $\delta \in (0,1)$, suppose there exist constants $\alpha_1 > \alpha_0$ satisfying (4.20) with $\kappa = \kappa_l f_\varepsilon(0)/2$, and the beta-min condition $\|\beta^*_{\mathcal{S}}\|_{\min} \geq (\alpha_0 + \alpha_1)\lambda$ with the penalty level $\lambda \asymp \sqrt{\log(p)/n}$. Then, with probability at least $1 - 2p^{-1} - 5n^{-1}$, $\widehat{\beta}^{(\ell)} = \widehat{\beta}^{\rm ora}$ for all $\ell \geq \lceil\log(s^{1/2}/\delta)/\log(1/\delta)\rceil$, provided that the bandwidth h and sample size n are subject to

$$\max\left\{\sqrt{\frac{s\log(p)}{n}}, \frac{s^2}{n\log(p)}\right\} \lesssim h \lesssim 1.$$

B Proof of Proposition 4.1 in Section 4.1

We derive an upper bound for $\|\beta^*_h - \beta^*\|_{\mathbf{\Sigma}}$ via a localized analysis exploited by Fan et al. (2018). Define the local vicinity $\Theta_h = \beta^* + \mathbb{B}_{\mathbf{\Sigma}}(h)$ of β^*. To begin with, it is unclear that whether $\beta^*_h \in \mathrm{argmin}_{\beta\in\mathbb{R}^p} Q_h(\beta)$ falls into this local region. Instead, we consider an intermediate vector $\beta^\dagger_h = (1-\eta)\beta^* + \eta\beta^*_h$, where $\eta = \sup\{u \in [0,1] : \beta^* + u(\beta^*_h - \beta^*) \in \Theta_h\}$, which is the large value between 0 and 1 such that the corresponding convex combination of β^* and β^*_h falls into Θ_h. If $\beta^*_h \notin \Theta_h$, then $\eta \in (0,1)$ and $\beta^\dagger_h$ falls onto the boundary of Θ_h, i.e., $\|\beta^\dagger_h - \beta^*\|_{\mathbf{\Sigma}} = h$; otherwise if $\beta^*_h \in \Theta_h$, $\eta = 1$ and hence $\beta^\dagger_h = \beta^*_h$.

By the convexity of $\beta \mapsto Q_h(\beta)$, the optimality of β^*_h, and Lemma F.2 in the supplementary material of Fan et al. (2018), we obtain that

$$\begin{aligned}0 &\leq \langle\nabla Q_h(\beta^\dagger_h) - \nabla Q_h(\beta^*), \beta^\dagger_h - \beta^*\rangle \\ &\leq \eta\cdot\langle\nabla Q_h(\beta^*_h) - \nabla Q_h(\beta^*), \beta^*_h - \beta^*\rangle = \langle-\nabla Q_h(\beta^*), \beta^\dagger_h - \beta^*\rangle.\end{aligned} \tag{B.1}$$

Applying the mean value theorem for vector-valued functions yields

$$\nabla Q_h(\beta^\dagger_h) - \nabla Q_h(\beta^*) = \int_0^1 \nabla^2 Q_h((1-t)\beta^* + t\beta^\dagger_h)\,\mathrm{d}t\,(\beta^\dagger_h - \beta^*), \tag{B.2}$$

where $\nabla^2 Q_h(\boldsymbol\beta) = \mathbb{E}\{K_h(\langle \boldsymbol x, \boldsymbol\beta - \boldsymbol\beta^*\rangle - \varepsilon)\boldsymbol x\boldsymbol x^{\mathrm T}\}$ for $\boldsymbol\beta \in \mathbb{R}^p$. With $\boldsymbol\delta := \boldsymbol\beta - \boldsymbol\beta^*$, note that

$$\mathbb{E}\{K_h(\boldsymbol x^{\mathrm T}\boldsymbol\delta - \varepsilon)|\boldsymbol x\} = \frac{1}{h}\int_{-\infty}^{\infty} K\Big(\frac{\boldsymbol x^{\mathrm T}\boldsymbol\delta - t}{h}\Big) f_{\varepsilon|\boldsymbol x}(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} K(u) f_{\varepsilon|\boldsymbol x}(\boldsymbol x^{\mathrm T}\boldsymbol\delta - hu)\,\mathrm{d}u.$$

By the Lipschitz continuity of $f_{\varepsilon|\boldsymbol x}(\cdot)$,

$$\mathbb{E}\{K_h(\boldsymbol x^{\mathrm T}\boldsymbol\delta - \varepsilon)|\boldsymbol x\} = f_{\varepsilon|\boldsymbol x}(0) + R_h(\boldsymbol\delta) \tag{B.3}$$

with $R_h(\boldsymbol\delta)$ satisfying $|R_h(\boldsymbol\delta)| \le l_0(|\boldsymbol x^{\mathrm T}\boldsymbol\delta| + \kappa_1 h)$. Substituting (B.3) into (B.1) and (B.2) yields

$$\begin{aligned}
&\langle \nabla Q_h(\boldsymbol\beta_h^\dagger) - \nabla Q_h(\boldsymbol\beta^*), \boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\rangle \\
&\ge \|\boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\|_{\mathbf J}^2 - \frac{l_0}{2}\mathbb{E}|\langle \boldsymbol x, \boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\rangle|^3 - l_0\kappa_1 h \cdot \|\boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\|_{\boldsymbol\Sigma}^2 \\
&\ge \|\boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\|_{\mathbf J}^2 - \frac{l_0}{2}\mu_3 \cdot \|\boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\|_{\boldsymbol\Sigma}^3 - l_0\kappa_1 h \cdot \|\boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\|_{\boldsymbol\Sigma}^2,
\end{aligned} \tag{B.4}$$

where $\mathbf J = \mathbb{E}\{f_{\varepsilon|\boldsymbol x}(0)\boldsymbol x\boldsymbol x^{\mathrm T}\}$.

On the other hand, we have

$$\langle -\nabla Q_h(\boldsymbol\beta^*), \boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\rangle \le \|\boldsymbol\Sigma^{-1/2}\nabla Q_h(\boldsymbol\beta^*)\|_2 \cdot \|\boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\|_{\boldsymbol\Sigma},$$

where $\nabla Q_h(\boldsymbol\beta^*) = \mathbb{E}\{\bar K(-\varepsilon/h) - \tau\}\boldsymbol x$. Using integration by parts and a Taylor series expansion yields

$$\begin{aligned}
\mathbb{E}\{\bar K(-\varepsilon/h)|\boldsymbol x\} &= \int_{-\infty}^{\infty} \bar K(-t/h)\,\mathrm{d}F_{\varepsilon|\boldsymbol x}(t) \\
&= -\frac{1}{h}\int_{-\infty}^{\infty} K(-t/h)F_{\varepsilon|\boldsymbol x}(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} K(u)F_{\varepsilon|\boldsymbol x}(-hu)\,\mathrm{d}u \\
&= \tau + \int_{-\infty}^{\infty} K(u)\int_0^{-hu}\{f_{\varepsilon|\boldsymbol x}(t) - f_{\varepsilon|\boldsymbol x}(0)\}\,\mathrm{d}t\,\mathrm{d}u,
\end{aligned}$$

from which it follows that $|\mathbb{E}\bar K(-\varepsilon/h) - \tau| \le \frac{l_0}{2}\kappa_2 h^2$. Consequently,

$$\|\boldsymbol\Sigma^{-1/2}\nabla Q_h(\boldsymbol\beta^*)\|_2 = \sup_{\boldsymbol u \in \mathbb{S}^{p-1}} \mathbb{E}\{\bar K(-\varepsilon/h) - \tau\}\langle \boldsymbol\Sigma^{-1/2}\boldsymbol x, \boldsymbol u\rangle \le \frac{l_0}{2}\kappa_2 h^2. \tag{B.5}$$

Putting together the pieces, we conclude that

$$\langle -\nabla Q_h(\boldsymbol\beta^*), \boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\rangle \le \frac{l_0}{2}\kappa_2 h^2 \cdot \|\boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\|_{\boldsymbol\Sigma}. \tag{B.6}$$

Recall that $f_{\varepsilon|\boldsymbol x}(0) \ge f_l > 0$ almost surely and $\boldsymbol\beta_h^\dagger \in \Theta_h$. Together, (B.1), (B.4) and (B.6) imply

$$f_l \cdot \|\boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\|_{\boldsymbol\Sigma}^2 \le (0.5\mu_3 + 0.5\kappa_2 + \kappa_1)l_0 h^2 \cdot \|\boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\|_{\boldsymbol\Sigma}.$$

Canceling $\|\boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\|_{\boldsymbol\Sigma}$ on both sides gives

$$\|\boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\|_{\boldsymbol\Sigma} \le \underbrace{(0.5\mu_3 + 0.5\kappa_2 + \kappa_1)}_{=:c_0}\frac{l_0 h}{f_l}h = \frac{c_0 l_0 h}{f_l}h.$$

Provided that $h < f_l/(c_0 l_0)$, $\boldsymbol\beta_h^\dagger$ falls in the interior of Θ_h, i.e., $\|\boldsymbol\beta_h^\dagger - \boldsymbol\beta^*\|_{\boldsymbol\Sigma} < h$. By the definition of $\boldsymbol\beta_h^\dagger$ in the beginning of the proof, we must have $\boldsymbol\beta_h^* \in \Theta_h$; otherwise if $\boldsymbol\beta_h^* \notin \Theta_h$, $\boldsymbol\beta_h^\dagger$ lies on the boundary of Θ_h, which leads to contradiction. Consequently, $\beta_h^* = \beta_h^\dagger$ satisfies the claimed bound (4.3). Moreover, by (B.4), $Q_h(\cdot)$ is strictly convex in a neighborhood of β_h^*, thus verifying the uniqueness claim.

Next, to investigate the leading term in the bias, define the remainder

$$\Delta_h = \Sigma^{-1/2}\{\nabla Q_h(\beta_h^*) - \nabla Q_h(\beta^*) - \mathbf{J}(\beta_h^* - \beta^*)\} = \Sigma^{-1/2}\mathbf{J}(\beta_h^* - \beta^*) - \Sigma^{-1/2}\nabla Q_h(\beta^*).$$

Once again, using the mean value theorem for vector-valued functions, we find that

$$\Delta_h = \left\{\Sigma^{-1/2}\int_0^1 \nabla^2 Q_h((1-t)\beta^* + t\beta_h^*)\,\mathrm{d}t\,\Sigma^{-1/2} - \mathbf{J}_0\right\}\Sigma^{1/2}(\beta_h^* - \beta^*), \tag{B.7}$$

where $\mathbf{J}_0 = \Sigma^{-1/2}\mathbf{J}\Sigma^{-1/2} = \mathbb{E}\{f_{\varepsilon|\boldsymbol{x}}(0)\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}\}$ and $\boldsymbol{z} = \Sigma^{-1/2}\boldsymbol{x}$. Under Conditions (B1) and (B2), we derive that

$$\begin{aligned}
&\left\|\Sigma^{-1/2}\int_0^1 \nabla^2 Q_h((1-t)\beta^* + t\beta_h^*)\,\mathrm{d}t\,\Sigma^{-1/2} - \mathbf{J}_0\right\|_2 \\
&= \left\|\mathbb{E}\int_0^1\int_{-\infty}^{\infty} K(u)\{f_{\varepsilon|\boldsymbol{x}}(t\langle\boldsymbol{x}, \beta_h^* - \beta^*\rangle - hu) - f_{\varepsilon|\boldsymbol{x}}(0)\}\,\mathrm{d}u\,\mathrm{d}t\,\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}\right\|_2 \\
&\le l_0 \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \mathbb{E}\int_0^1\int_{-\infty}^{\infty} K(u)(|t\langle\boldsymbol{x}, \beta_h^* - \beta^*\rangle| + h|u|)\,\mathrm{d}u\,\mathrm{d}t\,(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{u})^2 \\
&\le \frac{l_0}{2}\sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \mathbb{E}\{|\langle\boldsymbol{x}, \beta_h^* - \beta^*\rangle|(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{u})^2\} + l_0\kappa_1 h \\
&\le \frac{l_0}{2}\mu_3\|\beta_h^* - \beta^*\|_\Sigma + l_0\kappa_1 h.
\end{aligned}$$

This bound, together with (B.7), implies

$$\|\Delta_h\|_2 \le l_0(0.5\mu_3\|\beta_h^* - \beta^*\|_\Sigma + \kappa_1 h)\|\beta_h^* - \beta^*\|_\Sigma. \tag{B.8}$$

From the earlier bound (4.3), we see that $\|\Delta_h\|_2 \lesssim h^3$.

Turning to the gradient $\nabla Q_h(\beta^*) = \mathbb{E}\{\bar{K}(-\varepsilon/h) - \tau\}\boldsymbol{x}$, we apply a second-order Taylor series expansion to $F_{\varepsilon|\boldsymbol{x}}$ to conclude that

$$\begin{aligned}
\mathbb{E}\{\bar{K}(-\varepsilon/h)|\boldsymbol{x}\} - \tau &= \int_{-\infty}^{\infty} K(u)\int_0^{-hu}\{f_{\varepsilon|\boldsymbol{x}}(t) - f_{\varepsilon|\boldsymbol{x}}(0)\}\,\mathrm{d}t\,\mathrm{d}u \\
&= \frac{1}{2}\kappa_2 h^2 \cdot f'_{\varepsilon|\boldsymbol{x}}(0) + \int_{-\infty}^{\infty}\int_0^{-hu}\int_0^t K(u)\{f'_{\varepsilon|\boldsymbol{x}}(v) - f'_{\varepsilon|\boldsymbol{x}}(0)\}\,\mathrm{d}v\,\mathrm{d}t\,\mathrm{d}u.
\end{aligned}$$

Under the Lipschitz continuity assumption on $f'_{\varepsilon|\boldsymbol{x}}$, this further implies

$$\left\|\Sigma^{-1/2}\nabla Q_h(\beta^*) - \frac{1}{2}\kappa_2 h^2 \cdot \Sigma^{-1/2}\mathbb{E}\{f'_{\varepsilon|\boldsymbol{x}}(0)\boldsymbol{x}\}\right\|_2 \le \frac{l_1}{6}\kappa_3 h^3. \tag{B.9}$$

Finally, combining (B.8) and (B.9) proves (4.4). □

C Proofs of Results in Section 4.2

Recall from (4.10) that $\boldsymbol{w}_h^* = \boldsymbol{w}_h(\beta^*)$ and $\boldsymbol{w}_h(\beta) = \nabla\widehat{Q}_h(\beta) - \nabla Q_h(\beta)$, $\beta \in \mathbb{R}^p$. We start with the following two lemmas that will be needed in proving Proposition 4.2 and Theorem 4.1.

Lemma C.1. Let $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{x} \in \mathbb{R}^p$ be the standardized feature vector which is isotropic, i.e., $\mathbb{E}(\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}) = \mathbf{I}_p$. Under Condition (B3), the k-th ($k \geq 3$) absolute moments of all the one-dimensional marginals of $\boldsymbol{z}$ are uniformly bounded: $\mu_k := \sup_{\boldsymbol{u} \in \mathbb{S}^{p-1}} \mathbb{E}|\boldsymbol{z}^{\mathrm{T}}\boldsymbol{u}|^k \leq k! \upsilon_0^k$. In particular, $\mu_1 \leq \mu_2^{1/2} = 1$.

Lemma C.2. Assume Conditions (B1)–(B3) hold. Then, for any $t > 0$,

$$\|\boldsymbol{w}_h^*\|_\infty \leq \upsilon_0 \sigma_{\boldsymbol{x}} \left[\sqrt{\{\tau(1-\tau) + Ch^2\}\frac{2t}{n}} + \max(1-\tau, \tau)\frac{2t}{n} \right] \tag{C.1}$$

holds with probability at least $1 - 2pe^{-t}$, where $C = (\tau + 1)l_0\kappa_2$.

C.1 Proof of Proposition 4.2

Given $r, l > 0$, define the local cone-neighborhood of $\boldsymbol{\beta}^*$

$$\Theta = \Theta(r, l) = \{\boldsymbol{\beta} \in \mathbb{R}^p : \boldsymbol{\beta} - \boldsymbol{\beta}^* \in \mathbb{B}_{\boldsymbol{\Sigma}}(r) \cap \mathbb{C}_{\boldsymbol{\Sigma}}(l)\}, \tag{C.2}$$

where $\mathbb{B}_{\boldsymbol{\Sigma}}(r) = \{\boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}} \leq r\}$ and $\mathbb{C}_{\boldsymbol{\Sigma}}(l) = \{\boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}\|_1 \leq l\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}\}$. Since the smoothed quantile objective (2.4) is convex and twice continuously differentiable, it follows from (4.1) that

$$\begin{aligned} D(\boldsymbol{\beta}) &:= \langle \nabla \widehat{Q}_h(\boldsymbol{\beta}) - \nabla \widehat{Q}_h(\boldsymbol{\beta}^*), \boldsymbol{\beta} - \boldsymbol{\beta}^* \rangle \\ &= \frac{1}{n}\sum_{i=1}^n \left\{ \bar{K}\left(\frac{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} - y_i}{h}\right) - \bar{K}\left(\frac{-\varepsilon_i}{h}\right)\right\} \langle \boldsymbol{x}_i, \boldsymbol{\beta} - \boldsymbol{\beta}^* \rangle. \end{aligned} \tag{C.3}$$

For $i = 1, \ldots, n$, define the events $E_i = \{|\varepsilon_i| \leq h/2\} \cap \{|\langle \boldsymbol{x}_i, \boldsymbol{\beta} - \boldsymbol{\beta}^* \rangle| / \|\boldsymbol{\beta} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} \leq h/(2r)\}$, on which $|y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}| \leq h$ for any $\boldsymbol{\beta} \in \boldsymbol{\beta}^* + \mathbb{B}_{\boldsymbol{\Sigma}}(r)$. Since $\kappa_l = \min_{|u| \leq 1} K(u) > 0$, $D(\boldsymbol{\beta})$ can be lower bounded as

$$D(\boldsymbol{\beta}) \geq \frac{\kappa_l}{nh}\sum_{i=1}^n \langle \boldsymbol{x}_i, \boldsymbol{\beta} - \boldsymbol{\beta}^* \rangle^2 \mathbb{1}_{E_i}, \tag{C.4}$$

where $\mathbb{1}_{E_i}$ is the indicator function of E_i. Thus, it suffices to bound the right-hand side of the above inequality from below uniformly over $\boldsymbol{\beta} \in \Theta$.

To deal with the discontinuity, we use the following smoothing technique from Loh (2017), which turns the objective into a Lipschitz continuous empirical process. For $R > 0$, define the function

$$\varphi_R(u) = \begin{cases} u^2, & \text{if } |u| \leq \frac{R}{2}, \\ \{u - R\,\mathrm{sign}(u)\}^2, & \text{if } \frac{R}{2} \leq |u| \leq R, \\ 0, & \text{if } |u| > R, \end{cases}$$

which is R-Lipschitz continuous, and satisfies

$$\varphi_{cR}(cu) = c^2\varphi_R(u) \ \text{ for any } c > 0, \quad u^2\mathbb{1}(|u| \leq R/2) \leq \varphi_R(u) \leq u^2\mathbb{1}(|u| \leq R). \tag{C.5}$$

Together, (C.4) and (C.5) imply

$$\frac{D(\boldsymbol{\beta})}{\|\boldsymbol{\beta} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}}^2} \geq \kappa_l \cdot \underbrace{\frac{1}{nh}\sum_{i=1}^n \varphi_{h/(2r)}(\langle \boldsymbol{x}_i, \boldsymbol{\beta} - \boldsymbol{\beta}^* \rangle / \|\boldsymbol{\beta} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}}) \cdot \chi_i}_{=: D_0(\boldsymbol{\beta})}, \tag{C.6}$$

where $\chi_i = \mathbb{1}(|\varepsilon_i| \le h/2)$.

In the following, we bound the expectation $\mathbb{E}\{D_0(\boldsymbol{\beta})\}$ and the random fluctuation $D_0(\boldsymbol{\beta}) - \mathbb{E}\{D_0(\boldsymbol{\beta})\}$ over $\boldsymbol{\beta} \in \Theta$, respectively. For the binary variable χ_i, using Condition (B1) we have

$$|\mathbb{E}(\chi_i|\boldsymbol{x}_i) - hf_{\varepsilon|\boldsymbol{x}}(0)| \le \int_{-h/2}^{h/2} |f_{\varepsilon|\boldsymbol{x}}(t) - f_{\varepsilon|\boldsymbol{x}}(0)|\,\mathrm{d}t \le l_0 h^2/4. \tag{C.7}$$

Moreover, write $\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}^*$ for $\boldsymbol{\beta} \in \Theta$, and define the one-dimensional marginal $\xi_{\boldsymbol{\delta}} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}$ such that $\mathbb{E}(\xi_{\boldsymbol{\delta}}^2) = 1$. Provided $h \le f_l/(2l_0)$, it follows from (C.5) and (C.7) that

$$\mathbb{E}\{\varphi_{h/(2r)}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}) \cdot \chi_i\} = \mathbb{E}\{\varphi_{h/(2r)}(\xi_{\boldsymbol{\delta}}) \cdot \chi_i\} \ge \frac{7}{8} f_l h \cdot (1 - \mathbb{E}\{\xi_{\boldsymbol{\delta}}^2 \mathbb{1}_{|\xi_{\boldsymbol{\delta}}| > h/(4r)}\}).$$

For any $u > 0$, by the sub-exponential condition on $\boldsymbol{x} \in \mathbb{R}^p$, we have

$$\begin{aligned}
\mathbb{E}\{\xi_{\boldsymbol{\delta}}^2 \mathbb{1}(|\xi_{\boldsymbol{\delta}}| > u)\} &= 2\mathbb{E}\left\{\int_0^\infty t \cdot \mathbb{1}(|\xi_{\boldsymbol{\delta}}| > t) \cdot \mathbb{1}(|\xi_{\boldsymbol{\delta}}| > u)\,\mathrm{d}t\right\} \\
&= 2\int_0^u t \cdot \mathbb{E}\{\mathbb{1}(|\xi_{\boldsymbol{\delta}}| > t) \cdot \mathbb{1}(|\xi_{\boldsymbol{\delta}}| > u)\}\mathrm{d}t + 2\int_u^\infty t \cdot \mathbb{P}(|\xi_{\boldsymbol{\delta}}| > t)\,\mathrm{d}t \\
&= u^2 \mathbb{P}(|\xi_{\boldsymbol{\delta}}| > u) + 2\upsilon_0^2 \int_{u/\upsilon_0}^\infty t \cdot \mathbb{P}(|\xi_{\boldsymbol{\delta}}/\upsilon_0| \ge t)\,\mathrm{d}t \\
&\le u^2 e^{-u/\upsilon_0} + 2\upsilon_0^2 \int_{u/\upsilon_0}^\infty t e^{-t}\,\mathrm{d}t \\
&= (u^2 + 2\upsilon_0 u + 2\upsilon_0^2) e^{-u/\upsilon_0},
\end{aligned}$$

where the third equality follows from a change of variable. As long as $r \le h/(20\upsilon_0^2)$, taking $u = h/(4r) \ge 5\upsilon_0^2$ in the above bound yields $\mathbb{E}\{\xi_{\boldsymbol{\delta}}^2 \mathbb{1}(|\xi_{\boldsymbol{\delta}}| > h/(4r))\} < 1/4$. Consequently,

$$\inf_{\boldsymbol{\beta} \in \Theta} \mathbb{E}\{D_0(\boldsymbol{\beta})\} \ge \frac{21}{32} f_l \ \text{ as long as } \ 20\upsilon_0^2 r \le h \le f_l/(2l_0). \tag{C.8}$$

Next we evaluate the random fluctuation term

$$\Omega := \sup_{\boldsymbol{\delta} \in \mathbb{C}_{\boldsymbol{\Sigma}}(l)} |D_0(\boldsymbol{\delta}) - \mathbb{E}\{D_0(\boldsymbol{\delta})\}|. \tag{C.9}$$

Write $\omega_{\boldsymbol{\delta}}(\boldsymbol{x}_i, \varepsilon_i) = \varphi_{h/(2r)}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}) \cdot \chi_i/h$, so that $D_0(\boldsymbol{\delta}) = (1/n)\sum_{i=1}^n \omega_{\boldsymbol{\delta}}(\boldsymbol{x}_i, \varepsilon_i)$. By (C.6) and (C.7), and the fact that $\varphi_R(u) \le (R/2)|u|$, we have

$$0 \le \omega_{\boldsymbol{\delta}}(\boldsymbol{x}_i, \varepsilon_i) \le (4r)^{-2} h \quad \text{and} \quad \mathbb{E}\omega_{\boldsymbol{\delta}}^2(\boldsymbol{x}_i, \varepsilon_i) \le (4r)^{-2} \cdot 9 f_u h/8.$$

With the above preparations, we apply Theorem 7.3 in Bousquet (2003) (a refined Talagrand's inequality) to conclude that, for any $t > 0$,

$$\begin{aligned}
\Omega &\le \mathbb{E}\Omega + (\mathbb{E}\Omega)^{1/2} \frac{1}{2r}\sqrt{\frac{ht}{n}} + \frac{3}{2} f_u^{1/2} (4r)^{-1} \sqrt{\frac{ht}{n}} + \frac{h}{(4r)^2} \frac{t}{3n} \\
&\le \frac{5}{4}\mathbb{E}\Omega + \frac{3}{2} f_u^{1/2} (4r)^{-1} \sqrt{\frac{ht}{n}} + (4 + 1/3) \frac{ht}{(4r)^2 n}
\end{aligned} \tag{C.10}$$

with probability at least $1 - e^{-t}$, where the second step follows from the inequality that $ab \le a^2/4 + b^2$ for all $a, b \in \mathbb{R}$.

It then remains to bound the expectation $\mathbb{E}\Omega$. By Rademacher symmetrization,

$$\mathbb{E}\Omega \le 2\mathbb{E}\bigg\{ \sup_{\boldsymbol{\delta}\in\mathbb{C}_{\boldsymbol{\Sigma}}(l)} \frac{1}{n}\sum_{i=1}^n e_i \omega_{\boldsymbol{\delta}}(\boldsymbol{x}_i, \varepsilon_i)\bigg\},$$

where $e_1, \ldots, e_n$ are independent Rademacher random variables. Since $\chi_i = \mathbb{1}(|\varepsilon_i| \le h/2) \in \{0, 1\}$, $\omega_{\boldsymbol{\delta}}(\boldsymbol{x}_i, \varepsilon_i)$ can be written as $\omega_{\boldsymbol{\delta}}(\boldsymbol{x}_i, \varepsilon_i) = h^{-1}\varphi_{h/(2r)}(\chi_i \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}})$. By the Lipschitz continuity of $\varphi_R(\cdot)$, $\omega_{\boldsymbol{\delta}}(\boldsymbol{x}_i, \varepsilon_i)$ is a $(2r)^{-1}$-Lipschitz function in $\chi_i \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}$, i.e., for any sample $(\boldsymbol{x}_i, \varepsilon_i)$ and parameters $\boldsymbol{\delta}, \boldsymbol{\delta}' \in \mathbb{R}^p$,

$$|\omega_{\boldsymbol{\delta}}(\boldsymbol{x}_i, \varepsilon_i) - \omega_{\boldsymbol{\delta}'}(\boldsymbol{x}_i, \varepsilon_i)| \le \frac{1}{2r}|\chi_i \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}} - \chi_i \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}'/\|\boldsymbol{\delta}'\|_{\boldsymbol{\Sigma}}|. \tag{C.11}$$

Moreover, $\omega_{\boldsymbol{\delta}}(\boldsymbol{x}_i, \varepsilon_i) = 0$ for any $\boldsymbol{\delta}$ such that $\chi_i \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}} = 0$. To use Talagrand's contraction principle to bound the Rademacher complexity, define the subset $T \subseteq \mathbb{R}^n$

$$T = \{\boldsymbol{t} = (t_1, \ldots, t_n)^{\mathrm{T}} : t_i = \langle \chi_i \boldsymbol{x}_i, \boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}\rangle, i = 1, \ldots, n, \ \boldsymbol{\delta} \in \mathbb{C}_{\boldsymbol{\Sigma}}(l)\},$$

and contractions $\phi_i : \mathbb{R} \to \mathbb{R}$ as $\phi_i(t) = (2r/h) \cdot \varphi_{h/(2r)}(t)$. By (C.11), $|\phi(t) - \phi(s)| \le |t - s|$ for all $t, s \in \mathbb{R}$. Applying Talagrand's contraction principle (see, e.g., Theorem 4.12 and (4.20) in Ledoux and Talagrand (1991)), we have

$$\begin{aligned}
\mathbb{E}\Omega &\le 2\mathbb{E}\bigg\{ \sup_{\boldsymbol{\delta}\in\mathbb{C}_{\boldsymbol{\Sigma}}(l)} \frac{1}{n}\sum_{i=1}^n e_i \omega_{\boldsymbol{\delta}}(\boldsymbol{x}_i, \varepsilon_i)\bigg\} \\
&= \frac{1}{r}\mathbb{E}\bigg\{ \sup_{\boldsymbol{t}\in T} \frac{1}{n}\sum_{i=1}^n e_i \phi_i(t_i)\bigg\} \\
&\le \frac{1}{r}\mathbb{E}\bigg(\sup_{\boldsymbol{t}\in T} \frac{1}{n}\sum_{i=1}^n e_i t_i\bigg) \\
&= \frac{1}{r}\mathbb{E}\bigg\{ \sup_{\boldsymbol{\delta}\in\mathbb{C}_{\boldsymbol{\Sigma}}(l)} \frac{1}{n}\sum_{i=1}^n e_i \langle \chi_i \boldsymbol{x}_i, \boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}\rangle\bigg\} \\
&\le \frac{l}{r}\cdot \mathbb{E}\bigg\|\frac{1}{n}\sum_{i=1}^n e_i \chi_i \boldsymbol{x}_i\bigg\|_\infty,
\end{aligned} \tag{C.12}$$

where the last inequality follows from the cone constraint that $\|\boldsymbol{\delta}\|_1 \le l\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}$. The problem is then boiled down to bounding the expectation on the right-hand side of (C.12). For each $1 \le j \le p$, define the partial sum $S_j = \sum_{i=1}^n e_i \chi_i x_{ij}$, of which each summand satisfies $\mathbb{E}(e_i \chi_i x_{ij}) = 0$ and $\mathbb{E}(e_i \chi_i x_{ij})^2 \le \sigma_{jj} c_h := \sigma_{jj}(f_u h + l_0 h^2/4)$ due to (C.7). In addition, for $k = 3, 4, \ldots$,

$$\begin{aligned}
\mathbb{E}|e_i \chi_i x_{ij}|^k &\le \upsilon_0^k \sigma_{jj}^{k/2} c_h \cdot k \int_0^\infty t^{k-1}\mathbb{P}(|x_{ij}| \ge \upsilon_0 \sigma_{jj}^{1/2} t)\,\mathrm{d}t \\
&\le \upsilon_0^k \sigma_{jj}^{k/2} c_h \cdot k \int_0^\infty t^{k-1} e^{-t}\,\mathrm{d}t \\
&= k!\upsilon_0^k \sigma_{jj}^{k/2} c_h \le \frac{k!}{2}\cdot \upsilon_0^2 \sigma_{jj} c_h \cdot (2\upsilon_0 \sigma_{jj}^{1/2})^{k-2}.
\end{aligned}$$

Following the proof of Theorems 2.10 and 2.5 in Boucheron, Lugosi and Massart (2013), it can be shown that for all $\lambda \in (0, 1/c)$, $\log \mathbb{E}e^{\lambda S_j} \le \psi(\lambda) := \frac{v\lambda^2}{2(1-c\lambda)}$ and

$$\mathbb{E}\max_{1\le j\le p} |S_j| \le \inf_{\lambda\in(0,1/c)}\bigg\{\frac{\log(2p) + \psi(\lambda)}{\lambda}\bigg\} = \sqrt{2v\log(2p)} + c\log(2p),$$

where $v = \upsilon_0^2 \sigma_{\boldsymbol{x}}^2 c_h \cdot n$ and $c = 2\upsilon_0 \sigma_{\boldsymbol{x}}$. Re-arranging terms and using (C.12) yield

$$\mathbb{E}\Omega \le \upsilon_0 \sigma_{\boldsymbol{x}} \frac{l}{r} \left\{ \frac{3}{2} f_u^{1/2} \sqrt{\frac{h \log(2p)}{n}} + \frac{2 \log(2p)}{n} \right\}. \tag{C.13}$$

Consequently, it follows from (C.9), (C.10) with $t = \log(2p)$ and (C.13) that

$$\begin{aligned} \Omega &= \sup_{\boldsymbol{\delta} \in \mathbb{C}_{\boldsymbol{\Sigma}}(l)} |D_0(\boldsymbol{\delta}) - \mathbb{E}\{D_0(\boldsymbol{\delta})\}| \\ &\le \upsilon_0 \sigma_{\boldsymbol{x}} \frac{l}{r} \left\{ \frac{15}{8} f_u^{1/2} \sqrt{\frac{h \log(2p)}{n}} + \frac{5}{2} \frac{\log(2p)}{n} \right\} + \frac{3}{2} f_u^{1/2} (4r)^{-1} \sqrt{\frac{h \log(2p)}{n}} + (4 + 1/3) \frac{h \log(2p)}{(4r)^2 n} \end{aligned} \tag{C.14}$$

with probability at least $1 - (2p)^{-1}$.

Finally, from the bounds (C.3), (C.6), (C.8) and (C.14) we conclude that as long as $n \ge C f_u f_l^{-2} (l/r)^2 h \log(2p)$ for a sufficiently large C depending only on $(\upsilon_0, \sigma_{\boldsymbol{x}})$,

$$\inf_{\boldsymbol{\beta} \in \boldsymbol{\beta}^* + \mathbb{B}_{\boldsymbol{\Sigma}}(r) \cap \mathbb{C}_{\boldsymbol{\Sigma}}(l)} \frac{\langle \nabla \widehat{Q}_h(\boldsymbol{\beta}) - \nabla \widehat{Q}_h(\boldsymbol{\beta}^*), \boldsymbol{\beta} - \boldsymbol{\beta}^* \rangle}{\|\boldsymbol{\beta} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}}^2} \ge \frac{1}{2} \kappa_l f_l$$

holds with probability at least $1 - (2p)^{-1}$, as claimed. □

C.2 Proof of Theorem 4.1

Let $\mathcal{S} \subseteq [p]$ be the active set of $\boldsymbol{\beta}^*$ with cardinality $s = |\mathcal{S}|$. The symmetric Bregman divergence between $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_h$ and $\boldsymbol{\beta}^*$ is defined as

$$\langle \nabla \widehat{Q}_h(\widehat{\boldsymbol{\beta}}) - \nabla \widehat{Q}_h(\boldsymbol{\beta}^*), \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \rangle \ge 0. \tag{C.15}$$

The proof of Theorem 4.1 involves establishing upper and lower bounds for the symmetric Bregman divergence (C.15).

Step 1: Upper Bound. By the first-order optimality condition of (4.5), there exists a subgradient $\widehat{\boldsymbol{g}} \in \partial \|\widehat{\boldsymbol{\beta}}\|_1$ such that $\nabla \widehat{Q}_h(\widehat{\boldsymbol{\beta}}) + \lambda \widehat{\boldsymbol{g}} = \mathbf{0}$. Set $\widehat{\boldsymbol{\delta}} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$. By the definition of subgradient, we have

$$\begin{aligned} \langle \widehat{\boldsymbol{g}}, \boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}} \rangle &\le \|\boldsymbol{\beta}^*\|_1 - \|\widehat{\boldsymbol{\beta}}\|_1 = \|\boldsymbol{\beta}^*_{\mathcal{S}}\|_1 - \|\widehat{\boldsymbol{\delta}} + \boldsymbol{\beta}^*\|_1 \\ &= \|\boldsymbol{\beta}^*_{\mathcal{S}}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}^c}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}} + \boldsymbol{\beta}^*_{\mathcal{S}}\|_1 \le \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}^c}\|_1, \end{aligned} \tag{C.16}$$

where the last inequality holds by the reverse triangle inequality. Substituting the first-order optimality condition into (C.15) yields

$$\begin{aligned} &\langle \nabla \widehat{Q}_h(\widehat{\boldsymbol{\beta}}) - \nabla \widehat{Q}_h(\boldsymbol{\beta}^*), \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \rangle \\ &= \lambda \langle \widehat{\boldsymbol{g}}, \boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}} \rangle + \langle \nabla \widehat{Q}_h(\boldsymbol{\beta}^*) - \nabla Q_h(\boldsymbol{\beta}^*), \boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}} \rangle + \langle \nabla Q_h(\boldsymbol{\beta}^*), \boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}} \rangle \\ &\le \lambda (\|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_1 - \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}^c}\|_1) + \underbrace{\|\nabla \widehat{Q}_h(\boldsymbol{\beta}^*) - \nabla Q_h(\boldsymbol{\beta}^*)\|_\infty}_{=:\|\boldsymbol{w}_h^*\|_\infty} \|\widehat{\boldsymbol{\delta}}\|_1 + \underbrace{\|\boldsymbol{\Sigma}^{-1/2} \nabla Q_h(\boldsymbol{\beta}^*)\|_2}_{=:b_h^*} \|\widehat{\boldsymbol{\delta}}\|_{\boldsymbol{\Sigma}}, \end{aligned}$$

where $Q_h(\cdot)$ is the population smoothed quantile objective defined in (4.2). Here $\|\boldsymbol{w}_h^*\|_\infty$ is a stochastic term that determines the statistical error, and b_h^* is the (deterministic) smoothing bias satisfying $b_h^* \le l_0\kappa_2 h^2/2$ due to (B.5). Conditioned on the event $\{\lambda \ge 2\|\boldsymbol{w}_h^*\|_\infty\}$, we have

$$\begin{aligned}
&\langle \nabla \widehat{Q}_h(\widehat{\boldsymbol\beta}) - \nabla \widehat{Q}_h(\boldsymbol\beta^*), \widehat{\boldsymbol\beta} - \boldsymbol\beta^* \rangle \\
&\le \lambda(\|\widehat{\boldsymbol\delta}_{\mathcal S}\|_1 - \|\widehat{\boldsymbol\delta}_{\mathcal S^c}\|_1 + \|\widehat{\boldsymbol\delta}\|_1/2) + b_h^*\|\widehat{\boldsymbol\delta}\|_{\boldsymbol\Sigma} \\
&\le \frac{\lambda}{2}(3\|\widehat{\boldsymbol\delta}_{\mathcal S}\|_1 - \|\widehat{\boldsymbol\delta}_{\mathcal S^c}\|_1) + b_h^*\|\widehat{\boldsymbol\delta}\|_{\boldsymbol\Sigma} \\
&\le \frac{3}{2}s^{1/2}\lambda\|\widehat{\boldsymbol\delta}\|_2 + b_h^*\|\widehat{\boldsymbol\delta}\|_{\boldsymbol\Sigma}. \qquad \text{(C.17)}
\end{aligned}$$

Since $\langle \nabla \widehat{Q}_h(\widehat{\boldsymbol\beta}) - \nabla \widehat{Q}_h(\boldsymbol\beta^*), \widehat{\boldsymbol\beta} - \boldsymbol\beta^* \rangle \ge 0$, as a byproduct, we see from (C.17) that $\widehat{\boldsymbol\delta}$ satisfies the cone-type constraint $\|\widehat{\boldsymbol\delta}_{\mathcal S^c}\|_1 \le 3\|\widehat{\boldsymbol\delta}_{\mathcal S}\|_1 + 2\lambda^{-1}b_h^*\|\widehat{\boldsymbol\delta}\|_{\boldsymbol\Sigma}$, from which it follows that

$$\|\widehat{\boldsymbol\delta}\|_1 \le 4s^{1/2}\|\widehat{\boldsymbol\delta}\|_2 + 2\lambda^{-1}b_h^*\|\widehat{\boldsymbol\delta}\|_{\boldsymbol\Sigma} \le (4\gamma_p^{-1/2}s^{1/2} + l_0\kappa_2\lambda^{-1}h^2)\|\widehat{\boldsymbol\delta}\|_{\boldsymbol\Sigma}. \qquad \text{(C.18)}$$

We then let $h^2 \le s^{1/2}\lambda$ hereinafter, so that conditioned on $\{\lambda \ge 2\|\boldsymbol{w}_h^*\|_\infty\}$, $\widehat{\boldsymbol\beta} \in \boldsymbol\beta^* + \mathbb{C}_{\boldsymbol\Sigma}(l)$ with $l = (4\gamma_p^{-1/2} + l_0\kappa_2)s^{1/2}$.

Step 2: Lower Bound. Set $r = h/(20\upsilon_0^2)$. Recall from Proposition 4.2 that the RSC property only holds (with high probability) in a local neighborhood $\boldsymbol\beta^* + \mathbb{B}_{\boldsymbol\Sigma}(r) \cap \mathbb{C}_{\boldsymbol\Sigma}(l)$, to which $\widehat{\boldsymbol\beta}$ does not necessarily belong. Similarly to the proof of Proposition 4.1, we define $\eta = \sup\{u \in [0,1] : \boldsymbol\beta^* + u(\widehat{\boldsymbol\beta} - \boldsymbol\beta^*) \in \mathbb{B}_{\boldsymbol\Sigma}(r)\}$, and an intermediate vector $\widetilde{\boldsymbol\beta} = (1-\eta)\boldsymbol\beta^* + \eta\widehat{\boldsymbol\beta}$ that falls in $\boldsymbol\beta^* + \mathbb{B}_{\boldsymbol\Sigma}(r)$. By this definition, we have $\eta = 1$ if $\widehat{\boldsymbol\beta} \in \boldsymbol\beta^* + \mathbb{B}_{\boldsymbol\Sigma}(r)$, and $\eta \in (0,1)$ if $\widehat{\boldsymbol\beta} \notin \boldsymbol\beta^* + \mathbb{B}_{\boldsymbol\Sigma}(r)$. In the latter case, $\widetilde{\boldsymbol\beta}$ lies at the boundary of $\boldsymbol\beta^* + \mathbb{B}_{\boldsymbol\Sigma}(r)$. Since $\widetilde{\boldsymbol\beta} - \boldsymbol\beta^* = \eta(\widehat{\boldsymbol\beta} - \boldsymbol\beta^*)$, by (C.18) we also have $\widetilde{\boldsymbol\beta} \in \boldsymbol\beta^* + \mathbb{C}_{\boldsymbol\Sigma}(l)$ conditioned on $\{\lambda \ge 2\|\boldsymbol{w}_h^*\|_\infty\}$. Consequently, conditioned on $\{\lambda \ge 2\|\boldsymbol{w}_h^*\|_\infty\} \cap \mathcal{E}_{\rm rsc}(r,l,\kappa)$ with $\kappa = (\kappa_l f_l)/2$,

$$\langle \nabla \widehat{Q}_h(\widetilde{\boldsymbol\beta}) - \nabla \widehat{Q}_h(\boldsymbol\beta^*), \widetilde{\boldsymbol\beta} - \boldsymbol\beta^* \rangle \ge \frac{1}{2}\kappa_l f_l\,\|\widetilde{\boldsymbol\beta} - \boldsymbol\beta^*\|_{\boldsymbol\Sigma}^2. \qquad \text{(C.19)}$$

Step 3: Combining Lower and Upper Bounds. To bridge the upper and lower bounds obtained above, we apply (B.1) with $(Q_h, \boldsymbol\beta_h^*, \boldsymbol\beta_h^\dagger)$ replaced by $(\widehat{Q}_h, \widehat{\boldsymbol\beta}, \widetilde{\boldsymbol\beta})$ to conclude that

$$\langle \nabla \widehat{Q}_h(\widetilde{\boldsymbol\beta}) - \nabla \widehat{Q}_h(\boldsymbol\beta^*), \widetilde{\boldsymbol\beta} - \boldsymbol\beta^* \rangle \le \eta\langle \nabla \widehat{Q}_h(\widehat{\boldsymbol\beta}) - \nabla \widehat{Q}_h(\boldsymbol\beta^*), \widehat{\boldsymbol\beta} - \boldsymbol\beta^* \rangle.$$

This, combined with (C.17) and (C.19), implies that conditioned on $\{\lambda \ge 2\|\boldsymbol{w}_h^*\|_\infty\} \cap \mathcal{E}_{\rm rsc}(r,l,\kappa)$,

$$\frac{1}{2}\kappa_l f_l\,\|\widetilde{\boldsymbol\delta}\|_{\boldsymbol\Sigma}^2 \le \frac{3}{2}s^{1/2}\lambda\|\widetilde{\boldsymbol\delta}\|_2 + b_h^*\|\widetilde{\boldsymbol\delta}\|_{\boldsymbol\Sigma} \le \frac{3}{2}\gamma_p^{-1/2}s^{1/2}\lambda\|\widetilde{\boldsymbol\delta}\|_{\boldsymbol\Sigma} + \frac{1}{2}l_0\kappa_2 h^2\|\widetilde{\boldsymbol\delta}\|_{\boldsymbol\Sigma},$$

where $\widetilde{\boldsymbol\delta} = \widetilde{\boldsymbol\beta} - \boldsymbol\beta^*$. Canceling $\|\widetilde{\boldsymbol\delta}\|_{\boldsymbol\Sigma}$ and re-arranging the terms yield

$$\|\widetilde{\boldsymbol\delta}\|_{\boldsymbol\Sigma} \le \frac{1}{\kappa_l f_l}(3\gamma_p^{-1/2}s^{1/2}\lambda + l_0\kappa_2 h^2) \le \frac{1}{\kappa_l f_l}(3\gamma_p^{-1/2} + l_0\kappa_2)s^{1/2}\lambda. \qquad \text{(C.20)}$$

It remains to control the probability of the event $\{\lambda \ge 2\|\boldsymbol{w}_h^*\|_\infty\} \cap \mathcal{E}_{\rm rsc}(r,l,\kappa)$, where $l = (4\gamma_p^{-1/2} + l_0\kappa_2)s^{1/2}$ and $\kappa = (\kappa_l f_l)/2$. By Proposition 4.2 and Lemma C.2, we take

$$\lambda = 2\upsilon_0\sigma_{\boldsymbol x}\left\{\sqrt{\{\tau(1-\tau) + (1+\tau)l_0\kappa_2 h^2\}\frac{\log(2p)}{n}} + \max(1-\tau,\tau)\frac{2\log(2p)}{n}\right\}, \qquad \text{(C.21)}$$

so that $\{\lambda \geq 2\|\boldsymbol{w}_h^*\|_\infty\} \cap \mathcal{E}_{\rm rsc}(r, l, \kappa)$ occurs with probability at least $1 - p^{-1}$ as long as

$$\frac{\sigma_{\boldsymbol{x}}^2 f_u}{(\kappa_l f_l)^2} \frac{s \log(p)}{n} \lesssim h \leq f_l/(2l_0).$$

This certifies the error bound (C.20) for $\widetilde{\boldsymbol{\beta}}$. Assume further that

$$\frac{1}{\kappa_l f_l}(3\gamma_p^{-1/2} + l_0\kappa_2)s^{1/2}\lambda < r = \frac{h}{20\upsilon_0^2},$$

then $\widetilde{\boldsymbol{\beta}}$ falls in the interior of $\mathbb{B}_{\boldsymbol{\Sigma}}(r)$ with high probability. Via proof by contradiction, we must have $\eta = 1$ and thus $\widehat{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}$ also satisfies (C.20). This proves the claimed bounds. □

D Proofs of Results in Section 4.3

We first provide two high-level results on the cone property and ℓ_2-error bound of the weighted ℓ_1-penalized smoothed QR estimator that solves (2.7). Recall that $\boldsymbol{w}_h^* = \boldsymbol{w}_h(\boldsymbol{\beta}^*) \in \mathbb{R}^p$ and $b_h^* = b_h(\boldsymbol{\beta}^*)$, where $\boldsymbol{w}_h(\boldsymbol{\beta}) = \nabla \widehat{Q}_h(\boldsymbol{\beta}) - \nabla Q_h(\boldsymbol{\beta})$ and $b_h(\boldsymbol{\beta}) = \|\boldsymbol{\Sigma}^{-1/2} \nabla Q_h(\boldsymbol{\beta})\|_2$. Lemma D.1 provides conditions under which the optimal solution to the convex problem (2.7) falls in an ℓ_1-cone.

Lemma D.1. Let $\mathcal{T}$ be a subset of $[p]$ satisfying $\mathcal{S} \subseteq \mathcal{T}$, and let $\boldsymbol{\beta} \in \mathbb{R}^p$ be such that $\boldsymbol{\beta}_{\mathcal{T}^{\rm c}} = \mathbf{0}$. Conditioned on $\{\|\boldsymbol{\lambda}_{\mathcal{T}^{\rm c}}\|_{\min} > \|\boldsymbol{w}_h(\boldsymbol{\beta})\|_\infty\}$, any optimal solution $\widehat{\boldsymbol{\beta}}$ to (2.7) satisfies

$$\|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})_{\mathcal{T}^{\rm c}}\|_1 \leq \frac{\{\|\boldsymbol{\lambda}\|_\infty + \|\boldsymbol{w}_h(\boldsymbol{\beta})\|_\infty\}\|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})_{\mathcal{T}}\|_1 + b_h(\boldsymbol{\beta})\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_{\boldsymbol{\Sigma}}}{\|\boldsymbol{\lambda}_{\mathcal{T}^{\rm c}}\|_{\min} - \|\boldsymbol{w}_h(\boldsymbol{\beta})\|_\infty}.$$

Lemma D.2. Let $\mathcal{T}$ be a subset of $[p]$ such that $\mathcal{S} \subseteq \mathcal{T}$ and $k = |\mathcal{T}|$, and let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^{\rm T}$ satisfy $\|\boldsymbol{\lambda}\|_\infty \leq \lambda$, $\|\boldsymbol{\lambda}_{\mathcal{T}^{\rm c}}\|_{\min} \geq a\lambda$ for some $0 < a \leq 1$ and $\lambda \geq (s/\gamma_p)^{-1/2} b_h^*$. Conditioned on the event $\{\lambda \geq (2/a)\|\boldsymbol{w}_h^*\|_\infty\}$, any optimal solution $\widehat{\boldsymbol{\beta}}$ to (2.7) satisfies $\widehat{\boldsymbol{\beta}} \in \boldsymbol{\beta}^* + \mathbb{C}_{\boldsymbol{\Sigma}}(l)$, where $l = l(a, k) := (2 + 2/a)(k/\gamma_p)^{1/2} + (2/a)(s/\gamma_p)^{1/2}$. Moreover, let $r, \kappa > 0$ satisfy

$$\gamma_p^{-1/2}(0.5ak^{1/2} + 2s^{1/2})\lambda < r \cdot \kappa. \tag{D.1}$$

Then, conditioned on $\{\lambda \geq (2/a)\|\boldsymbol{w}_h^*\|_\infty\} \cap \mathcal{E}_{\rm rsc}(r, l, \kappa)$,

$$\begin{aligned} \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} &\leq \kappa^{-1}\{\gamma_p^{-1/2}(\|\boldsymbol{w}_{h,\mathcal{T}}^*\|_2 + \|\boldsymbol{\lambda}_{\mathcal{S}}\|_2) + b_h^*\} \\ &\leq \kappa^{-1}\gamma_p^{-1/2}(0.5ak^{1/2} + 2s^{1/2})\lambda. \end{aligned} \tag{D.2}$$

Lemma D.3 provides a probabilistic bound for the stochastic term $\|\boldsymbol{w}_{h,\mathcal{S}}^*\|_2$, which determines the oracle rate of convergence.

Lemma D.3. Assume that Conditions (B1)–(B3) hold. Then, for any $t > 0$,

$$\|\mathbf{S}^{-1/2}\boldsymbol{w}_{h,\mathcal{S}}^*\|_2 \leq 3\upsilon_0\left[\sqrt{\{\tau(1-\tau) + Ch^2\}\frac{2s+t}{n}} + \max(1-\tau, \tau)\frac{2s+t}{n}\right] \tag{D.3}$$

holds with probability at least $1 - e^{-t}$, where $C = (\tau + 1)l_0\kappa_2$ and $\mathbf{S} = \mathbb{E}(\boldsymbol{x}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}}^{\rm T}) \in \mathbb{R}^{s\times s}$.

D.1 Proof of Theorem 4.2

The proof is based a deterministic analysis conditioning on the event $\{\|\boldsymbol{w}_h^*\|_\infty \le 0.5q'(\alpha_0)\lambda\}$ for some $\lambda \ge (s/\gamma_p)^{-1/2} b_h^*$. We extend the argument used in the proof of Theorem 4.2 in Fan et al. (2018) with a more delicate treatment of the local RSC property and smoothing bias. With an initial estimator $\widehat{\boldsymbol{\beta}}^{(0)} = \mathbf{0}$, we have $\boldsymbol{\lambda}^{(0)} = (\lambda, \ldots, \lambda)^{\mathrm{T}} \in \mathbb{R}^p$. Applying Lemma D.2 with $\mathcal{T} = \mathcal{S}$ and $a = 1$ yields that, conditioned further on the event $\mathcal{E}_{\mathrm{rsc}}(r, l(1, s), \kappa)$,

$$\begin{aligned} \|\widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} &\le \kappa^{-1}\{\gamma_p^{-1/2}(\|\boldsymbol{w}_{h,\mathcal{S}}^*\|_2 + \|\boldsymbol{\lambda}_{\mathcal{S}}\|_2) + b_h^*\} \\ &\le \kappa^{-1}\{\gamma_p^{-1/2}(1 + 0.5q'(\alpha_0))s^{1/2}\lambda + b_h^*\} \\ &\le \kappa^{-1}\{0.5q'(\alpha_0) + 2\}(s/\gamma_p)^{1/2}\lambda, \end{aligned} \tag{D.4}$$

where $l(1, s) = 6(s/\gamma_p)^{1/2}$.

To improve the statistical rate of $\widehat{\boldsymbol{\beta}}^{(\ell)} = (\widehat{\beta}_1^{(\ell)}, \ldots, \widehat{\beta}_p^{(\ell)})^{\mathrm{T}}$ at step $\ell \ge 2$, we need to control the magnitude of the false discoveries of the solution obtained from the previous step, that is, $\max_{j \in \mathcal{S}^{\mathrm{c}}} |\widehat{\beta}_j^{(\ell-1)}|$. Recall that $\boldsymbol{\lambda}^{(\ell-1)} = (\lambda_1^{(\ell-1)}, \ldots, \lambda_p^{(\ell-1)})^{\mathrm{T}} = (q_\lambda'(|\widehat{\beta}_1^{(\ell-1)}|), \ldots, q_\lambda'(|\widehat{\beta}_p^{(\ell-1)}|))^{\mathrm{T}}$, where $q_\lambda(t) = \lambda^2 q(t/\lambda)$ for $t \ge 0$. Since $q'(\cdot)$ is monotone on $\mathbb{R}^+$, large magnitudes of $|\widehat{\beta}_j^{(\ell-1)}|$ indicate small values of $\lambda_j^{(\ell-1)}$. Motivated by this observation, we construct an augmented index set $\mathcal{T}_\ell$, satisfying $\mathcal{S} \subseteq \mathcal{T}_\ell \subseteq [p]$, in each step and control the magnitude of $\|\boldsymbol{\lambda}_{\mathcal{T}_\ell^{\mathrm{c}}}^{(\ell-1)}\|_{\min}$.

For $\ell = 1, 2, \ldots$, define the index set

$$\mathcal{T}_\ell = \mathcal{S} \cup \{1 \le j \le p : \lambda_j^{(\ell-1)} < q'(\alpha_0)\lambda\}, \tag{D.5}$$

which depends on $\widehat{\boldsymbol{\beta}}^{(\ell-1)}$. Let $c > 0$ be determined by equation (4.12). We claim that

$$|\mathcal{T}_\ell| < (c^2 + 1)s \quad \text{and} \quad \|\boldsymbol{\lambda}_{\mathcal{T}_\ell^{\mathrm{c}}}^{(\ell-1)}\|_{\min} \ge q'(\alpha_0)\lambda. \tag{D.6}$$

Indeed, if these two inequalities hold, applying Lemma D.2 with $a = q'(\alpha_0)$, $k = (c^2 + 1)s$ and $l = \{(2 + \frac{2}{q'(\alpha_0)})(c^2 + 1)^{1/2} + \frac{2}{q'(\alpha_0)}\}(s/\gamma_p)^{1/2}$ implies that, conditioned on $\mathcal{E}_{\mathrm{rsc}}(r, l, \kappa)$,

$$\|\widehat{\boldsymbol{\beta}}^{(\ell)} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} \le \kappa^{-1}\{\gamma_p^{-1/2}(\|\boldsymbol{w}_{h,\mathcal{T}_\ell}^*\|_2 + \|\boldsymbol{\lambda}_{\mathcal{S}}^{(\ell-1)}\|_2) + b_h^*\} \tag{D.7}$$

$$\begin{aligned} &< \kappa^{-1}\{0.5q'(\alpha_0)(c^2 + 1)^{1/2} + 2\}(s/\gamma_p)^{1/2}\lambda \\ &\le \gamma_p^{1/2}\alpha_0 c s^{1/2}\lambda = r_{\mathrm{opt}}, \end{aligned} \tag{D.8}$$

where we have used (4.11) and (4.12) in the second and third inequalities, respectively.

We now verify the claim (D.6) by induction on ℓ. The claim is trivial if $\ell = 1$, in which case $\boldsymbol{\lambda}^{(0)} = (\lambda, \ldots, \lambda)^{\mathrm{T}}$ and $\mathcal{T}_1 = \mathcal{S}$. Next, assume that for some integer $\ell \ge 1$, (D.6) holds and so does (D.8). First we show that $|\mathcal{T}_{\ell+1}| < (c^2 + 1)s$. For any $j \in \mathcal{T}_{\ell+1} \setminus \mathcal{S}$, $q_\lambda'(|\widehat{\beta}_j^{(\ell)}|) = \lambda_j^{(\ell)} < q'(\alpha_0)\lambda = q_\lambda'(\alpha_0\lambda)$, implying $|\widehat{\beta}_j^{(\ell)}| > \alpha_0\lambda$ by the monotonicity of q_λ' on $\mathbb{R}^+$. Recalling that $\beta_j^* = 0$ for $j \in \mathcal{T}_{\ell+1} \setminus \mathcal{S}$ and that the bound (D.8) holds for $\widehat{\boldsymbol{\beta}}^{(\ell)}$ by induction, we have

$$\begin{aligned} |\mathcal{T}_{\ell+1} \setminus \mathcal{S}|^{1/2} &< (\alpha_0\lambda)^{-1}\|\widehat{\boldsymbol{\beta}}_{\mathcal{T}_{\ell+1}\setminus\mathcal{S}}^{(\ell)}\|_2 = (\alpha_0\lambda)^{-1}\|(\widehat{\boldsymbol{\beta}}^{(\ell)} - \boldsymbol{\beta}^*)_{\mathcal{T}_{\ell+1}\setminus\mathcal{S}}\|_2 \\ &\le (\alpha_0\lambda)^{-1}\gamma_p^{-1/2}\|\widehat{\boldsymbol{\beta}}^{(\ell)} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} \le c s^{1/2}. \end{aligned} \tag{D.9}$$

Consequently, $|\mathcal{T}_{\ell+1}| = |\mathcal{S}| + |\mathcal{T}_{\ell+1} \setminus \mathcal{S}| < (c^2 + 1)s$, as claimed. Turning to $\|\boldsymbol{\lambda}_{\mathcal{T}_{\ell+1}^{\mathrm{c}}}^{(\ell)}\|_{\min}$, it follows from (D.5) that $\lambda_j^{(\ell)} \ge q'(\alpha_0)\lambda$ for each $j \in \mathcal{T}_{\ell+1}^{\mathrm{c}}$. This completes the proof of (D.6).

Thus far, we have shown that the bounds (D.7) and (D.8) hold for every $\ell \geq 1$. Specifically, the latter implies $\widehat{\beta}^{(\ell)} \in \beta^* + \mathbb{B}_{\Sigma}(r_{\rm opt})$ for every $\ell \geq 1$, where $r_{\rm opt} \asymp s^{1/2}\lambda$. Next we will show that, when the signal is sufficiently strong and if a concave penality q_λ is used, the error bound $r_{\rm opt}$ can be refined at each iteration. By (D.7), a key step is to derive sharper bounds on

$$\|\boldsymbol{\lambda}_{\mathcal{S}}^{(\ell-1)}\|_2 = \sqrt{\sum\nolimits_{j\in\mathcal{S}}\{\lambda_j^{(\ell-1)}\}^2} = \sqrt{\sum\nolimits_{j\in\mathcal{S}}\{q_\lambda'(|\widehat{\beta}_j^{(\ell-1)}|)\}^2} \quad \text{and} \quad \|\boldsymbol{w}_{h,\mathcal{T}_\ell}^*\|_2.$$

For each j, note that if $|\widehat{\beta}_j^{(\ell-1)} - \beta_j^*| \geq \alpha_0\lambda$, $\lambda_j^{(\ell-1)} \leq \lambda \leq \alpha_0^{-1}|\widehat{\beta}_j^{(\ell-1)} - \beta_j^*|$; otherwise if $|\widehat{\beta}_j^{(\ell-1)} - \beta_j^*| \leq \alpha_0\lambda$, $\lambda_j^{(\ell-1)} \leq q_\lambda'((|\beta_j^*| - \alpha_0\lambda)_+)$ due to the monotonicity of q_λ'. Therefore, we have

$$\|\boldsymbol{\lambda}_{\mathcal{S}}^{(\ell-1)}\|_2 \leq \|q_\lambda'((|\boldsymbol{\beta}_{\mathcal{S}}^*| - \alpha_0\lambda)_+)\|_2 + \alpha_0^{-1}\|(\widehat{\boldsymbol{\beta}}^{(\ell-1)} - \boldsymbol{\beta}^*)_{\mathcal{S}}\|_2.$$

For $\|\boldsymbol{w}_{h,\mathcal{T}_\ell}^*\|_2$, it follows from the triangle inequality and (D.9) that

$$\begin{aligned}
\|\boldsymbol{w}_{h,\mathcal{T}_\ell}^*\|_2 &\leq \|\boldsymbol{w}_{h,\mathcal{S}}^*\|_2 + |\mathcal{T}_\ell \setminus \mathcal{S}|^{1/2}\|\boldsymbol{w}_{h,\mathcal{S}^c}^*\|_\infty \\
&< \|\boldsymbol{w}_{h,\mathcal{S}}^*\|_2 + \frac{1}{\alpha_0\lambda}\|\boldsymbol{w}_{h,\mathcal{S}^c}^*\|_\infty\|(\widehat{\boldsymbol{\beta}}^{(\ell-1)} - \boldsymbol{\beta}^*)_{\mathcal{T}_\ell\setminus\mathcal{S}}\|_2 \\
&\leq \|\boldsymbol{w}_{h,\mathcal{S}}^*\|_2 + 0.5q'(\alpha_0)\alpha_0^{-1}\|(\widehat{\boldsymbol{\beta}}^{(\ell-1)} - \boldsymbol{\beta}^*)_{\mathcal{T}_\ell\setminus\mathcal{S}}\|_2.
\end{aligned}$$

Using the elementary inequality $a + b\cdot c/2 \leq \sqrt{(1+c^2/4)(a^2+b^2)}$ for $a, b, c \geq 0$, we obtain

$$\|\boldsymbol{\lambda}_{\mathcal{S}}^{(\ell-1)}\|_2 + \|\boldsymbol{w}_{h,\mathcal{T}_\ell}^*\|_2 \leq \|q_\lambda'((|\boldsymbol{\beta}_{\mathcal{S}}^*| - \alpha_0\lambda)_+)\|_2 + \|\boldsymbol{w}_{h,\mathcal{S}}^*\|_2 + \frac{\sqrt{1+\{q'(\alpha_0)/2\}^2}}{\alpha_0}\|(\widehat{\boldsymbol{\beta}}^{(\ell-1)} - \boldsymbol{\beta}^*)_{\mathcal{T}_\ell}\|_2.$$

Substituting this bound into (D.7) yields

$$\begin{aligned}
&\|\widehat{\boldsymbol{\beta}}^{(\ell)} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} \\
&\leq \kappa^{-1}\gamma_p^{-1/2}\{\|q_\lambda'((|\boldsymbol{\beta}_{\mathcal{S}}^*| - \alpha_0\lambda)_+)\|_2 + \|\boldsymbol{w}_{h,\mathcal{S}}^*\|_2\} + \kappa^{-1}b_h^* + \delta\cdot\|\widehat{\boldsymbol{\beta}}^{(\ell-1)} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}},
\end{aligned}$$

where $\delta = \sqrt{1+\{q'(\alpha_0)/2\}^2}/(\alpha_0\kappa\gamma_p) \in (0,1)$ by (4.11). This proves (4.13). In conjunction with (D.4), the second bound (4.14) follows immediately. $\square$

D.2 Proof of Theorem 4.3

The proof is based on Theorem 4.2, in conjunction with Proposition 4.2 and Lemma C.2. With the stated choice of regularization parameter $\lambda \asymp \sigma_{\boldsymbol{x}}\sqrt{\tau(1-\tau)\log(p)/n}$ and bandwidth constraint, applying Lemma C.2 with $t = 2\log(2p)$ implies that $\|\boldsymbol{w}_h^*\|_\infty \leq 0.5q'(\alpha_0)\lambda$ holds with probability at least $1-(2p)^{-1}$ provided that $h \leq \sqrt{\tau(1-\tau)/\{(1+\tau)l_0\kappa_2\}}$ and $n \geq \frac{\max(\tau,1-\tau)^2}{\tau(1-\tau)}\log(p)$.

Next, we apply Proposition 4.2 to control the probability of the event $\mathcal{E}_{\rm rsc}(r, l, \kappa)$ from Theorem 4.2, where

$$r = \frac{h}{20v_0^2}, \quad l = \left\{\left(2 + \frac{2}{q'(\alpha_0)}\right)(c^2+1)^{1/2} + \frac{2}{q'(\alpha_0)}\right\}(s/\gamma_p)^{1/2}, \quad \kappa = \frac{\kappa_l f_l}{2},$$

and the constant $c > 0$ is determined by equation (4.12). Provided that $nh \gtrsim f_u f_l^{-2} s\log(p)$, Proposition 4.2 guarantees that event $\mathcal{E}_{\rm rsc}(r, l, \kappa)$ holds with probability at least $1-(2p)^{-1}$.

Moreover, recall from (D.8) that $r_{\rm opt} = \gamma_p^{1/2}\alpha_0 c s^{1/2}\lambda$ and $b_h^* \le l_0\kappa_2 h^2/2$. As long as the bandwidth h is such that $r_{\rm opt} \le r$ and $b_h^* \le (s/\gamma_p)^{1/2}\lambda$, we can apply Theorem 4.2 to conclude that, conditioned on $\mathcal{E}_{\rm rsc}(r,l,\kappa) \cap \{\|\boldsymbol{w}_h^*\|_\infty \le 0.5q'(\alpha_0)\lambda\}$,

$$\|\widehat{\boldsymbol{\beta}}^{(\ell)} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} \le \delta^{\ell-1} r_{\rm opt} + (1-\delta)^{-1}\kappa^{-1}[\gamma_p^{-1/2}\{\|q'_\lambda((\boldsymbol{\beta}^*_{\mathcal{S}} - \alpha_0\lambda)_+)\|_2 + \|\boldsymbol{w}^*_{h,\mathcal{S}}\|_2\} + 0.5 l_0\kappa_2 h^2] \quad \text{(D.10)}$$

for any $\ell \ge 2$, where $\delta = \sqrt{5}/(\alpha_0\kappa_l f_l \gamma_p) \in (0,1)$ and $\|q'_\lambda((\boldsymbol{\beta}^*_{\mathcal{S}} - \alpha_0\lambda)_+)\|_2 = 0$ under the stated beta-min condition. Applying Lemma D.3 to the oracle term $\|\boldsymbol{w}^*_{h,\mathcal{S}}\|_2$, we obtain that with probability at least $1-e^{-t}$,

$$\|\mathbf{S}^{-1/2}\boldsymbol{w}^*_{h,\mathcal{S}}\|_2 \lesssim \upsilon_0\left\{\sqrt{\tau(1-\tau)\frac{s+t}{n}} + \max(\tau, 1-\tau)\frac{s+t}{n}\right\}.$$

Finally, turning to the first term on the right-hand of (D.10), noting that $r_{\rm opt} \asymp (\kappa_l f_l)^{-1} s^{1/2}\lambda$, we have

$$\delta^{\ell-1} s^{1/2}\lambda \lesssim (\kappa_l f_l)^{-1}\delta^{\ell-1}\sqrt{\frac{s\log(p)}{n}} \lesssim (\kappa_l f_l)^{-1}\sqrt{\frac{s}{n}},$$

provided that $\ell \ge \lceil \log(\log p)/\log(1/\delta)\rceil$. Putting together the pieces yields the claimed bounds in (4.17). □

E Proofs of Results in Section 4.4

Since the strong oracle property concerns the closeness between the estimator and the oracle, we modify Lemma D.2 to obtain the following result. Recall that $\gamma_p = \gamma_p(\boldsymbol{\Sigma}) \in (0,1]$ is the minimum eigenvalue of $\boldsymbol{\Sigma}$.

Lemma E.1. Let $\mathcal{T}$ be a subset of $[p]$ such that $\mathcal{S} \subseteq \mathcal{T}$ and $k = |\mathcal{T}|$, and let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^{\rm T}$ satisfy $\|\boldsymbol{\lambda}\|_\infty \le \lambda$, $\|\boldsymbol{\lambda}_{\mathcal{T}^c}\|_{\min} \ge a\lambda$ for some $0 < a \le 1$ and $\lambda > 0$. Conditioned on $\{\|\boldsymbol{w}_h^{\rm ora}\|_\infty \le 0.5a\lambda\}$, any optimal solution $\widehat{\boldsymbol{\beta}}$ of (2.6) falls in the ℓ_1-cone $\widehat{\boldsymbol{\beta}}^{\rm ora} + \mathbb{C}_{\boldsymbol{\Sigma}}(l)$, where $l = (2+2/a)(k/\gamma_p)^{1/2}$. Moreover, let $r, \kappa > 0$ satisfy

$$\gamma_p^{-1/2}(0.5ak^{1/2} + s^{1/2})\lambda < r\cdot\kappa. \quad \text{(E.1)}$$

Then, conditioned on $\{\|\boldsymbol{w}_h^{\rm ora}\|_\infty \le 0.5a\lambda\} \cap \mathcal{G}_{\rm rsc}(r,l,\kappa)$,

$$\|\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\rm ora}\|_{\boldsymbol{\Sigma}} \le \kappa^{-1}\gamma_p^{-1/2}(\|\boldsymbol{w}^{\rm ora}_{h,\mathcal{T}}\|_2 + \|\boldsymbol{\lambda}_{\mathcal{S}}\|_2) \le \kappa^{-1}\gamma_p^{-1/2}(0.5ak^{1/2} + s^{1/2})\lambda. \quad \text{(E.2)}$$

E.1 Proof of Theorem 4.4

For $\ell = 1, 2, \ldots$, let $\mathcal{T}_\ell = \mathcal{S} \cup \{j \in [p] : \lambda_j^{(\ell-1)} < q'(\alpha_0)\lambda\}$ be the index sets given in (D.5). Recall that $l = \{2 + 2/q'(\alpha_0)\}(c_1^2+1)^{1/2}(s/\gamma_p)^{1/2}$, where $c_1 > 0$ is determined by equation (4.21). Conditioned on $\{\|\boldsymbol{w}_h^{\rm ora}\|_\infty \le 0.5q'(\alpha_0)\lambda\} \cap \{\|\widehat{\boldsymbol{\beta}}^{\rm ora} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} \le r\} \cap \mathcal{G}_{\rm rsc}(r,l,\kappa)$, applying Lemma E.1 with $a = q'(\alpha_0)/2$ and following the same argument as in the proof of Theorem 4.2, we obtain that $|\mathcal{T}_\ell| < (c_1^2+1)s$ and

$$\begin{aligned}\|\widehat{\boldsymbol{\beta}}^{(\ell)} - \widehat{\boldsymbol{\beta}}^{\rm ora}\|_{\boldsymbol{\Sigma}} &\le \kappa^{-1}\gamma_p^{-1/2}\{\|\boldsymbol{\lambda}_{\mathcal{S}}^{(\ell-1)}\|_2 + \|\boldsymbol{w}^{\rm ora}_{h,\mathcal{T}_\ell}\|_2\} \\ &< \kappa^{-1}\gamma_p^{-1/2}\{0.5q'(\alpha_0)(c_1^2+1)^{1/2} + 1\}s^{1/2}\lambda = \gamma_p^{1/2}\alpha_0 c_1 s^{1/2}\lambda \le r. \end{aligned} \quad \text{(E.3)}$$

Furthermore, define a sequence of index sets

$$\mathcal{S}_\ell = \{j \in [p] : |\widehat{\beta}_j^{(\ell)} - \beta_j^*| > \alpha_0 \lambda\}, \quad \ell = 0, 1, 2, \ldots.$$

Given the initialization $\widehat{\boldsymbol{\beta}}^{(0)} = \mathbf{0}$, the stated beta-min condition ensures $\mathcal{S}_0 = \mathcal{S}$.

In order to establish the equivalence between $\widehat{\boldsymbol{\beta}}^{(\ell)}$ and $\widehat{\boldsymbol{\beta}}^{\rm ora}$, we need to derive sharper bounds on $\|\boldsymbol{\lambda}_{\mathcal{S}}^{(\ell-1)}\|_2$ and $\|\boldsymbol{w}_{h,\mathcal{T}_\ell}^{\rm ora}\|_2$ in (E.3). For $\|\boldsymbol{\lambda}_{\mathcal{S}}^{(\ell-1)}\|_2$, from the monotonicity of q_λ' we see that $\lambda_j^{(\ell-1)} = q_\lambda'(|\widehat{\beta}_j^{(\ell-1)}|) \le q_\lambda'(|\beta_j^*| - \alpha_0\lambda)$ if $j \in \mathcal{S} \cap \mathcal{S}_{\ell-1}^{\rm c}$, and $\lambda_j^{(\ell-1)} \le \lambda$ for all the remaining j. Combined with the beta-min condition $\|\boldsymbol{\beta}_{\mathcal{S}}^*\|_{\min} \ge (\alpha_0 + \alpha_1)\lambda$, we obtain

$$\|\boldsymbol{\lambda}_{\mathcal{S}}^{(\ell-1)}\|_2 \le \|q_\lambda'(|\boldsymbol{\beta}_{\mathcal{S}}^*| - \alpha_0\lambda)\|_2 + \lambda|\mathcal{S} \cap \mathcal{S}_{\ell-1}|^{1/2} = \lambda|\mathcal{S} \cap \mathcal{S}_{\ell-1}|^{1/2}.$$

Turning to $\|\boldsymbol{w}_{h,\mathcal{T}_\ell}^{\rm ora}\|_2$, recall that $\boldsymbol{w}_{h,\mathcal{S}}^{\rm ora} = \mathbf{0}$ and hence

$$\|\boldsymbol{w}_{h,\mathcal{T}_\ell}^{\rm ora}\|_2 = \|\boldsymbol{w}_{h,\mathcal{T}_\ell \setminus \mathcal{S}}^{\rm ora}\|_2 \le \|\boldsymbol{w}_h^{\rm ora}\|_\infty |\mathcal{T}_\ell \setminus \mathcal{S}|^{1/2}.$$

For each $j \in \mathcal{T}_\ell \setminus \mathcal{S}$, $\beta_j^* = 0$ and $\lambda_j^{(\ell-1)} = q_\lambda'(|\widehat{\beta}_j^{(\ell-1)}|) < q'(\alpha_0)\lambda = q_\lambda'(\alpha_0\lambda)$. Hence, $|\widehat{\beta}_j^{(\ell-1)} - \beta_j^*| = |\widehat{\beta}_j^{(\ell-1)}| > \alpha_0\lambda$, indicating $j \in \mathcal{S}_{\ell-1} \setminus \mathcal{S}$. Therefore, we have $\mathcal{T}_\ell \setminus \mathcal{S} \subseteq \mathcal{S}_{\ell-1} \setminus \mathcal{S}$, from which it follows that

$$\|\boldsymbol{w}_{h,\mathcal{T}_\ell}^{\rm ora}\|_2 \le \|\boldsymbol{w}_h^{\rm ora}\|_\infty |\mathcal{S}_{\ell-1} \setminus \mathcal{S}|^{1/2}.$$

Since $\|\widehat{\boldsymbol{\beta}}^{(\ell)} - \widehat{\boldsymbol{\beta}}^{\rm ora}\|_{\boldsymbol{\Sigma}} \ge \gamma_p^{1/2} \|\widehat{\boldsymbol{\beta}}^{(\ell)} - \widehat{\boldsymbol{\beta}}^{\rm ora}\|_2$, substituting the above bounds into (E.3) yields

$$\begin{aligned}
\|\widehat{\boldsymbol{\beta}}^{(\ell)} - \widehat{\boldsymbol{\beta}}^{\rm ora}\|_2 &\le \{|\mathcal{S} \cap \mathcal{S}_{\ell-1}|^{1/2} + 0.5q'(\alpha_0)|\mathcal{S}_{\ell-1} \setminus \mathcal{S}|^{1/2}\}(\kappa\gamma_p)^{-1}\lambda \\
&\le \frac{\sqrt{1 + \{q'(\alpha_0)/2\}^2}}{\kappa\gamma_p} |\mathcal{S}_{\ell-1}|^{1/2}\lambda, \quad \text{for every } \ell \ge 1.
\end{aligned} \tag{E.4}$$

By (E.4), in order to prove $\widehat{\boldsymbol{\beta}}^{(\ell)} = \widehat{\boldsymbol{\beta}}^{\rm ora}$ for some sufficiently large ℓ, it suffices to show that the set $\mathcal{S}_{\ell-1}$ is empty. By the definition of $\mathcal{S}_\ell$, $\min_{j \in \mathcal{S}_\ell} |\widehat{\beta}_j^{(\ell)} - \widehat{\beta}_j^{\rm ora}| > \alpha_0\lambda - \|\widehat{\boldsymbol{\beta}}^{\rm ora} - \boldsymbol{\beta}^*\|_\infty$. Provided that

$$\|\widehat{\boldsymbol{\beta}}^{\rm ora} - \boldsymbol{\beta}^*\|_\infty \le \left[\alpha_0 - \frac{\sqrt{1 + \{q'(\alpha_0)/2\}^2}}{\delta\kappa\gamma_p}\right]\lambda,$$

we have

$$|\mathcal{S}_\ell|^{1/2} < \frac{\|(\widehat{\boldsymbol{\beta}}^{(\ell)} - \widehat{\boldsymbol{\beta}}^{\rm ora})_{\mathcal{S}}\|_2}{\alpha_0\lambda - \|\widehat{\boldsymbol{\beta}}^{\rm ora} - \boldsymbol{\beta}^*\|_\infty} \le \delta|\mathcal{S}_{\ell-1}|^{1/2}, \quad \ell \ge 1. \tag{E.5}$$

Since $\mathcal{S}_0 = \mathcal{S}$ with $|\mathcal{S}_0| = s$, we have $|\mathcal{S}_\ell|^{1/2} < \delta^\ell s^{1/2}$ for all $\ell \ge 1$. When $\ell \ge \lceil \log(s^{1/2})/\log(1/\delta) \rceil$, $|\mathcal{S}_\ell| < 1$ and hence $\mathcal{S}_\ell$ must be empty. Returning to the error bound (E.4), we conclude that $\widehat{\boldsymbol{\beta}}^{(\ell)} = \widehat{\boldsymbol{\beta}}^{\rm ora}$ for all $\ell \ge \lceil \log(s^{1/2})/\log(1/\delta) \rceil + 1$. This completes the proof. $\square$

E.2 Proof of Theorem 4.5

To apply the deterministic result in Theorem 4.4, we need the following two lemmas to control the probability of the events in (4.22). Specifically, Lemma E.2 ensures that the local RSC event $\mathcal{G}_{\rm rsc}(r, l, \kappa)$ holds with high probability, and Lemma E.3 characterizes all the stochastic quantities that involve the oracle estimator.

Lemma E.2. Let (r, l, h) satisfy

$$24\upsilon_1^2 r = h \le f_l/(2l_0) \quad \text{and} \quad nh \ge C f_u f_l^{-2} \max\{s, l^2 \log(p)\} \tag{E.6}$$

for some sufficiently large constant C depending only on $(\upsilon_1, \sigma_{\boldsymbol{x}})$. Then, the event $\mathcal{G}_{\rm rsc}(r, l, \kappa)$ holds with probability at least $1 - (2p)^{-1}$, where $\kappa = \kappa_l f_l/2$ and $\kappa_l = \min_{|u| \le 1} K(u)$.

Lemma E.3. Let $A_0 \ge 1$ be the constant in (4.26). For any $t \ge 0$, the oracle score $\boldsymbol{w}_h^{\rm ora} = \nabla \widehat{Q}_h(\widehat{\boldsymbol{\beta}}^{\rm ora}) \in \mathbb{R}^p$ and oracle estimator $\widehat{\boldsymbol{\beta}}^{\rm ora}$ satisfy the bounds

$$\|\boldsymbol{w}_h^{\rm ora}\|_\infty \lesssim \sqrt{\frac{\log(2p)}{n}} + A_0 \left\{ \sqrt{\frac{\log(s) + t}{n}} + \sqrt{\frac{s + \log(p)}{nh}} \sqrt{\frac{s + t}{n}} + h^2 \right\} \tag{E.7}$$

and

$$\|(\widehat{\boldsymbol{\beta}}^{\rm ora} - \boldsymbol{\beta}^*)_{\mathcal{S}}\|_\infty \lesssim \frac{s + t}{h^{1/2} n} + h^2 + \sqrt{\frac{\log(s) + t}{n}} \tag{E.8}$$

with probability at least $1 - p^{-1} - 5e^{-t}$, provided that the sample size n and bandwidth h are subject to $\max\{\sqrt{(s + t)/n}, \sqrt{\log(p)/n}\} \lesssim h \lesssim 1$.

Returning to the main thread, we are now ready to use Theorem 4.4 to establish the strong oracle property. Set $r = h/(24\upsilon_1^2)$, $l = \{2 + \frac{2}{q'(\alpha_0)}\}(c_1^2 + 1)^{1/2}(s/\gamma_p)^{1/2}$, $\kappa = \kappa_l f_l/2$, and choose the bandwidth $h \asymp \{\log(p)/n\}^{1/4}$ so that $r \asymp \{\log(p)/n\}^{1/4}$. Together, Lemma E.2 and Lemma E.3 with $t = \log(n)$ imply that, with probability at least $1 - 2p^{-1} - 5n^{-1}$, the following bounds

$$\|\boldsymbol{w}_h^{\rm ora}\|_\infty \lesssim \sqrt{\frac{\log(p)}{n}}, \quad \|\widehat{\boldsymbol{\beta}}^{\rm ora} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} \lesssim \sqrt{\frac{s + \log(n)}{n}} \quad \text{and} \quad \|\widehat{\boldsymbol{\beta}}^{\rm ora} - \boldsymbol{\beta}^*\|_\infty \lesssim \sqrt{\frac{\min\{s, \log(p)\}}{n}}$$

hold provided the sample size obeys $n \gtrsim \max\{s^{8/3}/(\log p)^{5/3}, \log(p)\}$.

Finally, as required by (4.22) in Theorem 4.4, if we choose the regularization parameter $\lambda = C\sqrt{\log(p)/n}$ for a sufficiently large C, then the events in (4.22) hold with probability at least $1 - 2p^{-1} - 5n^{-1}$ under the scaling $n \gtrsim \max\{s^{8/3}/(\log p)^{5/3}, s^2 \log(p)\}$. We have thus verified all the requirements in Theorem 4.4, hence certifying the strong oracle property. □

F Proofs of Results in Section A

F.1 Proof of Proposition A.1

By (4.1) and the uniqueness of b_h, $m''(\alpha) = \int_{-\infty}^{\infty} K(u) f_\varepsilon(\alpha - hu)\, du$ satisfies that $m''(b_h) > 0$. Recall that β_1^* denotes the intercept and $\boldsymbol{x} = (1, \boldsymbol{x}_-^{\rm T})^{\rm T}$ with $\boldsymbol{x}_- \in \mathbb{R}^{p-1}$. For any $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\rm T} \in \mathbb{R}^p$,

$$\begin{aligned} Q_h(\boldsymbol{\beta}) &= \mathbb{E}\ell_h(\varepsilon - (\beta_1 - \beta_1^*) - \boldsymbol{x}_-^{\rm T}(\boldsymbol{\beta}_- - \boldsymbol{\beta}_-^*)) \\ &= \mathbb{E}_{\boldsymbol{x}}\mathbb{E}\{\ell_h(\varepsilon - (\beta_1 - \beta_1^*) - \boldsymbol{x}_-^{\rm T}(\boldsymbol{\beta}_- - \boldsymbol{\beta}_-^*))|\boldsymbol{x}\} \\ &= \mathbb{E}\{m(\beta_1 - \beta_1^* + \boldsymbol{x}_-^{\rm T}(\boldsymbol{\beta}_- - \boldsymbol{\beta}_-^*))\} \\ &\ge m(b_h) = \mathbb{E}\ell_h(\varepsilon - b_h) = Q_h(\boldsymbol{\beta}^\star), \end{aligned} \tag{F.1}$$

where $\boldsymbol{\beta}^\star = (\beta_1^* + b_h, \boldsymbol{\beta}_-^{*\rm T})^{\rm T} \in \mathbb{R}^p$. This implies that $Q_h(\boldsymbol{\beta}^\star) = \min_{\boldsymbol{\beta} \in \mathbb{R}^p} Q_h(\boldsymbol{\beta})$. Furthermore, compute the Hessian matrix $\nabla^2 Q_h(\boldsymbol{\beta}) = \mathbb{E}\{K_h(\boldsymbol{x}^{\rm T}(\boldsymbol{\beta} - \boldsymbol{\beta}^*) - \varepsilon)\boldsymbol{x}\boldsymbol{x}^{\rm T}\}$. In particular, $\nabla^2 Q_h(\boldsymbol{\beta}^\star) =$

$\mathbb{E}\{K_h(b_h - \varepsilon)\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\} = m''(b_h)\mathbb{E}(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})$ is positive definite, so that $\boldsymbol{\beta}^\star$ is the unique minimizer of $\boldsymbol{\beta} \mapsto Q_h(\boldsymbol{\beta})$. This ensures that $\boldsymbol{\beta}_h^* = \boldsymbol{\beta}^\star$, as claimed. □

Next we characterize the order of b_h as a function of h. Similarly to the proof of Proposition 4.1, we define $\widetilde{b}$ as follows: if $|b_h| \le \kappa_2^{1/2} h$, set $\widetilde{b} = b_h$; otherwise if $|b_h| > h$, set $\widetilde{b} = \eta b_h$ for some $\eta \in (0,1)$ so that $|\widetilde{b}| = \kappa_2^{1/2} h$. By (B.1),

$$0 \le \{m'(\widetilde{b}) - m'(0)\}\widetilde{b} \le \{m'(b_h) - m'(0)\}\widetilde{b} = -m'(0)\widetilde{b}.$$

For the left-hand side,

$$\begin{aligned} m'(\widetilde{b}) - m'(0) &= \int_0^{\widetilde{b}} m''(t)\,\mathrm{d}t = \int_0^{\widetilde{b}} \int_{-\infty}^{\infty} K(u) f_\varepsilon(t - hu)\,\mathrm{d}u\,\mathrm{d}t \\ &= f_\varepsilon(0)\widetilde{b} + \int_0^{\widetilde{b}} \int_{-\infty}^{\infty} K(u)\{f_\varepsilon(t-hu) - f_\varepsilon(0)\}\,\mathrm{d}u\,\mathrm{d}t, \end{aligned}$$

implying

$$\{m'(\widetilde{b}) - m'(0)\}\widetilde{b} \ge f_\varepsilon(0)\widetilde{b}^2 - \frac{l_0}{2}|\widetilde{b}|^3 - l_0\kappa_1 h \cdot \widetilde{b}^2.$$

For the right-hand side, we have $|m'(0)| = |\int_{-\infty}^{\infty} K(u)\{F_\varepsilon(-hu) - F_\varepsilon(0)\}\,\mathrm{d}u| \le \frac{l_0}{2}\kappa_2 h^2$. Combining the above upper and lower bounds, we find that

$$f_\varepsilon(0)\widetilde{b}^2 \le \frac{l_0}{2}\kappa_2 h^2 |\widetilde{b}| + \frac{l_0}{2}|\widetilde{b}|^3 + l_0\kappa_1 h\widetilde{b}^2 \le (\kappa_2 + \kappa_1\kappa_2^{1/2}) l_0 h^2 |\widetilde{b}|,$$

where the first inequality uses the fact that $|\widetilde{b}| \le \kappa_2^{1/2} h$. Canceling $|\widetilde{b}|$ gives

$$|\widetilde{b}| \le \underbrace{(\kappa_2^{1/2} + \kappa_1)}_{=c_1} \frac{l_0 h}{f_\varepsilon(0)} \cdot \kappa_2^{1/2} h.$$

As long as $c_1 l_0 h < f_\varepsilon(0)$, the above inequality implies $|\widetilde{b}| < \kappa_2^{1/2} h$. By the definition of $\widetilde{b}$, we must have $\widetilde{b} = b_h$; otherwise $|\widetilde{b}| = \kappa_2^{1/2} h$ which leads to contradiction. This completes the proof of (A.1). □

F.2 Proof of Theorem A.1

By the definition of $\boldsymbol{\beta}_h^*$, we have $\nabla Q_h(\boldsymbol{\beta}_h^*) = \mathbf{0}$. Replacing $\boldsymbol{\beta}^*$ by $\boldsymbol{\beta}_h^*$ in (4.10), the smoothing error term b_h^* now becomes zero. Modifying the proof of Theorem 4.2 accordingly, the conclusions therein remain valid, but now with $b_h^* = 0$ and

$$\boldsymbol{w}_h^* = \boldsymbol{w}_h(\boldsymbol{\beta}_h^*) = \frac{1}{n}\sum_{i=1}^n \{\bar{K}((b_h - \varepsilon_i)/h) - \tau\}\boldsymbol{x}_i.$$

Once again, the key is to show that event $\mathcal{E}_{\mathrm{rsc}}(r, l, \kappa) \cap \{\|\boldsymbol{w}_h^*\|_\infty \le 0.5q'(\alpha_0)\lambda\}$ holds with high probability, where $\mathcal{E}_{\mathrm{rsc}}$ is given in (4.7) with $\boldsymbol{\beta}^*$ replaced by $\boldsymbol{\beta}_h^*$.

Proceed similarly to the proof of Lemma C.2 and Lemma D.3, we obtain that with probability at least $1 - (2p)^{-1}$,

$$\|\boldsymbol{w}_h^*\|_\infty \lesssim \sigma_{\boldsymbol{x}}\left\{\sqrt{\frac{\log(2p)}{n}} + \frac{\log(2p)}{n}\right\}, \tag{F.2}$$

and for any $t > 0$,

$$\|\mathbf{S}^{-1/2}\boldsymbol{w}^*_{h,\mathcal{S}}\|_2 \lesssim \sqrt{\frac{s+t}{n}} + \frac{s+t}{n} \tag{F.3}$$

holds with probability at least $1 - e^{-t}$,

Next, in order to show that Proposition 4.2 remains valid if $\boldsymbol{\beta}^*$ is replaced by $\boldsymbol{\beta}^*_h$, it suffices to change the definition of the event E_i in (C.4) to

$$E_i = \{|\varepsilon_i - b_h| \le h/2\} \cap \left\{ \frac{|\langle \boldsymbol{x}_i, \boldsymbol{\beta} - \boldsymbol{\beta}^*_h \rangle|}{\|\boldsymbol{\beta} - \boldsymbol{\beta}^*_h\|_{\boldsymbol{\Sigma}}} \le \frac{h}{2r} \right\}.$$

Moreover, note that

$$|\mathbb{E}\mathbb{1}(|\varepsilon_i - b_h| \le h/2) - hf_0| \le \int_{-h/2}^{h/2} |f_\varepsilon(t + b_h) - f_\varepsilon(0)| \, \mathrm{d}t \le l_0 h^2/4 + l_0 b_h h.$$

Keep all other statements the same, we obtain that with probability at least $1 - (2p)^{-1}$, the event $\mathcal{E}_{\rm rsc}(r, l, \kappa)$ with $r = h/(20\upsilon_0^2)$, $l = \{(2 + \frac{2}{q'(\alpha_0)})(c^2+1)^{1/2} + \frac{2}{q'(\alpha_0)}\}(s/\gamma_p)^{1/2}$ and $\kappa = \kappa_l f_0/2$ holds as long as $\sigma^2_{\boldsymbol{x}} f_0^{-1} s \log(p)/n \lesssim h \lesssim f_0$.

With a penalty level $\lambda \asymp \sigma_{\boldsymbol{x}} \sqrt{\log(p)/n}$, we conclude from Theorem 4.2 that with probability at least $1 - p^{-1} - e^{-t}$,

$$\|\widehat{\boldsymbol{\beta}}^{(\ell)} - \boldsymbol{\beta}^*_h\|_{\boldsymbol{\Sigma}} \lesssim \delta^{\ell-1} f_0^{-1} \sqrt{\frac{s\log(p)}{n}} + (1-\delta)^{-1} f_0^{-1} \sqrt{\frac{s+t}{n}}$$

holds for every $\ell \ge 2$, provided that $n \gtrsim s\log(p) + t$ and $\sigma_{\boldsymbol{x}} f_0^{-1} \sqrt{s\log(p)/n} \lesssim h \lesssim f_0$, where $\delta = \sqrt{4 + \{q'(\alpha_0)\}^2/(\alpha_0\kappa_l f_0 \gamma_p)}$. This completes the proof by letting $\ell \ge \lceil \log(\log p)/\log(1/\delta) \rceil$. □

F.3 Proof of Proposition A.2

Define the oracle smoothed quantile loss and its population counterpart as

$$\widehat{Q}^{\rm ora}_h(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \ell_h(y_i - \boldsymbol{x}^{\rm T}_{i,\mathcal{S}}\boldsymbol{\beta}) \quad \text{and} \quad Q^{\rm ora}_h(\boldsymbol{\beta}) = \mathbb{E}\widehat{Q}^{\rm ora}_h(\boldsymbol{\beta}), \ \ \boldsymbol{\beta} \in \mathbb{R}^s.$$

With some abuse of notation, we write $\boldsymbol{\beta}^*_h = \boldsymbol{\beta}^*_{h,\mathcal{S}} \in \mathbb{R}^s$ and $\widehat{\boldsymbol{\beta}}^{\rm ora} \in \mathrm{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^s} \widehat{Q}^{\rm ora}_h(\boldsymbol{\beta})$. The concentration bound (A.2) follows from the same argument that was used to prove (4.24). In particular, the restricted strong convexity of $\widehat{Q}^{\rm ora}_h$ around $\boldsymbol{\beta}^*_h$ is established similarly as in the proof of Theorem A.1, and the h^2-term vanishes because $\nabla Q^{\rm ora}_h(\boldsymbol{\beta}^*_h) = \mathbf{0}$.

To prove (A.3), define the stochastic process

$$\Delta(\boldsymbol{\beta}) = \mathbf{S}^{-1/2}\{\nabla\widehat{Q}^{\rm ora}_h(\boldsymbol{\beta}) - \nabla\widehat{Q}^{\rm ora}_h(\boldsymbol{\beta}^*_h) - \mathbf{D}_h(\boldsymbol{\beta} - \boldsymbol{\beta}^*_h)\}, \ \ \boldsymbol{\beta} \in \mathbb{R}^s, \tag{F.4}$$

where by the independence of $\boldsymbol{x}$ and ε, $\mathbf{D}_h = \nabla^2 Q^{\rm ora}_h(\boldsymbol{\beta}^*_h) = m''(b_h)\cdot\mathbf{S}$. We will bound the supremum $\sup_{\boldsymbol{\beta}\in\boldsymbol{\beta}^*_h + \mathbb{B}_{\mathbf{S}}(r)} \|\Delta(\boldsymbol{\beta}) - \mathbb{E}\Delta(\boldsymbol{\beta})\|_2$ using the same argument as in the proof of Theorem 4.2 in He *et al.* (2021). It then suffices to evaluate $\mathbb{E}\Delta(\boldsymbol{\beta})$. By the mean value theorem for vector-valued functions,

$$\begin{aligned}\mathbb{E}\Delta(\boldsymbol{\beta}) &= \mathbf{S}^{-1/2}\int_0^1 \nabla^2 Q^{\rm ora}_h((1-t)\boldsymbol{\beta}^*_h + t\boldsymbol{\beta})\,\mathrm{d}t\,(\boldsymbol{\beta} - \boldsymbol{\beta}^*_h) - \mathbf{S}^{-1/2}\mathbf{D}_h(\boldsymbol{\beta} - \boldsymbol{\beta}^*_h) \\ &= \left\{\mathbf{S}^{-1/2}\int_0^1 \nabla^2 Q^{\rm ora}_h((1-t)\boldsymbol{\beta}^*_h + t\boldsymbol{\beta})\,\mathrm{d}t\,\mathbf{S}^{-1/2} - m''(b_h)\mathbf{I}_s\right\}\mathbf{S}^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}^*_h). \end{aligned} \tag{F.5}$$

Note that, for every $\boldsymbol{\beta} \in \mathbb{R}^s$,

$$\nabla^2 Q_h^{\mathrm{ora}}(\boldsymbol{\beta}) = \mathbb{E}\{K_h(\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\beta} - y)\boldsymbol{x}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\} = \mathbb{E}\bigg\{\int_{-\infty}^{\infty} K(u) f_\varepsilon(\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}(\boldsymbol{\beta} - \boldsymbol{\beta}_h^*) + b_h - hu)\,\mathrm{d}u \cdot \boldsymbol{x}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\bigg\}.$$

Moreover, write $\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}_h^*$ for $\boldsymbol{\beta} \in \boldsymbol{\beta}_h^* + \mathbb{B}_{\mathbf{S}}(r)$ so that

$$\nabla^2 Q_h^{\mathrm{ora}}((1-t)\boldsymbol{\beta}_h^* + t\boldsymbol{\beta}) = \mathbb{E}\bigg\{\int_{-\infty}^{\infty} K(u) f_\varepsilon(\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta} \cdot t + b_h - hu)\,\mathrm{d}u \cdot \boldsymbol{x}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\bigg\}.$$

Consequently, for any $t \in [0, 1]$,

$$\begin{aligned}
&\Big\|\mathbf{S}^{-1/2}\nabla^2 Q_h^{\mathrm{ora}}((1-t)\boldsymbol{\beta}_h^* + t\boldsymbol{\beta})\,\mathbf{S}^{-1/2} - m''(b_h)\mathbf{I}_s\Big\|_2 \\
&= \bigg\|\mathbf{S}^{-1/2}\mathbb{E}\bigg[\int K(u)\{f_\varepsilon(\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta} \cdot t + b_h - hu) - f_\varepsilon(b_h - hu)\}\,\mathrm{d}u\,\boldsymbol{x}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\bigg]\mathbf{S}^{-1/2}\bigg\|_2 \\
&\le l_0 t \cdot \sup_{\|\boldsymbol{u}\|_2=1} \mathbb{E}\{\langle \boldsymbol{x}_{\mathcal{S}}, \mathbf{S}^{-1/2}\boldsymbol{u}\rangle^2 |\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta}|\} \\
&\le l_0 t \cdot \bigg(\sup_{\|\boldsymbol{u}\|_2=1} \mathbb{E}\langle \boldsymbol{x}_{\mathcal{S}}, \mathbf{S}^{-1/2}\boldsymbol{u}\rangle^4\bigg)^{1/2}\Big(\mathbb{E}\langle \boldsymbol{x}_{\mathcal{S}}, \boldsymbol{\delta}\rangle^2\Big)^{1/2} \\
&\le \mu_4^{1/2} l_0 r t,
\end{aligned}$$

where $\mu_4 = \sup_{\boldsymbol{u} \in \mathbb{S}^{p-1}} \mathbb{E}\langle \boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}, \boldsymbol{u}\rangle^4$. Together with (F.5), this leads to

$$\sup_{\boldsymbol{\beta} \in \boldsymbol{\beta}_h^* + \mathbb{B}_{\mathbf{S}}(r)} \|\mathbb{E}\Delta(\boldsymbol{\beta})\|_2 \le \frac{l_0}{2}\mu_4^{1/2} r^2. \tag{F.6}$$

Furthermore, observe that

$$m''(b_h) = \mathbb{E}\{K_h(b_h - \varepsilon)\} = \int_{-\infty}^{\infty} K(u) f_\varepsilon(b_h - hu)\,\mathrm{d}u \ge f_\varepsilon(0) - l_0(b_h + \kappa_1 h) \ge \frac{1}{2} f_\varepsilon(0),$$

where the last inequality holds provided that h is sufficiently small. Combining this with (F.6) and (B.31) of He *et al.* (2021), we conclude that for any $r, t > 0$,

$$\sup_{\boldsymbol{\beta} \in \boldsymbol{\beta}_h^* + \mathbb{B}_{\mathbf{S}}(r)} \Big\|\mathbf{S}^{-1/2}\{\nabla \widehat{Q}_h^{\mathrm{ora}}(\boldsymbol{\beta}) - \nabla \widehat{Q}_h^{\mathrm{ora}}(\boldsymbol{\beta}_h^*) - \mathbf{D}_h(\boldsymbol{\beta} - \boldsymbol{\beta}_h^*)\}\Big\|_2 \lesssim \bigg(\sqrt{\frac{s+t}{nh}} + r\bigg) r \tag{F.7}$$

with probability at least $1 - e^{-t}$ as long as $\sqrt{(s+t)/n} \lesssim h \lesssim 1$. Taking $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}^{\mathrm{ora}}$ and $r \asymp \sqrt{(s+t)/n}$, (A.3) follows from (A.2) and the fact that $\nabla \widehat{Q}_h^{\mathrm{ora}}(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}) = \mathbf{0}$. □

F.4 Proof of Theorem A.2

Similarly to the proof of Theorem 4.5, the proof of Theorem A.2 is based on Lemmas E.2 and E.3 with slight modifications. In the proof of Lemma E.2, change the event E_i used in (G.19) to

$$E_i = \{|\varepsilon_i - b_h| \le h/2\} \cap \{|\langle \boldsymbol{x}_i, \boldsymbol{\beta}_2 - \boldsymbol{\beta}_h^*\rangle| \le h/4\} \cap \{|\langle \boldsymbol{x}_i, \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\rangle| \le \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|_{\boldsymbol{\Sigma}} \cdot h/(4r)\},$$

and keep all other arguments, the conclusions of Lemma E.2 remain valid.

Recall that $\boldsymbol{w}_h^{\mathrm{ora}} = \nabla \widehat{Q}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}})$. The following result refines Lemma E.3 under the additional independence assumption.

Lemma F.1. Let $A_1 \geq 1$ be the constant in (A.4). For any $t \geq 0$, the oracle score $\boldsymbol{w}_h^{\rm ora} = \nabla \widehat{Q}_h(\widehat{\boldsymbol{\beta}}^{\rm ora}) \in \mathbb{R}^p$ and oracle estimator $\widehat{\boldsymbol{\beta}}^{\rm ora}$ satisfy the bounds

$$\|\boldsymbol{w}_h^{\rm ora}\|_\infty \lesssim \sqrt{\frac{\log(p)}{n}} + A_1 \left\{ \sqrt{\frac{\log(s)+t}{n}} + \frac{(s+t)^{1/2}(s+\log p)^{1/2}}{h^{1/2} n} \right\} \tag{F.8}$$

and

$$\|\widehat{\boldsymbol{\beta}}^{\rm ora} - \boldsymbol{\beta}^*\|_\infty \lesssim \frac{s+t}{h^{1/2}n} + \sqrt{\frac{\log(s)+t}{n}} \tag{F.9}$$

with probability at least $1 - p^{-1} - 5e^{-t}$, provided that the sample size n and bandwidth h are subject to $\sqrt{(s+t)/n} \lesssim h \lesssim 1$ and $h \gtrsim \sqrt{(s+\log p)/n}$.

The rest of the proof then proceeds similarly to the proof of Theorem 4.5, and thus is omitted. □

G Proof of Auxiliary Lemmas

G.1 Proof of Lemma C.1

Condition (B3) ensures that $\mathbb{P}(|\boldsymbol{z}^{\rm T}\boldsymbol{u}| \geq \upsilon_0 t) \leq e^{-t}$ for any $t \geq 0$ and $\boldsymbol{u} \in \mathbb{S}^{p-1}$. For any $k \geq 1$, this implies

$$\mathbb{E}|\boldsymbol{z}^{\rm T}\boldsymbol{u}|^k = \upsilon_0^k k \int_0^\infty t^{k-1}\mathbb{P}(|\boldsymbol{z}^{\rm T}\boldsymbol{u}| \geq \upsilon_0 t)\,{\rm d}t \leq \upsilon_0^k k \int_0^\infty t^{k-1} e^{-t}\,{\rm d}t = k!\upsilon_0^k.$$

Taking the supremum over $\boldsymbol{u} \in \mathbb{S}^{p-1}$ proves the claimed bound. □

G.2 Proof of Lemma C.2

To facilitate the proof, let $\xi_i = \bar{K}(-\varepsilon_i/h) - \tau$. Taking $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ in the gradient function (4.1) yields

$$\nabla \widehat{Q}_h(\boldsymbol{\beta}^*) = \frac{1}{n}\sum_{i=1}^n \{\bar{K}(-\varepsilon_i/h) - \tau\}\boldsymbol{x}_i = \frac{1}{n}\sum_{i=1}^n \xi_i \boldsymbol{x}_i.$$

The upper bound for $\|\nabla \widehat{Q}_h(\boldsymbol{\beta}^*) - \nabla Q_h(\boldsymbol{\beta}^*)\|_\infty = \|(1/n)\sum_{i=1}^n \{\xi_i\boldsymbol{x}_i - \mathbb{E}(\xi_i\boldsymbol{x}_i)\}\|_\infty$ involves two quantities that are related to the kernel function: $\mathbb{E}\{\bar{K}^2(-\varepsilon/h)|\boldsymbol{x}\}$ and $\mathbb{E}\{\bar{K}(-\varepsilon/h)|\boldsymbol{x}\}$. We start with obtaining an upper bound for $\mathbb{E}\{\bar{K}^2(-\varepsilon/h)|\boldsymbol{x}\}$. By a change of variable and integration by parts, we obtain

$$\begin{aligned}
\mathbb{E}\{\bar{K}^2(-\varepsilon/h)|\boldsymbol{x}\} &= \int_{-\infty}^\infty \bar{K}^2(-u/h) f_{\varepsilon|\boldsymbol{x}}(u)\,{\rm d}u \\
&= h\int_{-\infty}^\infty \bar{K}^2(v) f_{\varepsilon|\boldsymbol{x}}(-vh)\,{\rm d}v \\
&= 2\int_{-\infty}^\infty K(v)\bar{K}(v) F_{\varepsilon|\boldsymbol{x}}(-vh)\,{\rm d}v.
\end{aligned} \tag{G.1}$$

By the fundamental theorem of calculus and the fact that $F_{\varepsilon|\boldsymbol{x}}(0) = \tau$, we have

$$\begin{aligned}
F_{\varepsilon|\boldsymbol{x}}(-vh) &= F_{\varepsilon|\boldsymbol{x}}(0) + \int_0^{-vh} f_{\varepsilon|\boldsymbol{x}}(t)\,{\rm d}t \\
&= \tau + (-hv) f_{\varepsilon|\boldsymbol{x}}(0) + \int_0^{-vh} \{f_{\varepsilon|\boldsymbol{x}}(t) - f_{\varepsilon|\boldsymbol{x}}(0)\}\,{\rm d}t.
\end{aligned} \tag{G.2}$$

Moreover, it can be shown that

$$a_K := \int_{-\infty}^{\infty} vK(v)\bar{K}(v)\,\mathrm{d}v = \int_{0}^{\infty} K(v)\{1-K(v)\}\,\mathrm{d}v > 0 \quad \text{and} \quad a_K \le \kappa_1, \tag{G.3}$$

where $\kappa_1 = \int |u|K(u)\,\mathrm{d}u$.

Substituting (G.2) into (G.1), and by (G.3), we obtain

$$\begin{aligned}
&\mathbb{E}\{\bar{K}^2(-\varepsilon/h)|\boldsymbol{x}\}\\
&= 2\tau \int_{-\infty}^{\infty} K(v)\bar{K}(v)\,\mathrm{d}v - 2hf_{\varepsilon|\boldsymbol{x}}(0)\int_{-\infty}^{\infty} vK(v)\bar{K}(v)\,\mathrm{d}v\\
&\quad + 2\int_{-\infty}^{\infty}\int_{0}^{-vh}\{f_{\varepsilon|\boldsymbol{x}}(t) - f_{\varepsilon|\boldsymbol{x}}(0)\}K(v)\bar{K}(v)\,\mathrm{d}t\,\mathrm{d}v\\
&\le \tau - 2a_K h f_{\varepsilon|\boldsymbol{x}}(0) + l_0h^2\cdot\int_{-\infty}^{\infty} v^2K(v)\bar{K}(v)\,\mathrm{d}v\\
&\le \tau + l_0\kappa_2h^2,
\end{aligned} \tag{G.4}$$

where the first inequality holds using the Lipschitz condition on $f_{\varepsilon|\boldsymbol{x}}$ in Condition (B1), and the last inequality holds by Condition (B2) on the kernel function. Through a similar calculation, $\mathbb{E}\{\bar{K}(-\varepsilon/h)|\boldsymbol{x}\} = \tau + \int_{-\infty}^{\infty}\int_{0}^{-hv}\{f_{\varepsilon|\boldsymbol{x}}(t) - f_{\varepsilon|\boldsymbol{x}}(0)\}K(v)\,\mathrm{d}t\,\mathrm{d}v$. The Lipschitz condition on $f_{\varepsilon|\boldsymbol{x}}$ then ensures that

$$\left|\mathbb{E}\{\bar{K}(-\varepsilon/h)|\boldsymbol{x}\} - \tau\right| \le \frac{l_0}{2}\kappa_2h^2. \tag{G.5}$$

Together, (G.4) and (G.5) imply

$$\mathbb{E}(\xi^2|\boldsymbol{x}) \le \tau(1-\tau) + (\tau+1)l_0\kappa_2h^2 = \tau(1-\tau) + Ch^2, \tag{G.6}$$

where $\xi = \bar{K}(-\varepsilon/h) - \tau$ and $C = (\tau+1)l_0\kappa_2$.

With the above preparations, we are now ready to prove (D.3). First, we use Bernstein's inequality to bound each $(1/n)\sum_{i=1}^n\{\xi_i x_{ij} - \mathbb{E}(\xi_i x_{ij})\}$, and then apply a union bound over $j = 1,\ldots,p$. Note that $\xi_1x_{1j} - \mathbb{E}(\xi_1x_{1j}),\ldots,\xi_nx_{nj} - \mathbb{E}(\xi_nx_{nj})$ are independent zero-mean random variables, and by (G.6),

$$\mathbb{E}(\xi_ix_{ij})^2 = \mathbb{E}_{\boldsymbol{x}}\{x_{ij}^2\cdot\mathbb{E}(\xi_i^2|\boldsymbol{x}_i)\} \le \tau(1-\tau)\sigma_{jj} + C\sigma_{jj}h^2.$$

Under Condition (B3), we have $\mathbb{P}(|x_{ij}| \ge \sigma_{jj}^{1/2}\upsilon_0t) \le e^{-t}$ for all $t \ge 0$. Noting that $|\xi_i| \le \max(1-\tau,\tau)$, we have for $k = 2,3,\ldots$,

$$\begin{aligned}
\mathbb{E}(|\xi_ix_{ij}|^k) &\le \max(1-\tau,\tau)^{k-2}\,\mathbb{E}_{\boldsymbol{x}}\{|x_{ij}|^k\cdot\mathbb{E}(\xi_i^2|\boldsymbol{x}_i)\}\\
&\le \max(1-\tau,\tau)^{k-2}\{\tau(1-\tau)+Ch^2\}\upsilon_0^k\sigma_{jj}^{k/2}\int_0^{\infty}\mathbb{P}(|\sigma_{jj}^{-1/2}x_{ij}| \ge \upsilon_0t)kt^{k-1}\,\mathrm{d}t\\
&\le \max(1-\tau,\tau)^{k-2}\{\tau(1-\tau)+Ch^2\}\upsilon_0^k\sigma_{jj}^{k/2}k\int_0^{\infty}t^{k-1}e^{-t}\,\mathrm{d}t\\
&= k!\max(1-\tau,\tau)^{k-2}\{\tau(1-\tau)+Ch^2\}\upsilon_0^k\sigma_{jj}^{k/2}\\
&\le \frac{k!}{2}\cdot\{\tau(1-\tau)+Ch^2\}\upsilon_0^2\sigma_{jj}\cdot\{2\max(1-\tau,\tau)\upsilon_0\sigma_{jj}^{1/2}\}^{k-2}.
\end{aligned} \tag{G.7}$$

Consequently, it follows from Bernstein's inequality that for every $t \geq 0$,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\{\xi_i x_{ij} - \mathbb{E}(\xi_i x_{ij})\}\right| \leq \upsilon_0 \sigma_{jj}^{1/2}\left[\sqrt{\{\tau(1-\tau) + Ch^2\}\frac{2t}{n}} + \max(1-\tau,\tau)\frac{2t}{n}\right]$$

with probability at least $1-2e^{-t}$. Finally, we apply a union bound to reach the conclusion (D.3). □

G.3 Proof of Lemma D.1

Since the objective function in (2.7) is convex, by the first-order optimality condition, there exists a subgradient $\widehat{\boldsymbol{g}} \in \partial\|\widehat{\boldsymbol{\beta}}\|_1$ such that $\nabla\widehat{Q}_h(\widehat{\boldsymbol{\beta}}) + \boldsymbol{\lambda}\circ\widehat{\boldsymbol{g}} = \mathbf{0}$. Using the fact that the subdifferential of a convex function is monotone increasing, for any $\boldsymbol{\beta}\in\mathbb{R}^p$, we have

$$\begin{aligned}
0 &= \langle\nabla\widehat{Q}_h(\widehat{\boldsymbol{\beta}}) + \boldsymbol{\lambda}\circ\widehat{\boldsymbol{g}}, \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rangle \\
&= \underbrace{\langle\nabla\widehat{Q}_h(\widehat{\boldsymbol{\beta}}) - \nabla\widehat{Q}_h(\boldsymbol{\beta}), \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rangle}_{\geq 0} + \langle\nabla\widehat{Q}_h(\boldsymbol{\beta}), \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rangle + \langle\boldsymbol{\lambda}\circ\widehat{\boldsymbol{g}}, \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rangle \\
&\geq \langle\nabla\widehat{Q}_h(\boldsymbol{\beta}) - \nabla Q_h(\boldsymbol{\beta}), \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rangle + \langle\nabla Q_h(\boldsymbol{\beta}), \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rangle + \langle\boldsymbol{\lambda}\circ\widehat{\boldsymbol{g}}, \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rangle \\
&\geq -\|\boldsymbol{w}_h(\boldsymbol{\beta})\|_\infty\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1 - b_h(\boldsymbol{\beta})\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_{\boldsymbol{\Sigma}} + \langle\boldsymbol{\lambda}\circ\widehat{\boldsymbol{g}}, \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rangle.
\end{aligned}$$

In addition, by the definition of the subgradient, $\langle\boldsymbol{\lambda}\circ\widehat{\boldsymbol{g}}, \widehat{\boldsymbol{\beta}}\rangle = \|\boldsymbol{\lambda}\circ\widehat{\boldsymbol{\beta}}\|_1$. Thus, for any $\boldsymbol{\beta}\in\mathbb{R}^p$ satisfying $\boldsymbol{\beta}_{\mathcal{T}^c} = \mathbf{0}$, we can decompose $\langle\boldsymbol{\lambda}\circ\widehat{\boldsymbol{g}}, \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rangle$ according to the subset $\mathcal{T}\subseteq[p]$ as

$$\begin{aligned}
\langle\boldsymbol{\lambda}\circ\widehat{\boldsymbol{g}}, \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rangle &= \langle(\boldsymbol{\lambda}\circ\widehat{\boldsymbol{g}})_{\mathcal{T}^c}, (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{\mathcal{T}^c}\rangle + \langle(\boldsymbol{\lambda}\circ\widehat{\boldsymbol{g}})_{\mathcal{T}}, (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{\mathcal{T}}\rangle \\
&\geq \|\boldsymbol{\lambda}_{\mathcal{T}^c}\|_{\min}\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{\mathcal{T}^c}\|_1 - \|\boldsymbol{\lambda}_{\mathcal{T}}\|_\infty\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})_{\mathcal{T}}\|_1.
\end{aligned}$$

Re-arranging the terms leads to the stated result. □

G.4 Proof of Lemma D.2

The proof of Lemma D.2 is based on Lemma D.1 and a similar localized analysis used in the proof of Theorem 4.1. Define an intermediate vector $\widetilde{\boldsymbol{\beta}} = (1-\eta)\boldsymbol{\beta}^* + \eta\widehat{\boldsymbol{\beta}}$, where $\eta = \sup\{u\in[0,1] : \boldsymbol{\beta}^* + u(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)\in\mathbb{B}_{\boldsymbol{\Sigma}}(r)\}$, and note that $\widetilde{\boldsymbol{\beta}}\in\boldsymbol{\beta}^*+\mathbb{B}_{\boldsymbol{\Sigma}}(r)$. If $\widehat{\boldsymbol{\beta}}\in\boldsymbol{\beta}^*+\mathbb{B}_{\boldsymbol{\Sigma}}(r)$, $\widetilde{\boldsymbol{\beta}}$ coincides with $\widehat{\boldsymbol{\beta}}$; otherwise, $\widetilde{\boldsymbol{\beta}}$ lies on the boundary of $\boldsymbol{\beta}^*+\mathbb{B}_{\boldsymbol{\Sigma}}(r)$ with η strictly less than 1.

We first show that $\widetilde{\boldsymbol{\beta}}\in\boldsymbol{\beta}^*+\mathbb{C}_{\boldsymbol{\Sigma}}(l(a,k))$. By a variant of (B.2) and the optimality of $\widehat{\boldsymbol{\beta}}$, we have

$$0 \leq \langle\nabla\widehat{Q}_h(\widetilde{\boldsymbol{\beta}}) - \nabla\widehat{Q}_h(\boldsymbol{\beta}^*), \widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\rangle \leq \eta\langle\nabla\widehat{Q}_h(\widehat{\boldsymbol{\beta}}) - \nabla\widehat{Q}_h(\boldsymbol{\beta}^*), \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\rangle \tag{G.8}$$

and $\nabla\widehat{Q}_h(\widehat{\boldsymbol{\beta}}) + \boldsymbol{\lambda}\circ\widehat{\boldsymbol{g}} = \mathbf{0}$ for some $\widehat{\boldsymbol{g}}\in\partial\|\widehat{\boldsymbol{\beta}}\|_1$. Lemma D.1 ensures that, conditioned on $\{\lambda \geq (2/a)\|\boldsymbol{w}_h^*\|_\infty\}$, $\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*$ obeys the following cone-type constraint:

$$\begin{aligned}
\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)_{\mathcal{T}^c}\|_1 &\leq \frac{(\|\boldsymbol{\lambda}\|_\infty + \|\boldsymbol{w}_h^*\|_\infty)\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)_{\mathcal{T}}\|_1 + b_h^*\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}}}{\|\boldsymbol{\lambda}_{\mathcal{T}^c}\|_{\min} - \|\boldsymbol{w}_h^*\|_\infty} \\
&\leq \left(1 + \frac{2\lambda}{\|\boldsymbol{\lambda}_{\mathcal{T}^c}\|_{\min}}\right)\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)_{\mathcal{T}}\|_1 + \frac{2}{\|\boldsymbol{\lambda}_{\mathcal{T}^c}\|_{\min}}b_h^*\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} \\
&\leq (1+2/a)\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)_{\mathcal{T}}\|_1 + 2b_h^*\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}}/(a\lambda),
\end{aligned}$$

where the second and third inequalities follow from the assumed condition on $\boldsymbol{\lambda}$, i.e., $\lambda \geq \|\boldsymbol{\lambda}\|_\infty$ and $\|\boldsymbol{\lambda}_{\mathcal{T}^c}\|_{\min} \geq a\lambda$. Since $\lambda \geq (s/\gamma_p)^{-1/2} b_h^*$, it follows that

$$\begin{aligned}
\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 &\leq (2 + 2/a)\|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{\mathcal{T}}\|_1 + 2b_h^*\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}}/(a\lambda) \\
&\leq (2 + 2/a)k^{1/2}\|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{\mathcal{T}}\|_2 + 2b_h^*\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}}/(a\lambda) \\
&\leq \{(2 + 2/a)(k/\gamma_p)^{1/2} + (2/a)(s/\gamma_p)^{1/2}\}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} \\
&=: l(a,k)\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}},
\end{aligned}$$

where $\gamma_p = \gamma_p(\boldsymbol{\Sigma})$ is the minimum eigenvalue of $\boldsymbol{\Sigma}$. Thus $\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \in \mathbb{B}_{\boldsymbol{\Sigma}}(r) \cap \mathbb{C}_{\boldsymbol{\Sigma}}(l(a,k))$. Conditioned on the event $\mathcal{E}_{\rm rsc}(r, l(a,k), \kappa)$,

$$\langle \nabla \widehat{Q}_h(\widetilde{\boldsymbol{\beta}}) - \nabla \widehat{Q}_h(\boldsymbol{\beta}^*), \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \rangle \geq \kappa \cdot \|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}}^2. \tag{G.9}$$

Turning to the right-hand side of (G.8), we have

$$\langle \nabla \widehat{Q}_h(\widehat{\boldsymbol{\beta}}) - \nabla \widehat{Q}_h(\boldsymbol{\beta}^*), \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \rangle = -\langle \nabla \widehat{Q}_h(\boldsymbol{\beta}^*), \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \rangle - \langle \boldsymbol{\lambda} \circ \widehat{\boldsymbol{g}}, \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \rangle =: \Pi_1 + \Pi_2. \tag{G.10}$$

It suffices to obtain upper bounds for Π_1 and Π_2. Recall that $\boldsymbol{w}_h^* = \nabla \widehat{Q}_h(\boldsymbol{\beta}^*) - \nabla Q_h(\boldsymbol{\beta}^*)$ and $\widehat{\boldsymbol{\delta}} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$. Then, $|\Pi_1|$ can be upper bounded by

$$|\Pi_1| = \langle \boldsymbol{w}_h^*, \widehat{\boldsymbol{\delta}} \rangle + \langle \nabla Q_h(\boldsymbol{\beta}^*), \widehat{\boldsymbol{\delta}} \rangle \leq \|\boldsymbol{w}_{h,\mathcal{T}}^*\|_2 \|\widehat{\boldsymbol{\delta}}_{\mathcal{T}}\|_2 + \|\boldsymbol{w}_{h,\mathcal{T}^c}^*\|_\infty \|\widehat{\boldsymbol{\delta}}_{\mathcal{T}^c}\|_1 + b_h^* \|\widehat{\boldsymbol{\delta}}\|_{\boldsymbol{\Sigma}}. \tag{G.11}$$

For Π_2, consider the decomposition

$$\langle \boldsymbol{\lambda} \circ \widehat{\boldsymbol{g}}, \widehat{\boldsymbol{\delta}} \rangle = \langle (\boldsymbol{\lambda} \circ \widehat{\boldsymbol{g}})_{\mathcal{S}}, \widehat{\boldsymbol{\delta}}_{\mathcal{S}} \rangle + \langle (\boldsymbol{\lambda} \circ \widehat{\boldsymbol{g}})_{\mathcal{T} \setminus \mathcal{S}}, \widehat{\boldsymbol{\delta}}_{\mathcal{T} \setminus \mathcal{S}} \rangle + \langle (\boldsymbol{\lambda} \circ \widehat{\boldsymbol{g}})_{\mathcal{T}^c}, \widehat{\boldsymbol{\delta}}_{\mathcal{T}^c} \rangle$$

and note that

$$\langle (\boldsymbol{\lambda} \circ \widehat{\boldsymbol{g}})_{\mathcal{S}}, \widehat{\boldsymbol{\delta}}_{\mathcal{S}} \rangle \geq -\|\boldsymbol{\lambda}_{\mathcal{S}}\|_2 \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_2.$$

Since $\boldsymbol{\beta}_{\mathcal{S}^c}^* = \mathbf{0}$ and $\widehat{\boldsymbol{g}} \in \partial\|\widehat{\boldsymbol{\beta}}\|_1$, $\langle (\boldsymbol{\lambda} \circ \widehat{\boldsymbol{g}})_{\mathcal{T} \setminus \mathcal{S}}, (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{\mathcal{T} \setminus \mathcal{S}} \rangle = \langle (\boldsymbol{\lambda} \circ \widehat{\boldsymbol{g}})_{\mathcal{T} \setminus \mathcal{S}}, \widehat{\boldsymbol{\beta}}_{\mathcal{T} \setminus \mathcal{S}} \rangle \geq 0$ and

$$\langle (\boldsymbol{\lambda} \circ \widehat{\boldsymbol{g}})_{\mathcal{T}^c}, (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{\mathcal{T}^c} \rangle = \langle (\boldsymbol{\lambda} \circ \widehat{\boldsymbol{g}})_{\mathcal{T}^c}, \widehat{\boldsymbol{\beta}}_{\mathcal{T}^c} \rangle = \|(\boldsymbol{\lambda} \circ \widehat{\boldsymbol{\beta}})_{\mathcal{T}^c}\|_1 \geq \|\boldsymbol{\lambda}_{\mathcal{T}^c}\|_{\min} \|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{\mathcal{T}^c}\|_1.$$

Combining the above equations, we conclude that

$$\Pi_2 \leq \|\boldsymbol{\lambda}_{\mathcal{S}}\|_2 \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_2 - \|\boldsymbol{\lambda}_{\mathcal{T}^c}\|_{\min} \|\widehat{\boldsymbol{\delta}}_{\mathcal{T}^c}\|_1. \tag{G.12}$$

Substituting (G.12) and (G.11) into (G.10) implies

$$\begin{aligned}
&\langle \nabla \widehat{Q}_h(\widehat{\boldsymbol{\beta}}) - \nabla \widehat{Q}_h(\boldsymbol{\beta}^*), \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \rangle \\
&\leq \|\boldsymbol{w}_{h,\mathcal{T}}^*\|_2 \|\widehat{\boldsymbol{\delta}}_{\mathcal{T}}\|_2 + \|\boldsymbol{\lambda}_{\mathcal{S}}\|_2 \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_2 + b_h^* \|\widehat{\boldsymbol{\delta}}\|_{\boldsymbol{\Sigma}} - (\|\boldsymbol{\lambda}_{\mathcal{T}^c}\|_{\min} - \|\boldsymbol{w}_{h,\mathcal{T}^c}^*\|_\infty) \|\widehat{\boldsymbol{\delta}}_{\mathcal{T}^c}\|_1.
\end{aligned} \tag{G.13}$$

Recall that $\eta \widehat{\boldsymbol{\delta}} = \widetilde{\boldsymbol{\delta}}$. Provided $\|\boldsymbol{\lambda}\|_\infty \leq \lambda$ and $\|\boldsymbol{\lambda}_{\mathcal{T}^c}\|_{\min} \geq a\lambda \geq 2\|\boldsymbol{w}_h^*\|_\infty$, substituting (G.9) and (G.13) into (G.8) yields

$$\begin{aligned}
\kappa \|\widetilde{\boldsymbol{\delta}}\|_{\boldsymbol{\Sigma}}^2 &\leq \|\boldsymbol{w}_{h,\mathcal{T}}^*\|_2 \|\widetilde{\boldsymbol{\delta}}_{\mathcal{T}}\|_2 + \|\boldsymbol{\lambda}_{\mathcal{S}}\|_2 \|\widetilde{\boldsymbol{\delta}}_{\mathcal{S}}\|_2 + b_h^* \|\widetilde{\boldsymbol{\delta}}\|_{\boldsymbol{\Sigma}} \\
&\leq \gamma_p^{-1/2}(k^{1/2}\|\boldsymbol{w}_h^*\|_\infty + s^{1/2}\|\boldsymbol{\lambda}\|_\infty)\|\widetilde{\boldsymbol{\delta}}\|_{\boldsymbol{\Sigma}} + b_h^* \|\widetilde{\boldsymbol{\delta}}\|_{\boldsymbol{\Sigma}} \\
&\leq \{\gamma_p^{-1/2}(k^{1/2}a/2 + s^{1/2})\lambda + b_h^*\}\|\widetilde{\boldsymbol{\delta}}\|_{\boldsymbol{\Sigma}},
\end{aligned}$$

from which it follows that $\|\widetilde{\boldsymbol{\delta}}\|_{\boldsymbol{\Sigma}} \leq \kappa^{-1}\{\gamma_p^{-1/2}(k^{1/2}a/2 + s^{1/2})\lambda + b_h^*\}$. The constraint (D.1) ensures that $\widetilde{\boldsymbol{\beta}}$ falls in the interior of $\boldsymbol{\beta}^* + \mathbb{B}_{\boldsymbol{\Sigma}}(r)$. Therefore, we must have $\eta = 1$ and $\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}$. □

G.5 Proof of Lemma D.3

Recall that $\boldsymbol{w}_h^* = \nabla \widehat{Q}_h(\boldsymbol{\beta}^*) - \nabla Q_h(\boldsymbol{\beta}^*)$ is the centered score function evaluated at $\boldsymbol{\beta}^*$ and that $\xi_i = \bar{K}(-\varepsilon_i/h) - \tau$. Thus, $\boldsymbol{w}_{h,\mathcal{S}}^* = (1/n)\sum_{i=1}^n \{\xi_i \boldsymbol{x}_{i,\mathcal{S}} - \mathbb{E}(\xi_i \boldsymbol{x}_{i,\mathcal{S}})\} \in \mathbb{R}^s$. We first obtain an upper bound for $\|\mathbf{S}^{-1/2}\boldsymbol{w}_{h,\mathcal{S}}^*\|_2 = \sup_{\|\boldsymbol{u}\|_2=1}\langle \boldsymbol{u}, \mathbf{S}^{-1/2}\boldsymbol{w}_{h,\mathcal{S}}^*\rangle$. Using a covering argument, for any $\epsilon \in (0,1)$, there exists an ϵ-net $\mathcal{N}_\epsilon$ of the unit sphere with cardinality $|\mathcal{N}_\epsilon| \le (1+2/\epsilon)^s$ such that $\|\mathbf{S}^{-1/2}\boldsymbol{w}_{h,\mathcal{S}}^*\|_2 \le (1-\epsilon)^{-1}\max_{\boldsymbol{u}\in\mathcal{N}_\epsilon}\langle \boldsymbol{u}, \mathbf{S}^{-1/2}\boldsymbol{w}_{h,\mathcal{S}}^*\rangle$. Thus, it suffices to obtain an upper bound for $\langle \boldsymbol{u}, \mathbf{S}^{-1/2}\boldsymbol{w}_{h,\mathcal{S}}^*\rangle$ for each direction $\boldsymbol{u} \in \mathcal{N}_\epsilon$.

For each direction $\boldsymbol{u} \in \mathcal{N}_\epsilon$, let $\gamma_{\boldsymbol{u},i} = \langle \boldsymbol{u}, \mathbf{S}^{-1/2}\{\xi_i \boldsymbol{x}_{i,\mathcal{S}} - \mathbb{E}(\xi_i \boldsymbol{x}_{i,\mathcal{S}})\}\rangle$. We employ the Bernstein's inequality to bound $(1/n)\sum_{i=1}^n \gamma_{\boldsymbol{u},i}$. Note that $\gamma_{\boldsymbol{u},i}$ has mean zero, and by (G.6), the variance can be upper bounded as

$$\mathrm{var}(\gamma_{\boldsymbol{u},i}) \le \{\tau(1-\tau) + Ch^2\} \cdot \mathbb{E}\langle \boldsymbol{u}, \mathbf{S}^{-1/2}\boldsymbol{x}_{i,\mathcal{S}}\rangle^2.$$

Under Condition (B3), we have $\mathbb{P}(|\langle \boldsymbol{u}, \mathbf{S}^{-1/2}\boldsymbol{x}_{i,\mathcal{S}}\rangle| \ge \upsilon_0 t) \le e^{-t}$ for all $t \ge 0$. Noting that $|\xi_i| \le \max(1-\tau,\tau)$, we have for $k = 2, 3, \ldots$,

$$\begin{aligned}
\mathbb{E}(|\langle \boldsymbol{u}, \mathbf{S}^{-1/2}\boldsymbol{x}_{i,\mathcal{S}}\xi_i\rangle|^k) &\le \max(1-\tau,\tau)^{k-2}\,\mathbb{E}_{\boldsymbol{x}}\{|\langle \boldsymbol{u}, \mathbf{S}^{-1/2}\boldsymbol{x}_{i,\mathcal{S}}\rangle|^k \cdot \mathbb{E}(\xi_i^2|\boldsymbol{x}_i)\} \\
&\le \max(1-\tau,\tau)^{k-2}\{\tau(1-\tau)+Ch^2\}\upsilon_0^k \int_0^\infty \mathbb{P}(|\langle \boldsymbol{u}, \mathbf{S}^{-1/2}\boldsymbol{x}_{i,\mathcal{S}}\rangle| \ge \upsilon_0 t) k t^{k-1}\,\mathrm{d}t \\
&\le \max(1-\tau,\tau)^{k-2}\{\tau(1-\tau)+Ch^2\}\upsilon_0^k k \int_0^\infty t^{k-1}e^{-t}\,\mathrm{d}t \\
&= k!\max(1-\tau,\tau)^{k-2}\{\tau(1-\tau)+Ch^2\}\upsilon_0^k \\
&\le \frac{k!}{2}\cdot\{\tau(1-\tau)+Ch^2\}\upsilon_0^2 \cdot \{2\max(1-\tau,\tau)\upsilon_0\}^{k-2}. \qquad \text{(G.14)}
\end{aligned}$$

Consequently, it follows from Bernstein's inequality that for every $t \ge 0$,

$$\frac{1}{n}\sum_{i=1}^n \gamma_{\boldsymbol{u},i} \le \upsilon_0\left[\sqrt{\{\tau(1-\tau)+Ch^2\}\frac{2t}{n}} + \max(1-\tau,\tau)\frac{2t}{n}\right]$$

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

Finally, we apply a union bound over all vectors $\boldsymbol{u} \in \mathcal{N}_\epsilon$ and obtain

$$\|\mathbf{S}^{-1/2}\boldsymbol{w}_{h,\mathcal{S}}^*\|_2 \le \frac{\upsilon_0}{1-\epsilon}\left[\sqrt{\{\tau(1-\tau)+Ch^2\}\frac{2t}{n}} + \max(1-\tau,\tau)\frac{2t}{n}\right] \qquad \text{(G.15)}$$

with probability at least $1-e^{\log(1+2/\epsilon)s-t}$. Selecting $\epsilon = 0.313$ and taking $t = 2s+y$ yield the claimed bound. □

G.6 Proof of Lemma E.1

The proof is similar to that of Lemma D.2. We therefore only provide the key steps. As before, construct an intermediate vector $\widetilde{\boldsymbol{\beta}} = (1-\eta)\widehat{\boldsymbol{\beta}}^{\mathrm{ora}} + \eta\widehat{\boldsymbol{\beta}}$ satisfying $\widetilde{\boldsymbol{\beta}} \in \widehat{\boldsymbol{\beta}}^{\mathrm{ora}} + \mathbb{B}_{\boldsymbol{\Sigma}}(r)$, where $\eta \in (0,1]$ is chosen such that (i) $\eta = 1$ if $\widehat{\boldsymbol{\beta}} \in \widehat{\boldsymbol{\beta}}^{\mathrm{ora}} + \mathbb{B}_{\boldsymbol{\Sigma}}(r)$, and (ii) $\eta \in (0,1)$ if $\widehat{\boldsymbol{\beta}} \notin \widehat{\boldsymbol{\beta}}^{\mathrm{ora}} + \mathbb{B}_{\boldsymbol{\Sigma}}(r)$. In the latter case, $\widetilde{\boldsymbol{\beta}}$ lies on the boundary of $\boldsymbol{\beta}^* + \mathbb{B}_{\boldsymbol{\Sigma}}(r)$.

We first show that $\widetilde{\boldsymbol{\beta}} \in \widehat{\boldsymbol{\beta}}^{\mathrm{ora}} + \mathbb{C}(l)$ conditioned on the event $\{\|\boldsymbol{w}_h^{\mathrm{ora}}\|_\infty \le 0.5a\lambda\}$, where $l = (2+2/a)(k/\gamma_p)^{1/2}$. By (B.1) and the optimality of $\widehat{\boldsymbol{\beta}}$, we have

$$\langle \nabla\widehat{Q}_h(\widetilde{\boldsymbol{\beta}}) - \nabla\widehat{Q}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}), \widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\mathrm{ora}}\rangle \le \eta\langle \nabla\widehat{Q}_h(\widehat{\boldsymbol{\beta}}) - \nabla\widehat{Q}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}), \widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\mathrm{ora}}\rangle \qquad \text{(G.16)}$$

and $\nabla \widehat{Q}_h(\widehat{\boldsymbol{\beta}}) + \boldsymbol{\lambda} \circ \widehat{\boldsymbol{g}} = \mathbf{0}$ for some $\widehat{\boldsymbol{g}} \in \partial \|\widehat{\boldsymbol{\beta}}\|_1$. Following the proof of Lemma D.1 with $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}^{\rm ora}$, it can be similarly shown that conditioned on $\{\|\boldsymbol{w}_h^{\rm ora}\|_\infty \leq 0.5a\lambda\}$,

$$\|(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\rm ora})_{\mathcal{T}^{\rm c}}\|_1 \leq \frac{\{\|\boldsymbol{\lambda}\|_\infty + \|\boldsymbol{w}_h^{\rm ora}\|_\infty\}\|(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\rm ora})_{\mathcal{T}}\|_1}{\|\boldsymbol{\lambda}_{\mathcal{T}^{\rm c}}\|_{\min} - \|\boldsymbol{w}_h^{\rm ora}\|_\infty}$$
$$\leq \left(1 + \frac{2\lambda}{\|\boldsymbol{\lambda}_{\mathcal{T}^{\rm c}}\|_{\min}}\right)\|(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\rm ora})_{\mathcal{T}}\|_1 \leq (1 + 2/a)\|(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\rm ora})_{\mathcal{T}}\|_1.$$

Here there is no bias term because $\boldsymbol{w}_h^{\rm ora}$ is the score function evaluated at $\widehat{\boldsymbol{\beta}}^{\rm ora}$ without subtracting the mean. Consequently, $\|\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\rm ora}\|_1 \leq (2 + 2/a)\|(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\rm ora})_{\mathcal{T}}\|_1 \leq (2 + 2/a)(k/\gamma_p)^{1/2}\|\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\rm ora}\|_{\boldsymbol{\Sigma}}$, implying $\widetilde{\boldsymbol{\beta}} \in \widehat{\boldsymbol{\beta}}^{\rm ora} + \mathbb{C}_{\boldsymbol{\Sigma}}(l)$. Furthermore, if the event $\mathcal{G}_{\rm rsc}(r, l, \kappa)$ occurs,

$$\langle \nabla \widehat{Q}_h(\widetilde{\boldsymbol{\beta}}) - \nabla \widehat{Q}_h(\widehat{\boldsymbol{\beta}}^{\rm ora}), \widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\rm ora} \rangle \geq \kappa \|\widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\rm ora}\|_{\boldsymbol{\Sigma}}^2. \tag{G.17}$$

Let $\widehat{\boldsymbol{\delta}} = \widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\rm ora}$ and $\widetilde{\boldsymbol{\delta}} = \widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\rm ora} = \eta \widehat{\boldsymbol{\delta}}$. For the right-hand side of (G.16), by a similar argument to that leads to (G.13), we obtain

$$\begin{aligned}&\langle \nabla \widehat{Q}_h(\widehat{\boldsymbol{\beta}}) - \nabla \widehat{Q}_h(\widehat{\boldsymbol{\beta}}^{\rm ora}), \widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^{\rm ora} \rangle \\ &\leq \|\boldsymbol{w}_{h,\mathcal{T}}^{\rm ora}\|_2 \|\widehat{\boldsymbol{\delta}}_{\mathcal{T}}\|_2 + \|\boldsymbol{\lambda}_{\mathcal{S}}\|_2 \|\widehat{\boldsymbol{\delta}}_{\mathcal{S}}\|_2 - (\|\boldsymbol{\lambda}_{\mathcal{T}^{\rm c}}\|_{\min} - \|\boldsymbol{w}_{h,\mathcal{T}^{\rm c}}^{\rm ora}\|_\infty)\|\widehat{\boldsymbol{\delta}}_{\mathcal{T}^{\rm c}}\|_1.\end{aligned} \tag{G.18}$$

Given the stated conditioning, it follows from (G.16), (G.17) and (G.18) that

$$\kappa \|\widetilde{\boldsymbol{\delta}}\|_{\boldsymbol{\Sigma}}^2 \leq \|\boldsymbol{w}_{h,\mathcal{T}}^{\rm ora}\|_2 \|\widetilde{\boldsymbol{\delta}}_{\mathcal{T}}\|_2 + \|\boldsymbol{\lambda}_{\mathcal{S}}\|_2 \|\widetilde{\boldsymbol{\delta}}_{\mathcal{S}}\|_2 \leq \gamma_p^{-1/2}(\|\boldsymbol{w}_{h,\mathcal{T}}^{\rm ora}\|_2 + \|\boldsymbol{\lambda}_{\mathcal{S}}\|_2)\|\widetilde{\boldsymbol{\delta}}\|_{\boldsymbol{\Sigma}}.$$

Canceling out a factor of $\|\widetilde{\boldsymbol{\delta}}\|_{\boldsymbol{\Sigma}}$ from both sides yields

$$\|\widetilde{\boldsymbol{\delta}}\|_{\boldsymbol{\Sigma}} \leq \kappa^{-1}\gamma_p^{-1/2}(0.5ak^{1/2} + s^{1/2})\lambda < r,$$

where the second inequality follows from (E.1). Consequently, $\widetilde{\boldsymbol{\beta}}$ falls in the interior of $\widehat{\boldsymbol{\beta}}^{\rm ora} + \mathbb{B}_{\boldsymbol{\Sigma}}(r)$, thus enforcing $\eta = 1$ and $\widehat{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}$. This proves the claimed bound (E.2). □

G.7 Proof of Lemma E.2

The proof is based on an argument similar to that in the proof of Lemma 4.2 and also Proposition 2 in Loh (2017). Since the bandwidth h plays a critical role in subsequent analysis, we provide details of the proof that highlight its connection with the sample size.

For each pair $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, write $\boldsymbol{\delta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$, and similarly to (C.3) and (C.4),

$$D(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) := \langle \nabla \widehat{Q}_h(\boldsymbol{\beta}_1) - \nabla \widehat{Q}_h(\boldsymbol{\beta}_2), \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 \rangle \geq \frac{\kappa_l}{nh} \sum_{i=1}^n (\boldsymbol{x}_i^{\rm T}\boldsymbol{\delta})^2 \mathbb{1}_{E_i}, \tag{G.19}$$

where with slight abuse of notation, $\mathbb{1}_{E_i}$ is the indicator function of the event

$$E_i = \{|\varepsilon_i| \leq h/2\} \cap \{|\langle \boldsymbol{x}_i, \boldsymbol{\beta}_2 - \boldsymbol{\beta}^* \rangle| \leq h/4\} \cap \{|\boldsymbol{x}_i^{\rm T}\boldsymbol{\delta}| \leq \|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}} \cdot h/(4r)\},$$

on which $\max\{|y_i - \boldsymbol{x}_i^{\rm T}\boldsymbol{\beta}_1|, |y_i - \boldsymbol{x}_i^{\rm T}\boldsymbol{\beta}_2|\} \leq h$ for all $\boldsymbol{\beta}_1 \in \boldsymbol{\beta}_2 + \mathbb{B}_{\boldsymbol{\Sigma}}(r)$. In addition to the function φ_R introduced in the proof of Lemma 4.2, we further define

$$\phi_R(u) = \mathbb{1}(|u| \leq R/2) + \{2 - (2/R)\,\mathrm{sign}(u)\}\mathbb{1}(R/2 \leq |u| \leq R),$$

which is a smoothed version of the indicator function $u \mapsto \mathbb{1}(|u| \le R)$ and satisfies $\mathbb{1}(|u| \le R/2) \le \phi_R(u) \le \mathbb{1}(|u| \le R)$. Consequently,

$$
\begin{aligned}
D(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) &\ge \kappa_l \cdot \|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}^2 \cdot \underbrace{\frac{1}{nh} \sum_{i=1}^n \chi_i \cdot \varphi_{h/(4r)}(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}) \phi_{h/4}(\langle \boldsymbol{x}_i, \boldsymbol{\beta}_2 - \boldsymbol{\beta}^* \rangle)}_{=: D_0(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)} \\
&= \kappa_l \cdot \|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}^2 \cdot \{\mathbb{E} D_0(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) + D_0(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) - \mathbb{E} D_0(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\},
\end{aligned} \tag{G.20}
$$

where $\chi_i = \mathbb{1}(|\varepsilon_i| \le h/2)$. Provided $h \le f_l/(2l_0)$, the earlier result (C.7) implies

$$
7 f_l h/8 \le \mathbb{E}(\chi_i | \boldsymbol{x}_i) \le 9 f_u h/8 \ \text{ almost surely.}
$$

To bound the mean $\mathbb{E} D_0(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ from below, applying (C.7) and inequalities $\varphi_R(u) \ge u^2 \mathbb{1}(|u| \le R/2)$ and $\phi_R(u) \ge \mathbb{1}(|u| \le R/2)$ yields

$$
\begin{aligned}
&\mathbb{E}\{\chi_i \cdot \varphi_{h/(4r)}(\boldsymbol{x}^{\mathrm{T}} \boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}) \phi_{h/4}(\langle \boldsymbol{x}, \boldsymbol{\beta}_2 - \boldsymbol{\beta}^* \rangle)\} \\
&\ge \frac{7}{8} f_l h \, \mathbb{E}\{\varphi_{h/(4r)}(\boldsymbol{x}^{\mathrm{T}} \boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}) \phi_{h/4}(\langle \boldsymbol{x}, \boldsymbol{\beta}_2 - \boldsymbol{\beta}^* \rangle)\} \\
&\ge \frac{7}{8} f_l h \Big(1 - \mathbb{E}\{\xi_{\boldsymbol{\delta}}^2 \mathbb{1}_{|\xi_{\boldsymbol{\delta}}| > h/(8r)}\} - \mathbb{E}\{\xi_{\boldsymbol{\delta}}^2 \mathbb{1}_{|\langle \boldsymbol{x}, \boldsymbol{\beta}_2 - \boldsymbol{\beta}^* \rangle| > h/8}\}\Big),
\end{aligned}
$$

where $\xi_{\boldsymbol{\delta}} = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}$ is such that $\mathbb{E}\xi_{\boldsymbol{\delta}}^2 = 1$. Under Condition (B3′) with $\upsilon_1 \ge 1$, for any $\boldsymbol{\delta} \in \mathbb{R}^p$ and $u > 0$ we have

$$
\mathbb{E}\{\xi_{\boldsymbol{\delta}}^2 \mathbb{1}(|\xi_{\boldsymbol{\delta}}| > u)\} \le 2u^2 e^{-u^2/2\upsilon_1^2} + 4\upsilon_1^2 \int_{u/\upsilon_1}^{\infty} t e^{-t^2/2} \, \mathrm{d}t = (2u^2 + 4\upsilon_1^2) e^{-u^2/2\upsilon_1^2}.
$$

Moreover, for $\boldsymbol{\beta}_2 \in \boldsymbol{\beta}^* + \mathbb{B}_{\boldsymbol{\Sigma}}(r/2)$,

$$
\mathbb{E}\{\xi_{\boldsymbol{\delta}}^2 \mathbb{1}_{|\langle \boldsymbol{x}, \boldsymbol{\beta}_2 - \boldsymbol{\beta}^* \rangle| > h/8}\} \le (\mathbb{E}\xi_{\boldsymbol{\delta}}^4)^{1/2} \mathbb{P}(|\langle \boldsymbol{x}, \boldsymbol{\beta}_2 - \boldsymbol{\beta}^* \rangle| > h/8)^{1/2} \le 4\sqrt{2} \upsilon_1^2 e^{-h^2/(8\upsilon_1 r)^2},
$$

where we have used the fact that $\mathbb{E}\xi_{\boldsymbol{\delta}}^4 \le 16\upsilon_1^4$. From the above three moment inequalities, we find that as long as $24\upsilon_1^2 r \le h$, or equivalently, $h/(8r) \ge 3\upsilon_1^2$,

$$
\mathbb{E} D_0(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) > 0.66 f_l \ \text{ for all } \ \boldsymbol{\beta}_1 \in \boldsymbol{\beta}_2 + \mathbb{B}_{\boldsymbol{\Sigma}}(r) \ \text{ and } \ \boldsymbol{\beta}_2 \in \boldsymbol{\beta}^* + \mathbb{B}_{\boldsymbol{\Sigma}}(r/2). \tag{G.21}
$$

To bound $|D_0(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) - \mathbb{E} D_0(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)|$ uniformly over $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \in \Lambda(r, l) = \{(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) : \boldsymbol{\beta} \in \boldsymbol{\beta}_1 + \mathbb{B}_{\boldsymbol{\Sigma}}(r) \cap \mathbb{C}_{\boldsymbol{\Sigma}}(l), \boldsymbol{\beta}_2 \in \boldsymbol{\beta}^* + \mathbb{B}_{\boldsymbol{\Sigma}}(r/2), \mathrm{supp}(\boldsymbol{\beta}_2) \subseteq \mathcal{S}\}$, define

$$
\Omega(r, l) = \sup_{(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \in \Lambda(r, l)} \{-D_0(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) + \mathbb{E} D_0(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\}.
$$

Write $D_0(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (1/n) \sum_{i=1}^n \omega_{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2}(\boldsymbol{x}_i, \varepsilon_i)$, where

$$
\omega_{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2}(\boldsymbol{x}_i, \varepsilon_i) = (\chi_i/h) \cdot \varphi_{h/(4r)}(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}) \phi_{h/4}(\langle \boldsymbol{x}_i, \boldsymbol{\beta}_2 - \boldsymbol{\beta}^* \rangle).
$$

Note that $\varphi_R(u) \le (R/2)|u|$ and $\phi_R(u) \in [0, 1]$. Then, for $h \le f_l/(2l_0)$,

$$
0 \le \omega_{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2}(\boldsymbol{x}_i, \varepsilon_i) \le (8r)^{-2} \cdot h \quad \text{and} \quad \mathbb{E}\omega_{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2}^2(\boldsymbol{x}_i, \varepsilon_i) \le (8r)^{-2} \cdot 9 f_u h/8.
$$

Again, using Bousquet's version of Talagrand's inequality yields that, for any $t > 0$,

$$\Omega(r,l) \le \mathbb{E}\Omega(r,l) + \sqrt{\frac{9f_u/8 + 2\mathbb{E}\Omega(r,l)}{(8r)^2}\frac{2ht}{n}} + \frac{h}{(8r)^2}\frac{t}{3n}$$
$$\le \frac{5}{4}\mathbb{E}\Omega(r,l) + \frac{3}{16}\sqrt{\frac{f_u ht}{r^2 n}} + \frac{13}{3}\frac{ht}{(8r)^2 n} \tag{G.22}$$

holds with probability at least $1 - e^{-t}$. To bound $\mathbb{E}\Omega(r,l)$, we proceed with a different method to that in the proof of Lemma 4.2. Using symmetrization with Rademacher random variables and by the connection between Gaussian and Rademacher complexities (see, e.g. Lemma 4.5 in Ledoux and Talagrand (1991)), we obtain

$$\mathbb{E}\Omega(r,l) \le 2 \cdot \sqrt{\frac{\pi}{2}} \cdot \mathbb{E}\Big\{ \sup_{(\beta_1,\beta_2)\in\Lambda(r,l)} \mathbb{G}_{\beta_1,\beta_2} \Big\}, \tag{G.23}$$

where $\mathbb{G}_{\beta_1,\beta_2} := (nh)^{-1}\sum_{i=1}^n g_i\chi_i \cdot \varphi_{h/(4r)}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}})\phi_{h/4}(\langle \boldsymbol{x}_i, \boldsymbol{\beta}_2 - \boldsymbol{\beta}^*\rangle)$ with $\boldsymbol{\delta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$, and g_i are independent standard normal random variables that are independent of the observations. Let $\mathbb{E}^*$ be the conditional expectation given $\{(y_i, \boldsymbol{x}_i)\}_{i=1}^n$. Note that $\{\mathbb{G}_{\beta_1,\beta_2}\}_{(\beta_1,\beta_2)\in\Lambda(r,l)}$ is a (conditional) Gaussian process and $\mathbb{G}_{\beta^*,\beta^*} = 0$. We then apply the Gaussian comparison theorem to bound $\mathbb{E}^*\{\sup_{(\beta_1,\beta_2)\in\Lambda(r,l)} \mathbb{G}_{\beta_1,\beta_2}\}$, from which an upper bound for $\mathbb{E}\{\sup_{(\beta_1,\beta_2)\in\Lambda(r,l)} \mathbb{G}_{\beta_1,\beta_2}\}$ follows immediately. For $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2), (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2') \in \Lambda(r,l)$, write $\boldsymbol{\delta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$ and $\boldsymbol{\delta}' = \boldsymbol{\beta}_1' - \boldsymbol{\beta}_2'$. Consequently,

$$\mathbb{G}_{\beta_1,\beta_2} - \mathbb{G}_{\beta_1',\beta_2'} = \mathbb{G}_{\beta_1,\beta_2} - \mathbb{G}_{\beta_2'+\delta,\beta_2'} + \mathbb{G}_{\beta_2'+\delta,\beta_2'} - \mathbb{G}_{\beta_1',\beta_2'}$$
$$= \frac{1}{nh}\sum_{i=1}^n g_i\chi_i \cdot \varphi_{h/(4r)}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}})\{\phi_{h/4}(\langle \boldsymbol{x}_i, \boldsymbol{\beta}_2 - \boldsymbol{\beta}^*\rangle) - \phi_{h/4}(\langle \boldsymbol{x}_i, \boldsymbol{\beta}_2' - \boldsymbol{\beta}^*\rangle)\}$$
$$+ \frac{1}{nh}\sum_{i=1}^n g_i\chi_i \cdot \phi_{h/4}(\langle \boldsymbol{x}_i, \boldsymbol{\beta}_2' - \boldsymbol{\beta}^*\rangle)\{\varphi_{h/(4r)}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}) - \varphi_{h/(4r)}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}'/\|\boldsymbol{\delta}'\|_{\boldsymbol{\Sigma}})\}.$$

Note that ϕ_R and φ_R are, respectively, $(2/R)$- and R-Lipschitz continuous, and $\varphi_R(u) \le (R/2)^2$. Consequently,

$$\mathbb{E}^*(\mathbb{G}_{\beta_1,\beta_2} - \mathbb{G}_{\beta_2'+\delta,\beta_2'})^2$$
$$\le \frac{1}{n^2}\sum_{i=1}^n \frac{h^2}{(8r)^4}\Big(\frac{8}{h}\Big)^2 \langle \boldsymbol{x}_i, \boldsymbol{\beta}_2 - \boldsymbol{\beta}_2'\rangle^2\chi_i = \Big(\frac{1}{8r^2 n}\Big)^2 \sum_{i=1}^n \langle \boldsymbol{x}_i, \boldsymbol{\beta}_2 - \boldsymbol{\beta}_2'\rangle^2\chi_i \tag{G.24}$$

and

$$\mathbb{E}^*(\mathbb{G}_{\beta_2'+\delta,\beta_2'} - \mathbb{G}_{\beta_1',\beta_2'})^2 \le \frac{1}{(nh)^2}\sum_{i=1}^n \{\varphi_{h/(4r)}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}}) - \varphi_{h/(4r)}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}'/\|\boldsymbol{\delta}'\|_{\boldsymbol{\Sigma}})\}^2\chi_i$$
$$\le \Big(\frac{1}{4rn}\Big)^2 \sum_{i=1}^n (\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\boldsymbol{\Sigma}} - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\delta}'/\|\boldsymbol{\delta}'\|_{\boldsymbol{\Sigma}})^2\chi_i. \tag{G.25}$$

Motivated by (G.24), (G.25) and the inequality

$$\mathbb{E}^*(\mathbb{G}_{\beta_1,\beta_2} - \mathbb{G}_{\beta_1',\beta_2'})^2 \le 2\mathbb{E}^*(\mathbb{G}_{\beta_1,\beta_2} - \mathbb{G}_{\beta_2'+\delta,\beta_2'})^2 + 2\mathbb{E}^*(\mathbb{G}_{\beta_2'+\delta,\beta_2'} - \mathbb{G}_{\beta_1',\beta_2'})^2,$$

we define another Gaussian process $\{\mathbb{Z}_{\beta_1,\beta_2}\}_{(\beta_1,\beta_2)\in\Lambda(r,l)}$ as

$$\begin{aligned}\mathbb{Z}_{\beta_1,\beta_2} &= \frac{\sqrt{2}}{8r^2n}\sum_{i=1}^n g_i'\langle \boldsymbol{x}_i, \boldsymbol{\beta}_2-\boldsymbol{\beta}^*\rangle\chi_i + \frac{\sqrt{2}}{4rn}\sum_{i=1}^n g_i''\frac{\langle \boldsymbol{x}_i, \boldsymbol{\beta}_1-\boldsymbol{\beta}_2\rangle}{\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\|_{\boldsymbol{\Sigma}}}\chi_i \\ &= \frac{\sqrt{2}}{8r^2n}\sum_{i=1}^n \langle g_i'\boldsymbol{x}_{i,\mathcal{S}}, (\boldsymbol{\beta}_2-\boldsymbol{\beta}^*)_{\mathcal{S}}\rangle\chi_i + \frac{\sqrt{2}}{4rn}\sum_{i=1}^n g_i''\frac{\langle \boldsymbol{x}_i, \boldsymbol{\beta}_1-\boldsymbol{\beta}_2\rangle}{\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\|_{\boldsymbol{\Sigma}}}\chi_i,\end{aligned}$$

such that $\mathbb{E}^*(\mathbb{G}_{\beta_1,\beta_2}-\mathbb{G}_{\beta_1',\beta_2'})^2 \le \mathbb{E}^*(\mathbb{Z}_{\beta_1,\beta_2}-\mathbb{Z}_{\beta_1',\beta_2'})^2$, where $g_1', g_1'', \ldots, g_n', g_n''$ are i.i.d. standard normal random variables that are independent of all the other variables. Applying Theorem 7.2.11 in Vershynin (2018)—Sudakov-Fernique's Gaussian comparison inequality, we obtain

$$\mathbb{E}^*\bigg\{\sup_{(\beta_1,\beta_2)\in\Lambda(r,l)}\mathbb{G}_{\beta_1,\beta_2}\bigg\} \le \mathbb{E}^*\bigg\{\sup_{(\beta_1,\beta_2)\in\Lambda(r,l)}\mathbb{Z}_{\beta_1,\beta_2}\bigg\}, \tag{G.26}$$

which remains valid if $\mathbb{E}^*$ is replaced by $\mathbb{E}$. To bound the supremum of $\mathbb{Z}_{\beta_1,\beta_2}$ over $(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)\in\Lambda(r,l)$, using the cone-like constraint $\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\|_1 \le l\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\|_{\boldsymbol{\Sigma}}$ and $\boldsymbol{\beta}_2\in\boldsymbol{\beta}^*+\mathbb{B}_{\boldsymbol{\Sigma}}(r/2)$, we deduce that

$$\begin{aligned}\mathbb{E}\bigg\{\sup_{(\beta_1,\beta_2)\in\Lambda(r,l)}\mathbb{Z}_{\beta_1,\beta_2}\bigg\} &\le \frac{\sqrt{2}}{16r}\mathbb{E}\bigg\|\frac{1}{n}\sum_{i=1}^n g_i'\chi_i\mathbf{S}^{-1/2}\boldsymbol{x}_{i,\mathcal{S}}\bigg\|_2 + \frac{\sqrt{2}l}{4r}\mathbb{E}\bigg\|\frac{1}{n}\sum_{i=1}^n g_i''\chi_i\boldsymbol{x}_i\bigg\|_\infty \\ &\le \frac{\sqrt{2}}{16r}\sqrt{\frac{9f_uh}{8}\frac{s}{n}} + \frac{\sqrt{2}l}{4r}\mathbb{E}\bigg\|\frac{1}{n}\sum_{i=1}^n g_i''\chi_i\boldsymbol{x}_i\bigg\|_\infty.\end{aligned} \tag{G.27}$$

This, combined with (G.23), (G.26) and (G.27), yields

$$\mathbb{E}\Omega(r,l) \le \sqrt{\pi}\left\{\frac{3}{16}\sqrt{\frac{hs}{2r^2n}} + \frac{l}{2r}\mathbb{E}\bigg(\max_{1\le j\le p}\bigg|\frac{1}{n}\sum_{i=1}^n g_i\chi_i x_{ij}\bigg|\bigg)\right\}. \tag{G.28}$$

It remains the bound the second term on the right-hand side of (G.28). Write $S_j = \sum_{i=1}^n g_i\chi_i x_{ij}$ for $j = 1,\ldots,p$. Under Condition (B3′), for each $1\le j\le p$ and $k\ge 3$ we have

$$\mathbb{E}|x_j|^k \le 2\upsilon_1^k\sigma_{jj}^{k/2}k\int_0^\infty t^{k-1}e^{-t^2/2}\,\mathrm{d}t = 2^{k/2}\upsilon_1^k\sigma_{jj}^{k/2}k\Gamma(k/2).$$

Let $g\sim\mathcal{N}(0,1)$ be independent of $\boldsymbol{x}$. By the Legendre duplication formula $\Gamma(k)\Gamma(k+1/2) = 2^{1-2k}\sqrt{\pi}\,\Gamma(2k)$, we have

$$\mathbb{E}|gx_j|^k \le 2^{k/2}\frac{\Gamma(\frac{k+1}{2})}{\sqrt{\pi}}\cdot 2^{k/2}\upsilon_1^k\sigma_{jj}^{k/2}k\Gamma(k/2) = 2\upsilon_1^k\sigma_{jj}^{k/2}k!.$$

Also recall that $\mathbb{E}(\chi_i|\boldsymbol{x}_i) \le c_h := 9f_u h/8$. Hence, for any $0 \le \lambda < (2\upsilon_1\sigma_{jj}^{1/2})^{-1}$,

$$\begin{aligned}
\mathbb{E}e^{\lambda g\chi_i x_j} &= 1 + \frac{1}{2}\lambda^2\mathbb{E}(\chi_i x_j)^2 + \sum_{k=3}^{\infty}\frac{\lambda^k}{k!}\mathbb{E}(g\chi_i x_j)^k \\
&\le 1 + \frac{1}{2}c_h\sigma_{jj}\lambda^2 + 2c_h\sum_{\ell=2}^{\infty}\frac{\lambda^{2\ell}}{(2\ell)!}\upsilon_1^{2\ell}\sigma_{jj}^{\ell}(2\ell)! \\
&= 1 + \frac{1}{2}c_h\sigma_{jj}\lambda^2 + 2c_h\sum_{\ell=2}^{\infty}\upsilon_1^{2\ell}\sigma_{jj}^{\ell}\lambda^{2\ell} \\
&\le 1 + \frac{1}{2}\upsilon_1^2 c_h\sigma_{jj}\sum_{k=2}^{\infty}\lambda^k(2\upsilon_1\sigma_{jj}^{1/2})^{k-2} \\
&\le 1 + \frac{1}{2}\frac{\upsilon_1^2 c_h\sigma_{jj}\lambda^2}{1 - 2\upsilon_1\sigma_{jj}^{1/2}\lambda}.
\end{aligned}$$

It then follows that $\log\mathbb{E}e^{\lambda S_j} \le \frac{1}{2}\frac{\lambda^2\cdot\upsilon_1^2 c_h\sigma_{jj}n}{1-2\upsilon_1\sigma_{jj}^{1/2}\lambda}$ for any $\lambda \in (0, (2\upsilon_1\sigma_{jj}^{1/2})^{-1})$. By symmetry, the same bound applies to $-S_j$. Applying Corollary 2.6 in Boucheron, Lugosi and Massart (2013), we obtain

$$\mathbb{E}\Bigg(\max_{1\le j\le p}\bigg|\frac{1}{n}\sum_{i=1}^{n}g_i\chi_i x_{ij}\bigg|\Bigg) = \mathbb{E}\max_{1\le j\le p}|S_j/n| \le \upsilon_1\sigma_{\boldsymbol{x}}\Bigg\{\frac{3}{2}\sqrt{\frac{f_u h\log(2p)}{n}} + \frac{2\log(2p)}{n}\Bigg\}. \tag{G.29}$$

Finally, take $r = h/(24\upsilon_1^2)$. Combining (G.28), (G.29) with the concentration bound (G.22), we conclude that with probability at least $1-p^{-1}$, $\Omega(r,l) \le 0.16 f_l$ as long as $nh \gtrsim f_u f_l^{-2}\max\{s, l^2\log(p)\}$. This, together with (G.19), (G.20) and (G.21), proves the claim. □

G.8 Proof of Lemma E.3

Let $\boldsymbol{w}_h(\boldsymbol{\beta}) = \nabla\widehat{Q}_h(\boldsymbol{\beta}) - \nabla Q_h(\boldsymbol{\beta})$ be the centered score function as in (4.10). From the decomposition $\nabla\widehat{Q}_h(\widehat{\boldsymbol{\beta}}^{\rm ora}) = \boldsymbol{w}_h(\widehat{\boldsymbol{\beta}}^{\rm ora}) - \boldsymbol{w}_h(\boldsymbol{\beta}^*) + \nabla Q_h(\widehat{\boldsymbol{\beta}}^{\rm ora}) + \boldsymbol{w}_h(\boldsymbol{\beta}^*)$, we have

$$\|\boldsymbol{w}_h^{\rm ora}\|_\infty \le \|\boldsymbol{w}_h(\widehat{\boldsymbol{\beta}}^{\rm ora}) - \boldsymbol{w}_h(\boldsymbol{\beta}^*)\|_\infty + \|\nabla Q_h(\widehat{\boldsymbol{\beta}}^{\rm ora})\|_\infty + \|\boldsymbol{w}_h(\boldsymbol{\beta}^*)\|_\infty.$$

For $\|\boldsymbol{w}_h(\boldsymbol{\beta}^*)\|_\infty = \|\boldsymbol{w}_h^*\|_\infty$, applying Lemma C.2 to $\|\boldsymbol{w}_{h,\mathcal{S}}^*\|_\infty$ and $\|\boldsymbol{w}_{h,\mathcal{S}^c}^*\|_\infty$ separately, we obtain that the following bounds

$$\|\boldsymbol{w}_{h,\mathcal{S}}^*\|_\infty \lesssim \sigma_{\boldsymbol{x}}\sqrt{\frac{\log(2s)+t}{n}} \quad\text{and}\quad \|\boldsymbol{w}_{h,\mathcal{S}^c}^*\|_\infty \lesssim \sigma_{\boldsymbol{x}}\sqrt{\frac{\log(2p)}{n}} \tag{G.30}$$

hold with probability at least $1 - e^{-t}$ and $1 - (2p)^{-1}$, respectively, provided $n \gtrsim \log(2p) + t$.

In the following, we control the other two terms $\|\boldsymbol{w}_h(\widehat{\boldsymbol{\beta}}^{\rm ora}) - \boldsymbol{w}_h(\boldsymbol{\beta}^*)\|_\infty$ and $\|\nabla Q_h(\widehat{\boldsymbol{\beta}}^{\rm ora})\|_\infty$, separately, via empirical process arguments. The main difficulty is that the oracle $\widehat{\boldsymbol{\beta}}^{\rm ora}$ is also random and does not have a closed-form expression like the least squares estimator.

Step 1: Bounding $\|\boldsymbol{w}_h(\widehat{\boldsymbol{\beta}}^{\rm ora}) - \boldsymbol{w}_h(\boldsymbol{\beta}^*)\|_\infty$. Define the oracle local neighborhood $\Theta_{\mathcal{S}}^*(r) = \{\boldsymbol{\beta} \in \boldsymbol{\beta}^* + \mathbb{B}_{\boldsymbol{\Sigma}}(r) : \boldsymbol{\beta}_{\mathcal{S}^c} = \mathbf{0}\}$. Conditioned on the event $\{\widehat{\boldsymbol{\beta}}^{\rm ora} \in \boldsymbol{\beta}^* + \mathbb{B}_{\boldsymbol{\Sigma}}(r)\}$,

$$\|\boldsymbol{w}_h(\widehat{\boldsymbol{\beta}}^{\rm ora}) - \boldsymbol{w}_h(\boldsymbol{\beta}^*)\|_\infty \le \sup_{\boldsymbol{\beta}\in\Theta_{\mathcal{S}}^*(r)}\|\boldsymbol{w}_h(\boldsymbol{\beta}) - \boldsymbol{w}_h(\boldsymbol{\beta}^*)\|_\infty. \tag{G.31}$$

For $j \in [p]$, let $\mathbf{e}_j \in \mathbb{R}^p$ be the canonical basis vectors in $\mathbb{R}^p$. For every $\boldsymbol{\beta} \in \Theta^*_{\mathcal{S}}(r)$, write $\boldsymbol{\delta} = (\boldsymbol{\beta} - \boldsymbol{\beta}^*)_{\mathcal{S}} \in \mathbb{R}^s$, and note that $\|\boldsymbol{\beta} - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} = \|\boldsymbol{\delta}\|_{\mathbf{S}}$, where $\mathbf{S} = \boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}$. Hence,

$$\sup_{\boldsymbol{\beta} \in \Theta^*_{\mathcal{S}}(r)} \|\boldsymbol{w}_h(\boldsymbol{\beta}) - \boldsymbol{w}_h(\boldsymbol{\beta}^*)\|_\infty \leq \sigma_{\boldsymbol{x}} \max_{1 \leq j \leq p} \sup_{\|\boldsymbol{\delta}\|_{\mathbf{S}} \leq r} |W_j(\boldsymbol{\delta})|, \tag{G.32}$$

where $W_j(\boldsymbol{\delta}) = (1/n) \sum_{i=1}^n (w_{ij} - \mathbb{E} w_{ij})$, $w_{ij} = \bar{x}_{ij} \{\bar{K}((\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}\rangle - \varepsilon_i)/h) - \bar{K}(-\varepsilon_i/h)\}$ and $\bar{x}_{ij} = x_{ij}/\sigma_{jj}^{1/2}$.

We will apply a concentration result for empirical processes in Spokoiny (2012) to bound the local fluctuation $\sup_{\|\boldsymbol{\delta}\|_{\mathbf{S}} \leq r} W_j(\boldsymbol{\delta})$. To this end, we need to control the exponential moments of $W_j(\boldsymbol{\delta})$. Note that

$$\begin{aligned}
\mathbb{E}\bigg\{ \bar{K}\bigg(\frac{\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}\rangle - \varepsilon_i}{h} \bigg) \bigg| \boldsymbol{x} \bigg\} &= \int_{-\infty}^{\infty} \bar{K}((\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}\rangle - t)/h) f_{\varepsilon_i|\boldsymbol{x}_i}(t)\, \mathrm{d}t \\
&= h \int_{-\infty}^{\infty} \bar{K}(u) f_{\varepsilon_i|\boldsymbol{x}_i}(\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}\rangle - uh)\, \mathrm{d}u \\
&= \int_{-\infty}^{\infty} F_{\varepsilon_i|\boldsymbol{x}_i}(\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}\rangle - uh) K(u)\, \mathrm{d}u.
\end{aligned}$$

Similarly, $\mathbb{E}\{\bar{K}(-\varepsilon_i/h)|\boldsymbol{x}\} = \int_{-\infty}^{\infty} F_{\varepsilon_i|\boldsymbol{x}_i}(-uh)K(u)\, \mathrm{d}u$. Under Conditions (B1′) and (B2′), we have $|w_{ij}| \leq h^{-1} |\bar{x}_{ij} \langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}\rangle|$ and $|\mathbb{E}(w_{ij})| \leq f_u \mathbb{E}|\bar{x}_{ij}\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}\rangle| \leq f_u \|\boldsymbol{\delta}\|_{\mathbf{S}}$. Moreover, by Minkowski's integral inequality, it can be shown that

$$\mathbb{E}(w_{ij}^2|\boldsymbol{x}_i) \leq f_u h^{-1} \bar{x}_{ij}^2 \langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}\rangle^2.$$

The above bounds together imply

$$\mathbb{E}\{(w_{ij} - \mathbb{E}w_{ij})^2|\boldsymbol{x}_i\} \leq 2(f_u^2 \|\boldsymbol{\delta}\|_{\mathbf{S}}^2 + f_u h^{-1} \bar{x}_{ij}^2 \langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}\rangle^2).$$

For every $\lambda \in \mathbb{R}$ and $\boldsymbol{\delta} \in \mathbb{R}^s$, write $\lambda_* = \lambda/\|\boldsymbol{\delta}\|_{\mathbf{S}}$ and $\boldsymbol{\delta}_* = \boldsymbol{\delta}/\|\boldsymbol{\delta}\|_{\mathbf{S}}$. Then, by the elementary inequality $|e^u - 1 - u| \leq (u^2/2)e^{u \vee 0}$, we obtain

$$\begin{aligned}
&\mathbb{E} e^{\lambda W_j(\boldsymbol{\delta})/\|\boldsymbol{\delta}\|_{\mathbf{S}}} = \prod_{i=1}^n \mathbb{E} e^{\frac{\lambda_*}{n}(w_{ij} - \mathbb{E}w_{ij})} \\
&\leq \prod_{i=1}^n \mathbb{E}\bigg\{ 1 + \frac{\lambda_*^2}{2n^2} (w_{ij} - \mathbb{E}w_{ij})^2 e^{\frac{|\lambda_*|}{n}|w_{ij} - \mathbb{E}w_{ij}|} \bigg\} \\
&\leq \prod_{i=1}^n \bigg\{ 1 + \frac{\lambda_*^2}{2n^2} e^{\frac{|\lambda| f_u}{n}} \mathbb{E}(w_{ij} - \mathbb{E}w_{ij})^2 e^{\frac{|\lambda|}{nh}|\bar{x}_{ij}\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}_*\rangle|} \bigg\} \\
&\leq \prod_{i=1}^n \bigg\{ 1 + \frac{\lambda^2 f_u^2}{n^2} e^{\frac{|\lambda| f_u}{n}} \mathbb{E} e^{\frac{|\lambda|}{nh}|\bar{x}_{ij}\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}_*\rangle|} + \frac{\lambda^2 f_u}{n^2 h} e^{\frac{|\lambda| f_u}{n}} \mathbb{E} \bar{x}_{ij}^2 \langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}_*\rangle^2 e^{\frac{|\lambda|}{nh}|\bar{x}_{ij}\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}_*\rangle|} \bigg\}.
\end{aligned} \tag{G.33}$$

Applying Hölder's inequality to the exponential moments on the right-hand side of (G.33) yields that, for any $t > 0$,

$$\mathbb{E} \bar{x}_{ij}^2 \langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}_*\rangle^2 e^{t|\bar{x}_{ij}\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}_*\rangle|} \leq (\mathbb{E}\bar{x}_{ij}^4 e^{t\bar{x}_{ij}^2})^{1/2} \cdot (\mathbb{E}\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}_*\rangle^4 e^{t\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}_*\rangle^2})^{1/2}$$

and

$$\mathbb{E} e^{t|\bar{x}_{ij}\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}_*\rangle|} \leq (\mathbb{E} e^{t\bar{x}_{ij}^2})^{1/2} \cdot (\mathbb{E} e^{t\langle \boldsymbol{x}_{i,\mathcal{S}}, \boldsymbol{\delta}_*\rangle^2})^{1/2}.$$

For any unit vector $\boldsymbol{u} \in \mathbb{S}^{p-1}$, let $Z_{\boldsymbol{u}} = (\boldsymbol{z}^{\mathrm{T}}\boldsymbol{u})^2/(4v_1^2)$, where $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}$. By Condition (B3′), $\mathbb{P}(Z_{\boldsymbol{u}} \geq u) \leq 2e^{-2u}$ for any $u \geq 0$. It can be shown that

$$\mathbb{E}e^{Z_{\boldsymbol{u}}} = 1 + \int_0^\infty e^u \mathbb{P}(Z_{\boldsymbol{u}} \geq u)\mathrm{d}u \leq 3 \quad \text{and} \quad \mathbb{E}Z_{\boldsymbol{u}}^2 e^{Z_{\boldsymbol{u}}} = \int_0^\infty (u^2 + 2u)e^u \mathbb{P}(Z_{\boldsymbol{u}} \geq u)\mathrm{d}u \leq 8.$$

Substituting the above moment bounds into (G.33), we find that for any $|\lambda| \leq \min\{nh/(4v_1^2), n/f_u\}$,

$$\mathbb{E}e^{\lambda W_j(\boldsymbol{\delta})/\|\boldsymbol{\delta}\|_{\mathbf{S}}} \leq \prod_{i=1}^n \{1 + Cv_1^4 f_u/(n^2h)\} \leq e^{Cv_1^4 f_u/(nh)},$$

where $C > 0$ is an absolute constant. Similarly, it can be derived that for each pair $(\boldsymbol{\delta}, \boldsymbol{\delta}')$,

$$\mathbb{E}e^{\lambda\{W_j(\boldsymbol{\delta}) - W_j(\boldsymbol{\delta}')\}/\|\boldsymbol{\delta}-\boldsymbol{\delta}'\|_{\mathbf{S}}} \leq e^{Cv_1^4 f_u/(nh)} \quad \text{for all} \quad |\lambda| \leq \min\{nh/(4v_1^2), n/f_u\}.$$

The above inequality certifies condition $(\mathcal{E}d)$ in Spokoiny (2012) (see Section 2 in the supplement), so that Corollary 2.2 therein applies to the process $\{W_j(\boldsymbol{\delta}) : \|\boldsymbol{\delta}\|_{\mathbf{S}} \leq r\}$: with probability at least $1 - e^{-u}$,

$$\sup_{\boldsymbol{\beta} \in \Theta_{\mathcal{S}}^*(r)} \langle \boldsymbol{w}_h(\boldsymbol{\beta}) - \boldsymbol{w}_h(\boldsymbol{\beta}^*), \mathbf{e}_j \rangle = \sup_{\|\boldsymbol{\delta}\|_{\mathbf{S}} \leq r} W_j(\boldsymbol{\delta}) \lesssim v_1^2 f_u^{1/2} \sigma_{\boldsymbol{x}} r \sqrt{\frac{s+u}{nh}},$$

provided $nh \gtrsim (s+u)^{1/2}$. The same bound applies to $\sup_{\|\boldsymbol{\delta}\|_{\mathbf{S}} \leq r} -W_j(\boldsymbol{\delta})$ by a similar argument. Taking $u = 2\log(2p)$ in (G.32), and using the union bound, we obtain

$$\sup_{\boldsymbol{\beta} \in \Theta_{\mathcal{S}}^*(r)} \|\boldsymbol{w}_h(\boldsymbol{\beta}) - \boldsymbol{w}_h(\boldsymbol{\beta}^*)\|_\infty \lesssim \sigma_{\boldsymbol{x}} r \sqrt{\frac{s + \log p}{nh}} \tag{G.34}$$

with probability at least $1 - (2p)^{-1}$ provided $nh \gtrsim (s + \log p)^{1/2}$.

Step 2: Bounding $\|\nabla Q_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}})\|_\infty$. As before, we write $\boldsymbol{\delta} = (\boldsymbol{\beta} - \boldsymbol{\beta}^*)_{\mathcal{S}}$ for $\boldsymbol{\beta} \in \Theta_{\mathcal{S}}^*(r)$. In face, since the oracle score is such that $\boldsymbol{w}_{h,\mathcal{S}}^{\mathrm{ora}} = \mathbf{0}$, it suffices to bound $\|\nabla Q_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}})_{\mathcal{S}^c}\|_\infty$ instead. Similarly to (B.5), we have $\|\nabla Q_h(\boldsymbol{\beta}^*)\|_\infty \leq 0.5 l_0 \kappa_2 \sigma_{\boldsymbol{x}} h^2$. For any $\boldsymbol{\beta} \in \Theta_{\mathcal{S}}^*(r)$, note that

$$\nabla Q_h(\boldsymbol{\beta})_{\mathcal{S}^c} - \nabla Q_h(\boldsymbol{\beta}^*)_{\mathcal{S}^c} = \mathbb{E}\int_{-\infty}^{\infty} K(u)\{F_{\varepsilon|\boldsymbol{x}}(\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta} - uh) - F_{\varepsilon|\boldsymbol{x}}(-uh)\}\,\mathrm{d}u \cdot \boldsymbol{x}_{\mathcal{S}^c}.$$

Using the Taylor series expansion twice, we get

$$F_{\varepsilon|\boldsymbol{x}}(\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta} - uh) - F_{\varepsilon|\boldsymbol{x}}(-uh) = f_{\varepsilon|\boldsymbol{x}}(0) \cdot \boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta} + \int_0^{\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta}} \{f_{\varepsilon|\boldsymbol{x}}(t - hu) - f_{\varepsilon|\boldsymbol{x}}(0)\}\,\mathrm{d}t.$$

Together, the last two displays imply

$$\begin{aligned}
&\big\|\nabla Q_h(\boldsymbol{\beta})_{\mathcal{S}^c} - \nabla Q_h(\boldsymbol{\beta}^*)_{\mathcal{S}^c} - \mathbf{J}_{\mathcal{S}^c\mathcal{S}}\boldsymbol{\delta}\big\|_\infty \\
&\leq 0.5 l_0 \max_{j \in \mathcal{S}^c} \mathbb{E}|x_j|(\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta})^2 + \kappa_1 h \max_{j \in \mathcal{S}^c} \mathbb{E}|x_j \boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta}| \\
&\leq 0.5 l_0 \mu_4^{1/2} \sigma_{\boldsymbol{x}} \|\boldsymbol{\delta}\|_{\mathbf{S}}^2 + l_0 \kappa_1 \sigma_{\boldsymbol{x}} h \|\boldsymbol{\delta}\|_{\mathbf{S}},
\end{aligned}$$

where $\mathbf{J}_{\mathcal{S}^c\mathcal{S}} = \mathbb{E}\{f_{\varepsilon|\boldsymbol{x}}(0)\boldsymbol{x}_{\mathcal{S}^c}\boldsymbol{x}_{\mathcal{S}}\} \in \mathbb{R}^{(p-s)\times s}$. Putting together the pieces, we have shown that conditioned on $\{\widehat{\boldsymbol{\beta}}^{\mathrm{ora}} \in \boldsymbol{\beta}^* + \mathbb{B}_{\boldsymbol{\Sigma}}(r)\}$,

$$\big\|\nabla Q_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}})_{\mathcal{S}^c} - \mathbf{J}_{\mathcal{S}^c\mathcal{S}}(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}} - \boldsymbol{\beta}^*)_{\mathcal{S}}\big\|_\infty \leq 0.5 l_0 \sigma_{\boldsymbol{x}}(\mu_4^{1/2} r^2 + 2\kappa_1 h r + \kappa_2 h^2). \tag{G.35}$$

It remains to control $\|\mathbf{J}_{\mathcal{S}^c\mathcal{S}}(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}}\|_\infty$, which is closely related to the ℓ_∞-error of the oracle estimator. By (4.26),

$$
\begin{aligned}
&\big\|\mathbf{J}_{\mathcal{S}^c\mathcal{S}}(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}}\big\|_\infty = \big\|\mathbf{J}_{\mathcal{S}^c\mathcal{S}}(\mathbf{J}_{\mathcal{S}\mathcal{S}})^{-1}\mathbf{J}_{\mathcal{S}\mathcal{S}}(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}}\big\|_\infty \\
&\leq \max_{j\in\mathcal{S}^c}\|\mathbf{J}_{j\mathcal{S}}(\mathbf{J}_{\mathcal{S}\mathcal{S}})^{-1}\|_1 \cdot \|\mathbf{J}_{\mathcal{S}\mathcal{S}}(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}}\|_\infty \leq A_0 \cdot \|\mathbf{J}_{\mathcal{S}\mathcal{S}}(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}}\|_\infty. \qquad \text{(G.36)}
\end{aligned}
$$

Next we derive a sharper bound for the oracle error $(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}}$ under ℓ_∞-norm, instead of using the trivial bound $\|\mathbf{J}_{\mathcal{S}\mathcal{S}}(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}}\|_\infty \leq \|\mathbf{J}_{\mathcal{S}\mathcal{S}}(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}}\|_2$. By Proposition 4.3,

$$
\|(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}}\|_{\mathbf{S}} \lesssim f_l^{-1}\Bigg(\sqrt{\frac{s+t}{n}} + h^2\Bigg)
$$

and

$$
\Bigg\|\mathbf{D}(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}} + \underbrace{\frac{1}{n}\sum_{i=1}^n\{\bar{K}(-\varepsilon_i/h) - \tau\}\boldsymbol{x}_{i,\mathcal{S}}}_{=\nabla\widehat{Q}_h(\beta^*)_{\mathcal{S}}}\Bigg\|_{\mathbf{S}^{-1}} \lesssim \frac{s+t}{h^{1/2}n} + h\sqrt{\frac{s+t}{n}} + h^3
$$

hold with probability at least $1-3e^{-t}$, where $\mathbf{D} = \mathbf{J}_{\mathcal{S}\mathcal{S}}$. The latter, combined with an earlier bound in (G.30), implies

$$
\begin{aligned}
&\|\mathbf{D}(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}}\|_\infty \\
&\leq \|\mathbf{D}(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}} + \nabla\widehat{Q}_h(\beta^*)_{\mathcal{S}}\|_\infty + \|\nabla\widehat{Q}_h(\beta^*)_{\mathcal{S}}\|_\infty \\
&\leq \|\mathbf{D}(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}} + \nabla\widehat{Q}_h(\beta^*)_{\mathcal{S}}\|_2 + \|\boldsymbol{w}^*_{h,\mathcal{S}}\|_\infty + \|\nabla Q_h(\beta^*)_{\mathcal{S}}\|_\infty \\
&\leq \gamma_1(\mathbf{S})^{1/2}\cdot\|\mathbf{D}(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}} + \nabla\widehat{Q}_h(\beta^*)_{\mathcal{S}}\|_{\mathbf{S}^{-1}} + \|\boldsymbol{w}^*_{h,\mathcal{S}}\|_\infty + \|\nabla Q_h(\beta^*)_{\mathcal{S}}\|_\infty \\
&\lesssim \sqrt{\frac{\log(s)+t}{n}} + \frac{s+t}{h^{1/2}n} + h\sqrt{\frac{s+t}{n}} + h^2. \qquad \text{(G.37)}
\end{aligned}
$$

Combining the bounds (G.30), (G.31), (G.34), (G.35), (G.36) and (G.37), we find that, with probability at least $1-p^{-1}-4e^{-t}$,

$$
\|\nabla Q_h(\widehat{\beta}^{\rm ora})\|_\infty \lesssim \sqrt{\frac{\log(2p)}{n}} + A_0\Bigg\{\sqrt{\frac{\log(s)+t}{n}} + \frac{s+t}{h^{1/2}n} + h^2\Bigg\}
$$

provided that $\sqrt{(s\vee\log p+t)/n} \lesssim h \lesssim 1$. This proves (E.7).

Note that the ℓ_∞-error bound (G.37) does not imply the desired bound on $\|(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}}\|_\infty$ directly. Using the same arguments, we obtain

$$
\begin{aligned}
&\|(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}}\|_\infty \\
&\leq \|(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}} + \mathbf{D}^{-1}\nabla\widehat{Q}_h(\beta^*)_{\mathcal{S}}\|_\infty + \|\mathbf{D}^{-1}\nabla\widehat{Q}_h(\beta^*)_{\mathcal{S}}\|_\infty \\
&\leq \|(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}} + \mathbf{D}^{-1}\nabla\widehat{Q}_h(\beta^*)_{\mathcal{S}}\|_2 + \|\mathbf{D}^{-1}\boldsymbol{w}^*_{h,\mathcal{S}}\|_\infty + \|\mathbf{D}^{-1}\nabla Q_h(\beta^*)_{\mathcal{S}}\|_\infty \\
&\leq f_l^{-1}\gamma_s(\mathbf{S})^{-1/2}\cdot\|\mathbf{D}(\widehat{\beta}^{\rm ora} - \beta^*)_{\mathcal{S}} + \nabla\widehat{Q}_h(\beta^*)_{\mathcal{S}}\|_{\mathbf{S}^{-1}} \\
&\qquad + \|\mathbf{D}^{-1}\boldsymbol{w}^*_{h,\mathcal{S}}\|_\infty + f_l^{-1}\gamma_s(\mathbf{S})^{-1/2}\cdot\|\mathbf{S}^{-1/2}\nabla Q_h(\beta^*)_{\mathbf{S}}\|_2,
\end{aligned}
$$

where we have used the fact $f_l \cdot \gamma_s(\mathbf{S}) \le \gamma_s(\mathbf{D}) \le \gamma_1(\mathbf{D}) \le f_u \cdot \gamma_1(\mathbf{S})$. By (B.5), $\|\mathbf{S}^{-1/2}\nabla Q_h(\boldsymbol{\beta}^*)_{\mathbf{S}}\|_2 \le 0.5 l_0 \kappa_2 h^2$. For $\|\mathbf{D}^{-1}\boldsymbol{w}^*_{h,\mathcal{S}}\|_\infty$, note that

$$\|\mathbf{D}^{-1}\boldsymbol{w}^*_{h,\mathcal{S}}\|_\infty = \max_{1\le j\le s}\left|\frac{1}{n}\sum_{i=1}^n (1-\mathbb{E})\{\bar{K}(-\varepsilon_i/h)-\tau\}\langle \mathbf{S}^{-1/2}\boldsymbol{x}_{i,\mathcal{S}}, \mathbf{S}^{1/2}\mathbf{D}^{-1}\mathbf{e}_j\rangle\right|,$$

where $\mathbf{e}_j$'s are canonical basis vectors in $\mathbb{R}^s$, and satisfy $\|\mathbf{S}^{1/2}\mathbf{D}^{-1}\mathbf{e}_j\|_2 \le f_l^{-1}\gamma_s(\mathbf{S})^{-1/2}$. Following the proof of Lemma C.2, it can be similarly shown that, with probability at least $1-e^{-t}$,

$$\|\mathbf{D}^{-1}\boldsymbol{w}^*_{h,\mathcal{S}}\|_\infty \lesssim \sqrt{\frac{\log(s)+t}{n}}.$$

Putting together the pieces yields the stated result (E.8). □

G.9 Proof of Lemma F.1

The proof parallels that of Lemma E.3, and therefore we only provide an outline of the proof. Recall that $\boldsymbol{w}_h^{\mathrm{ora}} = \nabla\widehat{Q}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}})$, $\boldsymbol{w}_h(\boldsymbol{\beta}) = \nabla\widehat{Q}_h(\boldsymbol{\beta}) - \nabla Q_h(\boldsymbol{\beta})$ and $\nabla Q_h(\boldsymbol{\beta}_h^*) = \mathbf{0}$, where $Q_h(\boldsymbol{\beta}) = \mathbb{E}\widehat{Q}_h(\boldsymbol{\beta})$. From the decomposition $\nabla\widehat{Q}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}) = \boldsymbol{w}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}) - \boldsymbol{w}_h(\boldsymbol{\beta}_h^*) + \nabla Q_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}) + \boldsymbol{w}_h(\boldsymbol{\beta}_h^*)$ we see that

$$\|\boldsymbol{w}_h^{\mathrm{ora}}\|_\infty \le \|\boldsymbol{w}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}) - \boldsymbol{w}_h(\boldsymbol{\beta}_h^*)\|_\infty + \|\nabla Q_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}})\|_\infty + \|\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)\|_\infty.$$

In fact, since $\nabla\widehat{Q}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}})_{\mathcal{S}} = \mathbf{0}$, essentially we only need to control $\|(\boldsymbol{w}_h^{\mathrm{ora}})_{\mathcal{S}^c}\|_\infty$.

To bound $\|\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)\|_\infty$, treating $\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)_{\mathcal{S}} \in \mathbb{R}^s$ and $\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)_{\mathcal{S}^c} \in \mathbb{R}^{p-s}$ separately, it can be shown that as long as $n \gtrsim \log(2p)+t$,

$$\|\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty \lesssim \sigma_{\boldsymbol{x}}\sqrt{\frac{\log(2s)+t}{n}} \quad \text{and} \quad \|\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)_{\mathcal{S}^c}\|_\infty \lesssim \sigma_{\boldsymbol{x}}\sqrt{\frac{\log(2p)}{n}} \tag{G.38}$$

hold with probability at least $1-e^{-t}$ and $1-(2p)^{-1}$, respectively. Turning to $\|\boldsymbol{w}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}) - \boldsymbol{w}_h(\boldsymbol{\beta}_h^*)\|_\infty$, following the proof of (G.34) it can be similarly shown that with probability at least $1-(2p)^{-1}$,

$$\sup_{\boldsymbol{\beta}\in\Theta^*_{h,\mathcal{S}}(r)} \|\boldsymbol{w}_h(\boldsymbol{\beta}) - \boldsymbol{w}_h(\boldsymbol{\beta}^*)\|_\infty \lesssim \sigma_{\boldsymbol{x}} r\sqrt{\frac{s+\log(p)}{nh}} \tag{G.39}$$

provided $nh \gtrsim (s+\log p)^{1/2}$, where $\Theta^*_{h,\mathcal{S}}(r) = \{\boldsymbol{\beta} \in \boldsymbol{\beta}_h^* + \mathbb{B}_{\boldsymbol{\Sigma}}(r) : \boldsymbol{\beta}_{\mathcal{S}^c} = \mathbf{0}\}$.

It remains to bound $\|\nabla Q_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}})_{\mathcal{S}^c}\|_\infty$. Write $\boldsymbol{\delta} = (\boldsymbol{\beta}-\boldsymbol{\beta}_h^*)_{\mathcal{S}}$ for $\boldsymbol{\beta} \in \Theta^*_{h,\mathcal{S}}(r)$, and note that

$$\nabla Q_h(\boldsymbol{\beta}) - \underbrace{\nabla Q_h(\boldsymbol{\beta}_h^*)}_{=\mathbf{0}} = \mathbb{E}\int_{-\infty}^{\infty} K(u)\{F_\varepsilon(\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta} + b_h - uh) - F_\varepsilon(b_h - uh)\}\,\mathrm{d}u \cdot \boldsymbol{x}.$$

A Taylor expansion with integral remainder leads to

$$\begin{aligned} &F_\varepsilon(\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta} + b_h - uh) \\ &= F_\varepsilon(b_h - uh) + f_\varepsilon(b_h - uh)\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta} + \int_0^{\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta}} \{f_\varepsilon(t+b_h-uh) - f_\varepsilon(b_h-uh)\}\,\mathrm{d}t. \end{aligned}$$

Noting that $\int K(u)f_\varepsilon(b_h-uh)\,\mathrm{d}u=\int K_h(b_h-t)f_\varepsilon(t)\,\mathrm{d}t=\mathbb{E}K_h(b_h-\varepsilon)=m''(b_h)$, it follows that

$$\|\nabla Q_h(\boldsymbol{\beta})_{\mathcal{S}^c}-m''(b_h)\cdot\boldsymbol{\Sigma}_{\mathcal{S}^c\mathcal{S}}\boldsymbol{\delta}\|_\infty\le\frac{l_0}{2}\max_{j\in\mathcal{S}^c}\mathbb{E}\{|x_j|(\boldsymbol{x}_{\mathcal{S}}^{\mathrm{T}}\boldsymbol{\delta})^2\}\le\frac{l_0}{2}\mu_4^{1/2}\sigma_{\boldsymbol{x}}\|\boldsymbol{\delta}\|_{\mathbf{S}}^2.$$

Conditioned on the event $\{\|\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}-\boldsymbol{\beta}_h^*\|_{\boldsymbol{\Sigma}}\le r\}$, this implies

$$\|\nabla Q_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}})_{\mathcal{S}^c}\|_\infty\le m''(b_h)\cdot\|\boldsymbol{\Sigma}_{\mathcal{S}^c\mathcal{S}}(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}-\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty+\frac{l_0}{2}\mu_4^{1/2}\sigma_{\boldsymbol{x}}r^2. \tag{G.40}$$

Next we bound $\|\boldsymbol{\Sigma}_{\mathcal{S}^c\mathcal{S}}(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}-\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty$ using the Bahadur representation in Proposition A.2. Recall that $\mathbf{D}_h=m''(b_h)\mathbf{S}$ and $\mathbf{S}=\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}$, we have

$$\begin{aligned}&m''(b_h)\|\boldsymbol{\Sigma}_{\mathcal{S}^c\mathcal{S}}(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}-\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty=\|\boldsymbol{\Sigma}_{\mathcal{S}^c\mathcal{S}}\mathbf{S}^{-1}\mathbf{D}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}-\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty\\&\le\max_{j\in\mathcal{S}^c}\|\boldsymbol{\Sigma}_{j\mathcal{S}}(\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}})^{-1}\|_1\cdot\|\mathbf{D}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}-\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty\le A_1\|\mathbf{D}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}-\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty,\end{aligned}$$

where the last step is due to condition (A.4). In view of Proposition A.2, we write $\mathbf{D}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}-\boldsymbol{\beta}_h^*)_{\mathcal{S}}=-\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)_{\mathcal{S}}+\boldsymbol{r}_h$, where $\boldsymbol{r}_h\in\mathbb{R}^s$ is the remainder of the Bahadur representation. Together, (G.38) and Proposition A.2 imply that with probability at least $1-4e^{-t}$,

$$\begin{aligned}&\|\mathbf{D}_h(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}-\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty\le\|\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty+\|\boldsymbol{r}_h\|_\infty\\&\le\|\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty+\|\boldsymbol{r}_h\|_2\le\|\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty+\gamma_1(\mathbf{S})^{1/2}\|\mathbf{S}^{-1/2}\boldsymbol{r}_h\|_2\lesssim\sqrt{\frac{\log(s)+t}{n}}+\frac{s+t}{h^{1/2}n}\end{aligned} \tag{G.41}$$

and

$$\|(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}-\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_{\mathbf{S}}\lesssim\sqrt{\frac{s+t}{n}}, \tag{G.42}$$

provided that $\sqrt{(s+t)/n}\lesssim h\lesssim 1$. Combining (G.38)–(G.42) proves (F.8).

Back to oracle estimator, from the previous decomposition we have

$$\begin{aligned}&\|\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}-\boldsymbol{\beta}_h^*\|_\infty=\|(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}-\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty\\&\le\|(\widehat{\boldsymbol{\beta}}^{\mathrm{ora}}-\boldsymbol{\beta}_h^*)_{\mathcal{S}}+\mathbf{D}_h^{-1}\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty+\|\mathbf{D}_h^{-1}\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty\\&\le\frac{1}{m''(b_h)\sqrt{\gamma_s(\mathbf{S})}}\|\mathbf{S}^{-1/2}\boldsymbol{r}_h\|_2+\frac{1}{m''(b_h)}\|\mathbf{S}^{-1}\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty.\end{aligned}$$

Analogous to the first bound in (G.38), the ℓ_∞-norm of $\mathbf{S}^{-1}\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)_{\mathcal{S}}\in\mathbb{R}^s$ can also be bounded as

$$\|\mathbf{S}^{-1}\boldsymbol{w}_h(\boldsymbol{\beta}_h^*)_{\mathcal{S}}\|_\infty\lesssim\sqrt{\frac{\log(s)+t}{n}}$$

with probability at least $1-e^{-t}$. Combining this with (A.3) proves (F.9). □

References

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

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

Bradic, J., Fan, J. and Wang, W. (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *J. R. Statist. Soc.* B **73** 325–349.

Fan, J., Liu, H., Sun, Q. and Zhang, T. (2018). I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error. *Ann. Statist.* **96** 1348–1360.

Fernandes, M., Guerre, E. and Horta, E. (2021). Smoothing quantile regressions. *J. Bus. Econ. Statist.* **39** 338–357.

He, X., Pan, X., Tan, K. M. and Zhou, W.-X. (2021). Smoothed quantile regression with large-scale inference. *J. Econometrics*, to appear. https://doi.org/10.1016/j.jeconom.2021.07.010.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer-Verlag, Berlin.

Loh, P. (2017). Statistical consistency and asymptotic normality for high-dimensional robust M-estimators. *Ann. Statist.* **45** 866–896.

Spokoiny, V. (2012). Parametric estimation. Finite sample theory. *Ann. Statist.* **40** 2877–2909.

Vershynin, R. (2018). *High-Dimensional Probability*. Cambridge University Press, Cambridge.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.* **36** 1108–1126.