Supplement to “Smoothed Quantile Regression with Large-Scale Inference”

Xuming He*, Xiaoou Pan†, Kean Ming Tan‡, and Wen-Xin Zhou§

A One-step Conquer with Higher-order Kernels

As noted in Section 3.1, the smoothing bias is of order $h^2$ when a non-negative kernel is used. The ensuing empirical loss $\beta \mapsto (1/n) \sum_{i=1}^{n} (\rho_{\tau} * K_h)(y_i - \langle x_i, \beta \rangle)$ is not only twice-differentiable and convex, but also (provably) strongly convex in a local vicinity of $\beta^*$ with high probability. Kernel smoothing is ubiquitous in nonparametric statistics. The order of a kernel, $\nu$, is defined as the order of the first non-zero moment. The order of a symmetric kernel is always even. A kernel is called 

high-order if $\nu > 2$, which inevitably has negative parts and thus is no longer a probability density. Thus far we have focused on conquer with second-order kernels, and the resulting estimator achieves an $\ell_2$-error of the order $\sqrt{p/n + h^2}$.

Let $G(\cdot)$ be a higher-order symmetric kernel with order $\nu \geq 4$, and $b > 0$ be a bandwidth. Again, via convolution smoothing, we may consider a bias-reduced estimator that minimizes the empirical loss $\beta \mapsto \hat{Q}_b^G(\beta) := (1/n) \sum_{i=1}^{n} (\rho_{\tau} * G_b)(y_i - \langle x_i, \beta \rangle)$. This, however, leads to a non-convex optimization. Without further assumptions, finding a global minimum is computationally intractable: finding an $\epsilon$-suboptimal point for a $k$-times continuously differentiable loss function requires at least $\Omega((1/\epsilon)^{p/k})$ evaluations of the function and its first $k$ derivatives, ignoring problem-dependent constants; see Section 1.6 in Nemirovski and Yudin (1983). Instead, various gradient-based methods have been developed for computing stationary points, which are points $\beta$ with sufficiently small
gradient \( \| \nabla \hat{G}_b^i(\beta) \|_2 \leq \epsilon \), where \( \epsilon \geq 0 \) is optimization error. However, the equation \( \nabla \hat{G}_b^i(\beta) = 0 \) does not necessarily have a unique solution, whose statistical guarantees remain unknown.

Motivated by the classical one-step estimator (Bickel, 1975), we further propose a one-step conquer estimator using high-order kernels, which bypasses solving a large-scale non-convex optimization. To begin with, we choose two symmetric kernel functions, \( K : \mathbb{R} \rightarrow [0, \infty) \) with order two and \( G(\cdot) \) with order \( \nu \geq 4 \), and let \( h, b > 0 \) be two bandwidths. First, compute an initial conquer estimator \( \hat{\beta} \in \arg\min_{\beta \in \mathbb{R}} \hat{G}_b^K(\beta) \), where \( \hat{G}_b^K(\beta) = (1/n) \sum_{i=1}^n (\rho_r \ast K_b)(y_i - \langle x_i, \beta \rangle) \). Denote by \( \hat{r}_i = y_i - \langle x_i, \hat{\beta} \rangle \) for \( i = 1, \ldots, n \) the fitted residuals. Next, with slight abuse of notation, we define the one-step conquer estimator \( \tilde{\beta} \) as a solution to the equation \( \nabla^2 \hat{G}_b^i(\tilde{\beta})(\tilde{\beta} - \beta) = -\nabla \hat{G}_b^i(\beta) \), or equivalently,

\[
\left\{ \frac{1}{n} \sum_{i=1}^n G_b(\hat{r}_i)x_i x_i^T \right\}(\beta - \bar{\beta}) = \frac{1}{n} \sum_{i=1}^n \hat{G}(\hat{r}_i/b) + \tau - 1]x_i.
\]

(A.1)

where \( \hat{G}_b^K(\beta) = (1/n) \sum_{i=1}^n (\rho_r \ast G_b)(y_i - \langle x_i, \beta \rangle) \). Provided that \( \nabla^2 \hat{G}_b^i(\beta) \) is positive definite, the one-step conquer estimate \( \tilde{\beta} \) essentially performs a Newton-type step based on \( \hat{\beta} \):

\[
\tilde{\beta} = \hat{\beta} - (\nabla^2 \hat{G}_b^i(\hat{\beta})^{-1} \nabla \hat{G}_b^i(\hat{\beta}). \quad \text{(A.2)}
\]

In this case, \( \tilde{\beta} \) can be computed by the conjugate gradient method (Hestenes and Stiefel, 1952).

Theoretical properties of the one-step estimator \( \tilde{\beta} \) defined in (A.1), including the Bahadur representation and asymptotic normality with explicit Berry-Esseen bound, will be provided in on-line supplementary materials. For practical implementation, we consider higher-order Gaussian-based kernels. For \( r = 1, 2, \ldots \), the \((2r)\)-th order Gaussian kernels are

\[
G_{2r}(u) = \frac{(-1)^r \phi^{(2r-1)}(u)}{2^{r-1}(r-1)!} = \sum_{\ell=0}^{r-1} \frac{(-1)^\ell}{2^\ell \ell!} \phi^{(2\ell)}(u);
\]

see Section 2 of Wand and Schucany (1990). Integrating \( G_{2r}(\cdot) \) yields

\[
\tilde{G}_{2r}(v) = \int_{-\infty}^v G_{2r}(u) \, du = \sum_{\ell=0}^{r-1} \frac{(-1)^\ell}{2^\ell \ell!} \phi^{(2\ell-1)}(v).
\]

In fact, both \( G_{2r} \) and \( \tilde{G}_{2r} \) have simpler forms \( G_{2r}(u) = p_r(u)\phi(u) \) and \( \tilde{G}_{2r}(u) = \Phi(u) + P_r(u)\phi(u) \), where \( p_r(\cdot) \) and \( P_r(\cdot) \) are polynomials in \( u \). For example, \( p_1(u) = 1, P_1(u) = 0, P_2(u) = (-u^2 + 3)/2, P_2(u) = u/2, P_3(u) = (u^4 - 10u^2 + 15)/8, \) and \( P_3(u) = (-u^3 + 7u)/8 \). We refer to Oryshchenko
(2020) for more details when \( r \) is large.

## B Proofs

Recall that \( \Sigma = \mathbb{E}(xx^T) \) is positive definite. Throughout the proof, we write \( \Omega = \Sigma^{-1} \) and \( \|u\|_\Omega = \|\Sigma^{-1/2}u\|_2 \) for \( u \in \mathbb{R}^p \). By Hölder’s inequality, \( \langle u, v \rangle \leq \|u\|_\Sigma \cdot \|v\|_\Omega \) for any \( u, v \in \mathbb{R}^p \).

### B.1 Proof of Proposition 3.1

For every \( r > 0 \), define the ellipse \( \Theta(r) = \{ u \in \mathbb{R}^p : \|u\|_\Sigma \leq r \} \) and the local vicinity \( \Theta^* = \{ \beta \in \mathbb{R}^p : \beta - \beta^* \in \Theta(\kappa_2^1 h) \} \). Let \( \eta = \sup \{ u \in [0, 1] : u(\beta_h^* - \beta^*) \in \Theta(\kappa_2^1 h) \} \) and \( \tilde{\beta}^* = \beta^* + \eta(\beta_h^* - \beta^*) \).

By definition, \( \eta = 1 \) if \( \beta_h^* \in \Theta^* \) and \( \eta < 1 \) if \( \beta_h^* \notin \Theta^* \). In the latter case, \( \tilde{\beta}^* \in \partial \Theta^* \). By the convexity of \( \beta \mapsto Q_h(\beta) \) and Lemma C.1 in the supplementary material of Sun, Zhou and Fan (2020),

\[
0 \leq \langle \nabla Q_h(\tilde{\beta}^*) - \nabla Q_h(\beta^*), \tilde{\beta}^* - \beta^* \rangle \\
\leq \eta \cdot \langle \nabla Q_h(\beta_h^*) - \nabla Q_h(\beta^*), \beta_h^* - \beta^* \rangle = \langle -\nabla Q_h(\beta^*), \tilde{\beta}^* - \beta^* \rangle. \tag{B.1}
\]

It follows from the mean value theorem for vector-valued functions that

\[
\nabla Q_h(\tilde{\beta}^*) - \nabla Q_h(\beta^*) = \int_0^1 \nabla^2 Q_h((1-t)\beta^* + t\tilde{\beta}^*) \, dt (\tilde{\beta}^* - \beta^*), \tag{B.2}
\]

where \( \nabla^2 Q_h(\beta) = \mathbb{E}[K_h(y - (x, \beta))xx^T] \) for \( \beta \in \mathbb{R}^p \). With \( \delta = \beta - \beta^* \), note that

\[
\mathbb{E}[K_h(y - (x, \beta))] = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{u - (x, \delta)}{h}\right)f_{\epsilon|x}(u) \, du = \int_{-\infty}^{\infty} K(v)f_{\epsilon|x}((x, \delta) + hv) \, dv.
\]

By the Lipschitz continuity of \( f_{\epsilon|x}(\cdot) \),

\[
\mathbb{E}[K_h(y - (x, \beta))] = f_{\epsilon|x}(0) + R_h(\delta) \tag{B.3}
\]

with \( R_h(\delta) \) satisfying \( |R_h(\delta)| \leq l_0(|(x, \delta)| + \kappa_1 h) \). Substituting (B.3) into (B.1) and (B.2) yields

\[
\langle \nabla Q_h(\tilde{\beta}^*) - \nabla Q_h(\beta^*), \tilde{\beta}^* - \beta^* \rangle \\
\geq \|\tilde{\beta}^* - \beta^*\|^2_D - 0.5l_0\mathbb{E}|(x, \tilde{\beta}^* - \beta^*)|^3 - l_0\kappa_1 h \cdot \|\tilde{\beta}^* - \beta^*\|^3_\Sigma \\
\geq \|\tilde{\beta}^* - \beta^*\|^2_D - 0.5l_0m_3 \cdot \|\tilde{\beta}^* - \beta^*\|^3_\Sigma - l_0\kappa_1 h \cdot \|\tilde{\beta}^* - \beta^*\|^3_\Sigma. \tag{B.4}
\]
On the other hand, under model (3.1) we have

\[ \langle -\nabla Q_h(\beta^*), \tilde{\beta}^* - \beta^* \rangle \leq \|\nabla Q_h(\beta^*)\|_{\Omega} \cdot \|\tilde{\beta}^* - \beta^*\|_{\Sigma}. \]

where \( \nabla Q_h(\beta^*) = \mathbb{E}[\tilde{K}(-\varepsilon/h) - \tau]x \). By integration by parts and a Taylor series expansion,

\[
\mathbb{E}[\tilde{K}(-\varepsilon/h)|x] = \int_{-\infty}^{\infty} K(-t/h) dF_{\varepsilon|x}(t)
= -\frac{1}{h} \int_{-\infty}^{\infty} K(-t/h) F_{\varepsilon|x}(t) dt = \int_{-\infty}^{\infty} K(u) F_{\varepsilon|x}(-hu) du
= \tau + \int_{-\infty}^{\infty} K(u) \int_{0}^{hu} [f_{\varepsilon|x}(t) - f_{\varepsilon|x}(0)] dt du,
\]

from which it follows that \( |\mathbb{E}[\tilde{K}(-\varepsilon/h)|x] - \tau| \leq 0.5l_0\kappa_2 h^2 \). Consequently,

\[ ||\nabla Q_h(\beta^*)||_{\Omega} = \sup_{u \in \mathbb{R}^p} \mathbb{E}[\tilde{K}(-\varepsilon/h) - \tau](\Sigma^{-1/2}x, u) \leq 0.5l_0\kappa_2 h^2. \tag{B.5} \]

Putting together the pieces, we conclude that

\[ \langle -\nabla Q_h(\beta^*), \tilde{\beta}^* - \beta^* \rangle \leq 0.5l_0\kappa_2 h^2 \cdot ||\tilde{\beta}^* - \beta^*||_{\Sigma}. \tag{B.6} \]

Recall that \( f_{\varepsilon|x}(0) \geq \tilde{f} > 0 \) almost surely and \( ||\tilde{\beta}^* - \beta^*||_{\Sigma}^2 \leq \kappa_2 h^2 \). Combining this with (B.1), (B.4) and (B.6) yields

\[ \tilde{f} \cdot ||\tilde{\beta}^* - \beta^*||_{\Sigma}^2 \leq ||\tilde{\beta}^* - \beta^*||_{\Sigma}^2 \leq (0.5m_3 \kappa_2 + 0.5 \kappa_2 + \kappa_1 \kappa_2^{1/2})l_0 h^2 \cdot ||\tilde{\beta}^* - \beta^*||_{\Sigma}. \]

Canceling \( ||\tilde{\beta}^* - \beta^*||_{\Sigma} \) on both sides, we obtain

\[ ||\tilde{\beta}^* - \beta^*||_{\Sigma} \leq \tilde{L}^{-1} \left( (0.5m_3 \kappa_2^{1/2} + 0.5 \kappa_2^{1/2} + \kappa_1) l_0 h \cdot \kappa_2^{1/2} \right) = \tilde{L}^{-1} c_K l_0 h \cdot \kappa_2^{1/2} h. \]

Provided that \( h < \tilde{L}/(c_K l_0) \), \( \tilde{\beta}^* \) falls into the interior of \( \Theta^* \), i.e., \( ||\tilde{\beta}^* - \beta^*||_{\Sigma} < \kappa_2^{1/2} h \), thus enforcing \( n = 1 \) (otherwise, by construction \( \tilde{\beta}^* \) must lie on the boundary which leads to contradiction) and hence \( \tilde{\beta}^* = \beta_h^* \). In addition, by (B.4), \( Q_h(\cdot) \) is strictly convex in a neighborhood of \( \beta_h^* \) so that \( \beta_h^* \) is the unique minimizer and satisfies the stated bound (3.3).
Next, to investigate the leading term in the bias, define

$$\Delta = \Sigma^{-1/2} [\nabla Q_h(\beta_h^*) - \nabla Q_h(\beta^*) - D(\beta_h^* - \beta^*)]$$

and

$$D_0 = \Sigma^{-1/2} D \Sigma^{-1/2} = \mathbb{E} [f_{\hat{e}(x)}(0) w w^T]$$

where $w = \Sigma^{-1/2} x$. Again, by the mean value theorem for vector-valued functions,

$$\Delta = \left( \Sigma^{-1/2} \int_0^1 \nabla^2 Q_h((1 - t)\beta^* + t\beta_h^*) \, dt \right) \Sigma^{-1/2} D_0 \Sigma^{1/2} (\beta_h^* - \beta^*)$$

(B.7)

The Lipschitz continuity of $f_{\hat{e}(x)}(\cdot)$ ensures that

$$\left\| \Sigma^{-1/2} \int_0^1 \nabla^2 Q_h((1 - t)\beta^* + t\beta_h^*) \, dt \Sigma^{-1/2} - D_0 \right\|_2$$

$$\leq l_0 \sup_{u \in \mathcal{B}^{p-1}} \mathbb{E} \int_0^1 \int_{-\infty}^{\infty} K(u) |f_{\hat{e}(x)}(t(x, \beta_h^* - \beta^*) - hu) - f_{\hat{e}(x)}(0)| \, du \, dt \, w w^T$$

$$\leq 0.5l_0 \sup_{u \in \mathcal{B}^{p-1}} \mathbb{E} |(x, \beta_h^* - \beta^*)(w, u)^2| + l_0 \kappa_1 h$$

$$\leq 0.5l_0 m_3 \|\beta_h^* - \beta^*\|_\Sigma + l_0 \kappa_1 h.$$ 

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

$$\|\Delta\|_2 \leq l_0 (0.5m_3 \|\beta_h^* - \beta^*\|_\Sigma + \kappa_1 h) \|\beta_h^* - \beta^*\|_\Sigma.$$ 

(B.8)

Moreover, applying a second-order Taylor series expansion to $f_{\hat{e}(x)}(\cdot)$ yields

$$\mathbb{E}[\bar{K}(-\varepsilon/h)|x] - \tau$$

$$= \int_{-\infty}^{\infty} K(u) \int_0^{-hu} \{f_{\hat{e}(x)}(t) - f_{\hat{e}(x)}(0)\} \, dt \, du$$

$$= 0.5\kappa_2 h^2 \cdot f_{\hat{e}(x)}''(0) + \int_{-\infty}^{\infty} \int_0^{-hu} K(u) [f_{\hat{e}(x)}''(v) - f_{\hat{e}(x)}''(0)] \, dv \, dt \, du.$$

For $\nabla Q_h(\beta^*) = \mathbb{E}[\bar{K}(-\varepsilon/h) - \tau]x$, it follows that

$$\left\| \nabla Q_h(\beta^*) - \frac{1}{2} \kappa_2 h^2 \cdot \mathbb{E}[f_{\hat{e}(x)}''(0)] x \right\|_\Omega \leq \frac{1}{6} l_1 \kappa_3 h^3.$$ 

(B.9)

Combining (B.8) and (B.9) completes the proof of (3.4).
B.2 Proof of Theorem 3.1

Recall from (2.6) that $\hat{\beta}_h$ is the smoothed quantile regression estimator obtained by minimizing $\hat{Q}_h(\cdot)$. For any given $\tau \in (0, 1)$ and $h > 0$, by the optimality of $\hat{\beta}_h$ and convexity of $\hat{Q}_h(\cdot)$, we have $\nabla \hat{Q}_h(\hat{\beta}_h) = 0$ and $(\nabla \hat{Q}_h(\hat{\beta}_h) - \nabla \hat{Q}_h(\beta^*), \hat{\beta}_h - \beta^*) \geq 0$.

As in the proof of Proposition 3.1, we write $\Theta(t) = \{u \in \mathbb{R}^p : \|u\|_\Sigma \leq t\}$ for $t \geq 0$. For some $r > 0$ to be determined, define $\beta = \beta^* + \eta(\hat{\beta}_h - \beta^*)$ where $\eta := \sup\{u \in [0, 1] : u(\hat{\beta}_h - \beta^*) \in \Theta(r)\}$. Thus, by definition, $\eta = 1$ if $\hat{\beta}_h \in \beta^* + \Theta(r)$, and $\eta < 1$ if $\hat{\beta}_h \notin \beta^* + \Theta(r)$. In the latter case, we must have $\hat{\beta} \in \beta^* + \partial \Theta(\tau)$, where $\partial \Theta(\tau)$ denotes the boundary of $\Theta(\tau)$. Consider the symmetrized Bregman divergence associated with $\hat{Q}_h(\cdot)$:

$$D(\beta_1, \beta_2) = \langle \nabla \hat{Q}_h(\beta_1) - \nabla \hat{Q}_h(\beta_2), \beta_1 - \beta_2 \rangle, \quad \beta_1, \beta_2 \in \mathbb{R}^p.$$  \hspace{1cm} (B.10)

By Lemma C.1 in the supplement of Sun, Zhou and Fan (2020), the three points $\hat{\beta}_h$, $\hat{\beta}$, and $\beta^*$ satisfy $D(\hat{\beta}, \beta^*) \leq \eta D(\hat{\beta}, \beta^*)$. Combined with the facts $\hat{\beta} - \beta^* = \eta(\hat{\beta}_h - \beta^*) \in \Theta(r)$ and $\nabla \hat{Q}_h(\hat{\beta}_h) = 0$, we find that

$$\|\hat{\beta} - \beta^*\|_\Sigma^2 \cdot \frac{D(\hat{\beta}, \beta^*)}{\|\beta - \beta^*\|_\Sigma^2} \leq -\eta \langle \nabla \hat{Q}_h(\beta^*), \hat{\beta}_h - \beta^* \rangle \leq \|\nabla \hat{Q}_h(\beta^*)\|_\Omega \cdot \|\hat{\beta} - \beta^*\|_\Sigma.$$

Canceling $\|\hat{\beta} - \beta^*\|_\Sigma$ on both sides yields

$$\|\hat{\beta} - \beta^*\|_\Sigma \leq \frac{\|\nabla \hat{Q}_h(\beta^*)\|_\Omega}{\inf_{\beta \in \beta^* + \Theta(r)} D(\beta, \beta^*)/\|\beta - \beta^*\|_\Sigma^2}. \hspace{1cm} (B.11)$$

The following two lemmas provide, respectively, upper and lower bounds on $\|\nabla \hat{Q}_h(\beta^*)\|_\Omega$ and $\inf_{\beta \in \beta^* + \Theta(r)} D(\beta, \beta^*)/\|\beta - \beta^*\|_\Sigma^2$.

**Lemma B.1.** Assume that Conditions 3.1–3.4 hold. For any $t \geq 0$,

$$\|\nabla \hat{Q}_h(\beta^*)\|_\Omega \leq 1.46 \nu_0 \left\{ \nu_t \sqrt{\frac{4p + 2t}{n} + 2 \max(1 - \tau, \tau) \frac{2p + t}{n}} + 0.5 l_0 \kappa_2 h^2 \right\}$$

with probability at least $1 - e^{-t}$, where $\nu_t = \tau(1 - \tau) + (1 + \tau) l_0 \kappa_2 h^2$.

Under Condition 3.2, there exist constants $\bar{f}_h \geq \underline{f}_h$ such that

$$\underline{f}_h \leq \inf_{\|u\| \leq h/2} f_{\varepsilon|x}(u) \leq \sup_{\|u\| \leq h/2} f_{\varepsilon|x}(u) \leq \bar{f}_h \hspace{1cm} (B.13)$$
almost surely (over \( x \)). By the Lipschitz continuity, we can take \( \tilde{f}_h = \tilde{f} + l_0 h/2 \) and \( \mathcal{L}_h = \mathcal{L} - l_0 h/2 \).

Throughout the following, we assume \((B.13)\) holds. For every \( \delta \in (0, 1] \), we define \( \eta_\delta \geq 0 \) as

\[
\eta_\delta = \inf\{ \eta > 0 : \mathbb{E}\{ (u,w)^2 I(|(u,w)> \eta_\delta) \} \leq \delta \ \text{for all} \ u \in \mathbb{S}^{p-1} \},
\]

where \( w = \Sigma^{-1/2} x \) is the standardized covariate vector that satisfies \( \mathbb{E}(ww^T) = \mathbf{I}_p \), and hence \( \mathbb{E}(u,w)^2 = 1 \) for any \( u \in \mathbb{S}^{p-1} \). It can be shown that \( \eta_\delta \) depends only on \( \delta \) and \( \nu_0 \) in Condition 3.4, and the map \( \delta \mapsto \eta_\delta \) is non-increasing with \( \eta_\delta \downarrow 0 \) as \( \delta \uparrow 1 \).

**Lemma B.2.** Let \( 0 < h < 2\lfloor \sqrt{l_0} \rfloor \) and \( 0 < r \leq h/(4\eta_{1/4}) \) with \( \eta_{1/4} \) defined in \((B.14)\). Then, for any \( t > 0 \),

\[
\inf_{\beta \in \partial^{\Theta(r)} \kappa/\|\beta - \beta^*\|_{\Sigma}} D(\beta, \beta^*) \geq \frac{3}{4} 4L_h - \tilde{f}_h^{1/2} \left( \frac{5}{4} \sqrt{\frac{hp}{r^2 n}} + \sqrt{\frac{ht}{8r^2 n}} \right) - \frac{ht}{3r n} \quad (B.15)
\]

with probability at least \( 1 - e^{-t} \).

In view of \((B.11)\), \((B.12)\), and \((B.15)\), we take \( r = h/(4\eta_{1/4}) \) so that as long as \( (p + t)/n \leq h \leq 1 \),

\[
\| \tilde{\beta} - \beta^*\|_{\Sigma} \leq 3(\kappa L)^{-1} v_0 \left( C_{1/2}^{1/2} \sqrt{\frac{4p + 2t}{n}} + 2 \max(1 - \tau, \tau) \frac{2p + t}{n} \right) + (\kappa L)^{-1} l_0 \kappa_2 h^2 \quad (B.16)
\]

with probability at least \( 1 - 2e^{-t} \). With this choice of \( r \), we see that under the constraint \( \sqrt{(p + t)/n} \leq h \leq 1 \), the right-hand side of \((B.16)\) is strictly less than \( r \). In other words, conditioned on an event that occurs with high probability, \( \tilde{\beta} \) falls into the interior of \( \beta^* + \Theta(r) \), thus enforcing \( \eta = 1 \) and \( \tilde{\beta}_h = \beta \). The claimed bound for \( \tilde{\beta}_h \) then follows immediately.

**B.2.1 Proof of Lemma B.1**

Write \( Q_h(\beta) = \mathbb{E}(\tilde{Q}_h(\beta)) \), and define \( \xi_i = \tilde{R}(-\xi_i/h) - \tau \) for \( i = 1, \ldots, n \). By the triangle inequality and \((B.5)\), we have

\[
\| \nabla \tilde{Q}_h(\beta^*) \|_\Omega \leq \| \nabla \tilde{Q}_h(\beta^*) - \nabla Q_h(\beta^*) \|_\Omega + \| \nabla Q_h(\beta^*) \|_\Omega \leq \| \Sigma^{-1/2} \nabla \tilde{Q}_h(\beta^*) - \nabla Q_h(\beta^*) \|_2 + 0.5 l_0 \kappa_2 h^2.
\]

It suffices to obtain an upper bound for the centered score \( \Sigma^{-1/2} \{ \nabla \tilde{Q}_h(\beta^*) - \nabla Q_h(\beta^*) \} = (1/n) \sum_{i=1}^n \{ \xi_i w_i - \mathbb{E}(\xi_i w_i) \} \in \mathbb{R}^p \), where \( w_i = \Sigma^{-1/2} x_i \).

Using a covering argument, for any \( \epsilon \in (0, 1) \), there exists an \( \epsilon \)-net \( \mathcal{N}_\epsilon \) of the unit sphere with
cardinality $|\mathcal{N}| \leq (1 + 2/e)^p$ such that

$$||\Sigma^{-1/2}[\nabla \hat{Q}_h(\beta^*) - \nabla Q_h(\beta^*)]||_2 \leq (1 - \epsilon)^{-1} \max_{u \in \mathcal{N}} \langle u, \Sigma^{-1/2}[\nabla \hat{Q}_h(\beta^*) - \nabla Q_h(\beta^*)] \rangle.$$

For each unit vector $u \in \mathcal{N}$, define centered random variables $\gamma_{u,i} = \langle u, \xi w_i - \mathbb{E}(\xi_i w_i) \rangle$. It is easy to see that $|\epsilon| \leq \bar{\tau} := \max(1 - \tau, \tau)$. To bound $\mathbb{E}(\xi_i^2 | x_i)$, by a change of variable and integration by parts,

$$\mathbb{E}(\bar{K}^2(-v/h)|x) = 2 \int_{-\infty}^{\infty} K(v) \bar{K}(v) F_{el}(-vh) dv$$

$$= 2\tau \int_{-\infty}^{\infty} K(v) \bar{K}(v) dv - 2h f_{el}(0) \int_{-\infty}^{\infty} v K(v) \bar{K}(v) dv$$

$$+ 2 \int_{-\infty}^{\infty} \int_{0}^{-vh} \{f_{el}(t) - f_{el}(0)\} K(v) \bar{K}(v) dv dt$$

$$\leq \tau + l_0 \kappa_2 h^2,$$

where $\kappa_2$ and $l_0$ are the constants from Conditions 3.1 and 3.2. It then follows that $\mathbb{E}(\xi_i^2 | x_i) \leq \tau^2 := \tau(1 - \tau) + (1 + \tau) l_0 \kappa_2 h^2$. Hence, for $k = 2, 3, \ldots$,

$$\mathbb{E}(|\langle u, \xi_i w_i \rangle|^k) \leq \tau^{k-2} \mathbb{E}(|\langle u, w_i \rangle|^k \cdot \mathbb{E}(\xi_i^2 | x_i)|)$$

$$\leq \tau^{k-2} \nu_0 \int_{0}^{\infty} P(|\langle u, w_i \rangle| \geq \nu_0 t) k t^{k-1} dt$$

$$\leq \tau^{k-2} \nu_0 \int_{0}^{\infty} t^{-1} e^{-t} dt$$

$$= k! \cdot \tau^{k-2} \cdot \nu_0^k$$

$$\leq k! \cdot \tau^{k-2} \cdot (2\tau \nu_0)^{k-2}.$$

Consequently, it follows from Bernstein’s inequality that for every $u \geq 0$,

$$\frac{1}{n} \sum_{i=1}^{n} \gamma_{u,i} \leq \nu_0 \left( \nu \sqrt{\frac{2u}{n}} + \frac{2\tau u}{n} \right)$$

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

Finally, applying a union bound over $u \in \mathcal{N}$ yields

$$||\nabla \hat{Q}_h(\beta^*) - \nabla Q_h(\beta^*)||_\Omega \leq \frac{\nu_0}{1 - \epsilon} \left( \nu \sqrt{\frac{2u}{n}} + \frac{2\tau u}{n} \right)$$

8
with probability at least \(1 - e^{\log(1+2\epsilon)p^u}\). Taking \(\epsilon = 2/(e^2 - 1)\) and \(u = 2p + t (t \geq 0)\) proves the claimed result.

\(\square\)

### B.2.2 Proof of Lemma B.2

Recall that the empirical loss \(\widehat{Q}_h(\cdot)\) in (2.5) is convex and twice continuously differentiable with \(\nabla \widehat{Q}_h(\beta) = (1/n) \sum_{i=1}^n [K((\langle x_i, \beta \rangle - y_i)/h) - \tau|x_i|\) and \(\nabla^2 \widehat{Q}_h(\beta) = (1/n) \sum_{i=1}^n K_h((\langle x_i, \beta \rangle - y_i)x_i x_i^\top).

For the symmetrized Bregman divergence \(D : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^+\) defined in (B.10), we have

\[
D(\beta, \beta^*) = \frac{1}{n} \sum_{i=1}^n \left\{ \bar{K} \left( \frac{\langle x_i, \beta \rangle - y_i}{h} \right) - \bar{K} \left( \frac{-\varepsilon_i}{h} \right) \right\} \langle x_i, \beta - \beta^* \rangle.  \tag{B.17}
\]

Define the events \(E_i = \{|e_i| \leq h/2\} \cap \{|y_i - \langle x_i, \beta \rangle| \leq \|\beta - \beta^*\|_\Sigma \cdot h/(2r)\}\) for \(i = 1, \ldots, n\). For any \(\beta \in \beta^* + \Theta(r)\), note that \(|y_i - \langle x_i, \beta \rangle| \leq h\) on \(E_i\), implying

\[
D(\beta, \beta^*) \geq \frac{k_i}{nh} \sum_{i=1}^n \langle x_i, \beta - \beta^* \rangle^2 \mathbb{1}_{E_i}, \tag{B.18}
\]

where \(\mathbb{1}_{E_i}\) is the indicator function of \(E_i\) and \(k_i = \min_{|u| \leq 1} K(u)\). It then suffices to bound the right-hand side of the above inequality from below uniformly over \(\beta \in \beta^* + \Theta(r)\).

For \(R > 0\), define the function \(\varphi_R(u) = u^2 \mathbb{1}(|u| \leq R/2) + [u \text{ sign}(u) - R]^2 \mathbb{1}(R/2 < |u| \leq R)\), which is \(R\)-Lipschitz continuous and satisfies

\[
u^2 \mathbb{1}(|u| \leq R/2) \leq \varphi_R(u) \leq u^2 \mathbb{1}(|u| \leq R). \tag{B.19}
\]

Moreover, note that \(\varphi_{R}(cu) = c^2 \varphi_R(u)\) for any \(c > 0\) and \(\varphi_0(u) = 0\). For \(\beta \in \beta^* + \Theta(r)\), consider a change of variable \(\delta = \Sigma^{1/2} (\beta - \beta^*)/\|\beta - \beta^*\|_\Sigma \in \mathbb{S}^{p-1}\). Together, (B.18) and (B.19) imply

\[
\frac{D(\beta, \beta^*)}{\|\beta - \beta^*\|_\Sigma^2} \geq \frac{k_i}{nh} \sum_{i=1}^n \omega_i \cdot \varphi_{h/(2r)}(\langle w_i, \delta \rangle), \tag{B.20}
\]

where \(\omega_i := \mathbb{1}(|e_i| \leq h/2)\) with \(e_i = y_i - \langle x_i, \beta^* \rangle\), and \(w_i = \Sigma^{-1/2}x_i\).

Next, we bound the expectation \(\mathbb{E}[D_0(\delta)]\) and the random fluctuation \(D_0(\delta) - \mathbb{E}[D_0(\delta)]\), separately, starting with the former. By (B.13),

\[
f_{\delta}h \leq \mathbb{E}(\omega_i|x_i|) = \int_{-h/2}^{h/2} f_{\delta|x_i}(u) \, du \leq \bar{f}_h h. \tag{B.21}
\]
Moreover, define $\xi_\delta = \langle w, \delta \rangle$ such that $\mathbb{E}(\xi_\delta^2) = 1$. By (B.19) and (B.21),

$$\mathbb{E}[\omega_i \cdot \varphi_{h/2r}(\langle w_i, \delta \rangle)] \geq \bar{f}_h \cdot \mathbb{E}[\varphi_{h/2r}(\langle w_i, \delta \rangle)] \geq \bar{f}_h \cdot \left[ 1 - \mathbb{E}[\xi_\delta^2 \mid |\xi_\delta| > h/(4r)] \right].$$

For $0 < r \leq h/(4\eta_{1/4})$ with $\eta_{1/4}$ defined in (B.14), it follows that

$$\inf_{\delta \in \mathbb{S}^{p-1}} \mathbb{E}[D_0(\delta)] \geq \bar{f}_h \cdot \left[ 1 - \sup_{u \in \mathbb{S}^{p-1}} \mathbb{E}(w, u)^2 \mathbb{1}(|\langle w, u \rangle| \geq \eta_{1/4}) \right] \geq \frac{3}{4} \bar{f}_h. \quad (B.22)$$

Turning to the random fluctuation, we will use Theorem 7.3 in Bousquet (2003) (a refined Talagrand’s inequality) to bound

$$\Delta = \sup_{\delta \in \mathbb{S}^{p-1}} [D_0(\delta) - \mathbb{E}[D_0(\delta)]], \quad (B.23)$$

where $D_0(\delta) := -D_0(\delta)$. Note that $0 \leq \varphi_R(u) \leq \min((R/2)^2, (R/2)|u|)$ for all $u \in \mathbb{R}$ and $\omega_i \in [0, 1]$. Therefore,

$$0 \leq \chi_i := \langle \omega_i / h \cdot \varphi_{h/2r}(\langle w_i, \delta \rangle) \rangle \leq \omega_i \cdot \min\{h/(4r)^2, |\langle w_i, \delta \rangle|/(4r)\}.$$

This, combined with (B.21), yields $\mathbb{E}(\chi_i^2) \leq \bar{f}_h h/(4r)^2$. We then apply Theorem 7.3 in Bousquet (2003) and obtain that, for any $t > 0$,

$$\Delta \leq \mathbb{E}(\Delta) + (\mathbb{E}(\Delta))^{1/2} \sqrt{\frac{ht}{4r^2n}} + \sqrt{\frac{\bar{f}_h h}{4r^2n}} \leq \frac{\varepsilon}{4} \mathbb{E}(\Delta) + \sqrt{\frac{\bar{f}_h h}{4r^2n}} \leq \frac{\bar{f}_h h}{3r^2n} \quad (B.24)$$

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

It remains to bound the expected value $\mathbb{E}(\Delta)$. Recall that $\omega_i = \mathbb{1}\{\xi_i \leq h/2\} \in [0, 1]$, and hence

$$\omega_i \varphi_{h/2r}(\langle w_i, \delta \rangle) = \omega_i^2 \varphi_{h/2r}(\langle w_i, \delta \rangle) = \omega_i \varphi_{h/(2r)}(\langle \omega_i w_i, \delta \rangle) = \varphi_{h/(2r)}(\langle \omega_i w_i, \delta \rangle).$$

Then, we define

$$z_i = \omega_i w_i \quad \text{and} \quad \mathcal{E}(\delta; z_i) = \varphi_{h/(2r)}(\langle z_i, \delta \rangle), \quad \delta \in \mathbb{S}^{p-1}.$$

By Rademacher symmetrization,

$$\mathbb{E}(\Delta) \leq 2\mathbb{E}\left\{ \sup_{\delta \in \mathbb{S}^{p-1}} \frac{1}{nh} \sum_{i=1}^{n} e_i \cdot \mathcal{E}(\delta; z_i) \right\},$$

where $e_1, \ldots, e_n$ are independent Rademacher random variables. Since $\varphi_R(\cdot)$ is $R$-Lipschitz, $\mathcal{E}(\delta; z_i)$
is an \((h/2r)\)-Lipschitz function in \(\langle z_i, \delta \rangle\), i.e., for any sample \(z_i\) and parameters \(\delta, \delta' \in S^{p-1}\),

\[
|\mathcal{E}(\delta; z_i) - \mathcal{E}(\delta'; z_i)| \leq \frac{h}{2r} |\langle z_i, \delta \rangle - \langle z_i, \delta' \rangle|.
\]

Moreover, observe that \(\mathcal{E}(\delta; z_i) = 0\) for any \(\delta\) such that \(\langle z_i, \delta \rangle = 0\). With the above preparations, we are ready to use Talagrand’s contraction principle to bound \(\mathbb{E}(\Delta)\). Define the subset \(T \subseteq \mathbb{R}^n\) as

\[
T = \{t = (t_1, \ldots, t_n)^\top : t_i = \langle z_i, \delta \rangle, i = 1, \ldots, n, \delta \in S^{p-1}\},
\]

and contractions \(\phi_i : \mathbb{R} \to \mathbb{R}\) as \(\phi_i(t) = (2r/h) \cdot \varphi_{h/(2r)}(t)\). By (B.25), \(|\phi_i(t) - \phi_i(s)| \leq |t - s|\) 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

\[
\mathbb{E}(\Delta) \leq \frac{1}{r} \mathbb{E}\left\{\sup_{\delta \in S^{p-1}} \frac{1}{nh} \sum_{i=1}^{n} e_i \cdot \mathcal{E}(\delta; z_i)\right\} = \frac{1}{r} \mathbb{E}\left\{\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} e_i \phi_i(t)\right\} \leq \frac{1}{r} \mathbb{E}\left\{\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} e_i t_i\right\} \leq \bar{f}_h^{1/2} \sqrt{\frac{hp}{r^2n}}.
\]

This, combined with (B.23) and (B.24), yields

\[
\Delta \leq \bar{f}_h^{1/2} \left(\frac{5}{4} \sqrt{\frac{hp}{r^2n}} + \sqrt{\frac{ht}{8r^2n}}\right) + \frac{ht}{3r^2n} \quad \text{(B.26)}
\]

with probability at least \(1 - e^{-t}\).

Finally, combining (B.17), (B.20), (B.22) and (B.26) proves (B.15).

\(\square\)

### B.3 Proof of Theorem 3.2

We keep the notation used in the proof of Theorem 3.1, and for any \(t \geq 0\), let \(r = r(n, p, t) = \sqrt{(p + t)/n + h^2} > 0\) be such that \(\mathbb{P}\{\hat{\beta}_h \in \beta^* + \Theta(r)\} \geq 1 - 2e^{-t}\), provided \((p + t)/n \leq h \leq 1\). Define the vector-valued random process

\[
\Delta(\beta) = \mathbf{\Sigma}^{-1/2} [\nabla \hat{\mathcal{Q}}_h(\beta) - \nabla \mathcal{Q}_h(\beta^*) - \mathbf{D}(\beta - \beta^*)],
\]

(B.27)
where \( D = \mathbb{E}[f_{\text{ilex}}(0)xx^T] \). Since \( \hat{\beta}_h \) falls in a local neighborhood of \( \beta^* \) with high probability, it suffices to bound the local fluctuation \( \sup_{\beta \in \beta^* + \Theta(h)} \|\Delta(\beta)\|_2 \). By the triangle inequality,

\[
\sup_{\beta \in \beta^* + \Theta(h)} \|\Delta(\beta)\|_2 \leq \sup_{\beta \in \beta^* + \Theta(h)} \|\mathbb{E}\Delta(\beta)\|_2 + \sup_{\beta \in \beta^* + \Theta(h)} \|\Delta(\beta) - \mathbb{E}\Delta(\beta)\|_2 := I_1 + I_2. \tag{B.28}
\]

We now provide upper bounds for \( I_1 \) and \( I_2 \), respectively.

**Upper bound for \( I_1 \):** By the mean value theorem for vector-valued functions,

\[
\mathbb{E}\Delta(\beta) = \Sigma^{-1/2} \left( \int_0^1 \nabla^2 Q_h((1-t)\beta^* + t\beta) dt \right) - \Sigma^{-1/2} D(\beta - \beta^*)
\]

\[
= \left( \Sigma^{-1/2} \int_0^1 \nabla^2 Q_h((1-t)\beta^* + t\beta) dt \right) \Sigma^{-1/2} - D_0, \Sigma^{1/2}(\beta - \beta^*)
\]

where \( D_0 := \Sigma^{-1/2} D \Sigma^{-1/2} = \mathbb{E}[f_{\text{ilex}}(0)ww^T] \). By law of iterative expectation and by a change of variable,

\[
\Sigma^{-1/2} \nabla^2 Q_h(\beta) \Sigma^{-1/2} = \mathbb{E}[K_h(x, \beta - y)ww^T] = \mathbb{E}\left\{ \int_{-\infty}^\infty K(u)f_{\text{ilex}}(\langle x, \beta - y \rangle - hu) du \cdot ww^T \right\}
\]

For notational convenience, let \( v = \Sigma^{1/2}(\beta - \beta^*) \) with \( \beta \in \beta^* + \Theta(h) \), so that \( \|v\|_2 \leq r \) and

\[
\nabla^2 Q_h((1-t)\beta^* + t\beta) = \mathbb{E}\left\{ \int_{-\infty}^\infty K(u)f_{\text{ilex}}(t\langle w, v \rangle - hu) du \cdot ww^T \right\}
\]

By the Lipschitz continuity of \( f_{\text{ilex}}(\cdot) \), i.e. \( |f_{\text{ilex}}(u) - f_{\text{ilex}}(0)| \leq l_0|u| \) for all \( u \in \mathbb{R} \) almost surely for \( x \), we have

\[
\left\| \Sigma^{-1/2} \nabla^2 Q_h((1-t)\beta^* + t\beta) \Sigma^{-1/2} - D_0 \right\|_2 \leq l_0 t \sup_{\|u\|_2 = 1} \mathbb{E}((u, u)^2 |(w, v)|) + l_0 \kappa_1 h \sup_{\|u\|_2 = 1} \mathbb{E}(u, u)^2
\]

\[
\leq l_0 t \left( \sup_{\|u\|_2 = 1} \mathbb{E}(u, u)^3 \right)^{2/3} \left( \mathbb{E}(w, v)^3 \right)^{1/3} + l_0 \kappa_1 h \leq l_0 (m_3 r + \kappa_1 h),
\]

where the third inequality holds by the Cauchy-Schwarz inequality. Consequently,

\[
\sup_{\beta \in \beta^* + \Theta(h)} \|\mathbb{E}\Delta(\beta)\|_2 \leq l_0(0.5m_3 r + \kappa_1 h) \cdot r. \tag{B.29}
\]
Upper bound for $I_2$: Next, we provide an upper bound for $\Delta(\beta) - \mathbb{E}\Delta(\beta)$. Define the centered gradient process $G(\beta) = \Sigma^{-1/2} [\nabla \hat{Q}_h(\beta) - \nabla \tilde{Q}_h(\beta)]$, so that $\Delta(\beta) - \mathbb{E}\Delta(\beta) = G(\beta) - G(\beta^*)$. Again, by a change of variable $v = \Sigma^{1/2}(\beta - \beta^*)$, we have

$$
\sup_{\beta \in \beta^* + \Theta(r)} \|\Delta(\beta) - \mathbb{E}\Delta(\beta)\|_2 \leq \sup_{\beta \in \beta^* + \Theta(r)} \|G(\beta) - G(\beta^*)\|_2
$$

$$
= \sup_{\|v\|_2 \leq \Delta_0(v)} \| G(\beta^* + \Sigma^{-1/2} v) - G(\beta^*) \|_2.
$$

We will employ Theorem A.3 in Spokoiny (2013) to bound the supremum $\sup_{\|v\|_2 \leq \Delta_0(v)} \|\Delta(\beta)\|_2$, where $\Delta_0(\cdot)$ defined above satisfies $\Delta_0(0) = 0$ and $\mathbb{E}\Delta_0(v) = 0$. Taking the gradient with respect to $v$ yields

$$
\nabla \Delta_0(v) = \frac{1}{n} \sum_{i=1}^{n} \left( K_{i,v} w_i w_i^T - \mathbb{E}(K_{i,v} w_i w_i^T) \right),
$$

where $K_{i,v} := K_h(\langle w_i, v \rangle - \epsilon_i)$ satisfies $0 \leq K_{i,v} \leq \kappa_a h^{-1}$. For any $u, v \in \mathbb{S}^{p-1}$ and $\lambda \in \mathbb{R}$, using the elementary inequality $|e^u - 1 - u| \leq u^2 e|u|/2$, we obtain

$$
\mathbb{E}\exp\left[ \lambda h^{1/2}(u, \nabla \Delta_0(v) v) \right]
\leq \prod_{i=1}^{n} \left\{ 1 + \frac{\lambda^2}{2n} e^{\lambda h^{1/2}(w_i, u) \langle w_i, v \rangle} \mathbb{E}[K_{i,v}(w_i, u)(w_i, v) - \mathbb{E}(K_{i,v}(w_i, u)(w_i, v)) e^{\lambda h^{1/2}(w_i, u) \langle w_i, v \rangle} ] \right\}
\leq \prod_{i=1}^{n} \left\{ 1 + \frac{\lambda^2}{2n} e^{\lambda h^{1/2}(w_i, u) \langle w_i, v \rangle} \mathbb{E}[K_{i,v}(w_i, u)(w_i, v) - \mathbb{E}(K_{i,v}(w_i, u)(w_i, v)) e^{\lambda h^{1/2}(w_i, u) \langle w_i, v \rangle} ] \right\},
$$

(B.30)

where we use the bound $\mathbb{E}(w_i, u) \langle w_i, v \rangle \leq (\mathbb{E}(w_i, u)^2)^{1/2}(\mathbb{E}(w_i, v)^2)^{1/2} = 1$ in the second inequality. Moreover, the first and second conditional moments of $K_{i,v}$ can be rewritten as follows:

$$
\mathbb{E}(K_{i,v}|x_i) = \frac{1}{h} \int_{-\infty}^{\infty} K\left( \frac{\langle w_i, v \rangle - t}{h} \right) f_{\epsilon_i(\tau)|x_i}(t) \, dt = \int_{-\infty}^{\infty} K(u) f_{\epsilon_i(\tau)|x_i}(\langle w_i, v \rangle - hu) \, du;
$$

$$
\mathbb{E}(K^2_{i,v}|x_i) = \frac{1}{h^2} \int_{-\infty}^{\infty} K^2\left( \frac{\langle w_i, v \rangle - t}{h} \right) f_{\epsilon_i(\tau)|x_i}(t) \, dt = \frac{1}{h} \int_{-\infty}^{\infty} K^2(u) f_{\epsilon_i(\tau)|x_i}(\langle w_i, v \rangle - hu) \, du,
$$

from which it follows that $\mathbb{E}(K_{i,v}|x_i) \leq \bar{f}$ and $\mathbb{E}(K^2_{i,v}|x_i) \leq \kappa_a \bar{f} h^{-1}$ almost surely.
By the Cauchy-Schwarz inequality and the inequality \( ab \leq a^2/2 + b^2/2, a, b \in \mathbb{R} \), we have

\[
\mathbb{E}(\langle w_i, u \rangle \langle w_i, v \rangle)^2 e^{\langle w_i, u \rangle \langle w_i, v \rangle} \\
\leq \mathbb{E}((\langle w_i, u \rangle \langle w_i, v \rangle)^2 e^{\langle w_i, u \rangle^2 + \frac{1}{4} \langle w_i, v \rangle^2} \\
\leq \left( \mathbb{E}(\langle w_i, u \rangle^4) e^{\langle w_i, u \rangle^2} \right)^{1/2} \left( \mathbb{E}(\langle w_i, v \rangle^4) e^{\langle w_i, v \rangle^2} \right)^{1/2}, \quad \text{valid for any } t > 0.
\]

Given a unit vector \( u \), let \( \chi = \langle w, u \rangle^2/(2\nu_1)^2 \) so that under Condition 3.5, \( \mathbb{P}(\chi \geq u) \leq 2e^{-2u} \) for any \( u \geq 0 \). It follows that \( \mathbb{E}(e^{\chi}) = 1 + \int_0^\infty e^{2u} \mathbb{P}(\chi \geq u) du \leq 1 + 2 \int_0^\infty e^{-u} du = 3 \), and

\[
\mathbb{E}(\chi^2 e^{\chi}) = \int_0^\infty (u^2 + 2u)e^{2u} \mathbb{P}(\chi \geq u) du \leq 2 \int_0^\infty (u^2 + 2u)e^{-u} du = 8.
\]

Taking the supremum over \( u \in \mathbb{S}^{p-1} \), we have

\[
\sup_{u \in \mathbb{S}^{p-1}} \mathbb{E}(e^{\langle w, u \rangle^2/(2\nu_1)^2}) \leq 3 \quad \text{and} \quad \sup_{u \in \mathbb{S}^{p-1}} \mathbb{E}(\langle w, u \rangle^4 e^{\langle w, u \rangle^2/(2\nu_1)^2}) \leq 8(2\nu_1)^4.
\]

Substituting the above bounds into (B.30) yields that, for any \( |\lambda| \leq \min\{n^{1/2}h/(4\kappa_0\nu_1^2), n^{1/2}/\bar{f}\} \),

\[
\mathbb{E} \exp\left[n^{1/2} \langle u, \nabla \Delta_0(v) \rangle v \right] \\
\leq n \left[ 1 + \frac{e\lambda^2}{2n} \mathbb{E}(K_{t,v}(w_i, u)\langle w_i, v \rangle - \mathbb{E}(K_{t,v}(w_i, u)\langle w_i, v \rangle))^2 e^{\langle w_i, u \rangle \langle w_i, v \rangle}/(4\nu_1^2) \right] \\
\leq n \left[ 1 + \frac{e\lambda^2}{n} \mathbb{E}(K_{t,v}(w_i, u)\langle w_i, v \rangle)^2 e^{\langle w_i, u \rangle \langle w_i, v \rangle}/(4\nu_1^2) \\
+ \frac{e\lambda^2}{n} \mathbb{E}(\langle w_i, u \rangle \langle w_i, v \rangle)^2 e^{\langle w_i, u \rangle \langle w_i, v \rangle}/(4\nu_1^2) \right] \\
\leq n \left[ 1 + C_0^2 \frac{\lambda^2}{2nh} \right] \leq \exp(C_0^2 \lambda^2/(2h)),
\]

where \( C_0 > 0 \) depends only on \( (\nu_1, \kappa_0, \bar{f}) \). We have thus verified condition (A.4) in Spokoiny (2013) with \( g = \min\{h/(4\kappa_0\nu_1^2), 1/\bar{f}\}(n/2)^{1/2} \) and \( v_0 = C_0 h^{-1/2} \). Applying Theorem A.3 therein, we obtain that with probability at least \( 1 - e^{-t} \),

\[
\sup_{\|v\|_2 \leq r} \|\Delta_0(v)\|_2 \leq 6C_0 r \sqrt{\frac{4p + 2r}{nh}}
\]

as long as \( h \geq 8\kappa_0\nu_1^2 \sqrt{(2p + t)/n} \) and \( n \geq 4\bar{f}^2(2p + t) \).
Together with (B.28) and (B.29), this implies that with probability at least \( 1 - e^{-t} \),

\[
\sup_{\beta \in \beta^* + \Theta(r)} \| \Delta(\beta) \|_2 \leq 6C_0 r \sqrt{\frac{4p + 2t}{nh}} + l_0(0.5m_3r + \kappa_1h)r. \tag{B.31}
\]

Recall from the beginning of the proof that \( \widehat{\beta}_h \in \beta^* + \Theta(r) \) with probability at least \( 1 - 2e^{-t} \) with \( r = r(n, p, t) \approx \sqrt{(p + t)/n} + h^2 \). Combined with (B.31), we conclude that with probability at least \( 1 - 3e^{-t} \), \( \| \Delta(\widehat{\beta}_h) \|_2 \leq (p + t)/(h^{1/2}n) + h \sqrt{(p + t)/n} + h^3 \), as claimed. \( \square \)

### B.4 Proof of Theorem 3.3

Let \( \alpha \in \mathbb{R}^p \) be an arbitrary vector defining a linear functional of interest. Given \( h = h_n > 0 \), define \( S_n = n^{-1/2} \sum_{i=1}^n \gamma_i \xi_i \) and its centered version \( S_n^0 = S_n - \mathbb{E}(S_n) \), where \( \xi_i = \tau - \bar{K}(-\epsilon_i/h) \) and \( \gamma_i = \langle D^{-1} \alpha, x_i \rangle \). By the Lipschitz continuity of \( f_{\xi|x} \) around 0 and the fundamental theorem of calculus, it can be shown that \( \mathbb{E}(\gamma_i) \leq 0.5l_0\kappa_2 h^2 \), from which it follows by the law of iterated expectation that \( \mathbb{E}(\gamma_i|x_i) \leq 0.5l_0\kappa_2 \| D^{-1} \alpha \|_\Sigma \cdot h^2 \).

Let \( \delta_n = (p + \log n)/n \). Then, applying (B.27) and (B.31) with \( t = \log n \) and the triangle inequality, we obtain that under the constraint \( \delta_n^{1/2} \leq h \leq 1 \),

\[
\begin{align*}
\left| n^{1/2}(\alpha, \widehat{\beta}_h - \beta^*) - S_n^0 \right| & = n^{1/2} \left| \langle \Sigma^{1/2} D^{-1} \alpha, \Sigma^{1/2}(\widehat{\beta}_h - \beta^*) - \Sigma^{-1/2} \sum_{i=1}^n \tau - \bar{K}(-\epsilon_i/h) x_i \rangle \right| + \| \mathbb{E}(S_n) \| \\
& \leq c_1 \| D^{-1} \alpha \|_\Sigma \cdot n^{1/2}(h^{1/2}\delta_n + h^2) \tag{B.32}
\end{align*}
\]

with probability at least \( 1 - 3n^{-1} \) for some constant \( c_1 > 0 \).

For the centered partial sum \( S_n^0 = S_n - \mathbb{E}(S_n) = n^{-1/2} \sum_{i=1}^n (1 - \mathbb{E})\gamma_i \xi_i \), we have \( \text{var}(S_n^0) = \text{var}(S_n) = \mathbb{E}(\gamma^2 - \mathbb{E}(\gamma))^2 \), where \( \gamma = \langle D^{-1} \alpha, x \rangle \) and \( \xi = \tau - \bar{K}(-\epsilon/h) \). By the Berry-Esseen inequality (see, e.g., Tyurin (2011))

\[
\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left[ S_n^{0} \leq \text{var}(S_n^{0})^{1/2} x \right] - \Phi(x) \right| \leq \frac{\mathbb{E}(|\gamma| - \mathbb{E}(|\gamma|)^3)^2}{2(\mathbb{E}(|\gamma|)^2 - \mathbb{E}(|\gamma|)^3)^{3/2}} \sqrt{n}, \tag{B.33}
\]

We have shown that \( \mathbb{E}(\gamma^3) \leq \| D^{-1} \alpha \|_\Sigma \cdot h^2 \). Following a similar argument in the proof of Lemma B.1, we have \( \mathbb{E}(\xi^2|x) \leq \tau(1 - \tau) + h^2 \). For \( h \) sufficiently small, \( \text{var}(S_n) = \{ \tau(1 - \tau) + O(h) \} \| D^{-1} \alpha \|^2_\Sigma \) and \( \mathbb{E}(\gamma^3) \leq \max(\tau, 1 - \tau) \mathbb{E}(\xi^2)^3) \leq m_3 \{ \tau(1 - \tau) + O(h^2) \} \| D^{-1} \alpha \|^3_\Sigma \). Substituting these bounds into
Let $G$ be an application of Lemma A.7 in the supplement of Spokoiny and Zhilova (2015), for some constant $c_2 > 0$. Set $\sigma_r^2 = \tau(1 - \tau) \|D^{-1}a\|_{\Sigma}^2$, and we have shown that $|\text{var}(S_n)/\sigma_r^2 - 1| \leq h$. By an application of Lemma A.7 in the supplement of Spokoiny and Zhilova (2015), for sufficiently small $h$, we have

$$\sup_{x \in \mathbb{R}} |\text{var}(S_n)^{1/2} x - \Phi(x)| \leq c_2 n^{-1/2}$$

(B.34) yields

$$\sup_{x \in \mathbb{R}} |\text{var}(S_n)^{1/2} x - \Phi(x)| \leq c_2 n^{-1/2}$$

for some constant $c_2 > 0$. Setting $\sigma_r^2 = \tau(1 - \tau) \|D^{-1}a\|_{\Sigma}^2$, we have shown that $|\text{var}(S_n)/\sigma_r^2 - 1| \leq h$.

Before proceeding, we note that the constants $c_1 - c_3$ appeared above are all independent of $a$. Let $G \sim N(0, 1)$. Putting together the above derivations, for any $x \in \mathbb{R}$ and $a \in \mathbb{R}^p$, we obtain

$$\mathbb{P}(n^{1/2}\langle a, \tilde{\beta}_h - \beta^* \rangle \leq x)$$

$$\leq \mathbb{P}(S_n^0 \leq x + c_1 \|D^{-1}a\|_{\Sigma} \cdot n^{1/2}(h^{-1/2}\delta_n + h^2)) + 3n^{-1}$$

$$\leq \mathbb{P}(\text{var}(S_n)^{1/2} G \leq x + c_1 \|D^{-1}a\|_{\Sigma} \cdot n^{1/2}(h^{-1/2}\delta_n + h^2)) + c_2 n^{-1/2} + 3n^{-1}$$

$$\leq \mathbb{P}(\sigma_r G \leq x + c_1 \|D^{-1}a\|_{\Sigma} \cdot n^{1/2}(h^{-1/2}\delta_n + h^2)) + c_2 n^{-1/2} + c_3 h + 3n^{-1}$$

$$\leq \mathbb{P}(\sigma_r G \leq x) + c_1 (2\pi)^{-1/2} \|D^{-1}a\|_{\Sigma} \cdot n^{1/2}(h^{-1/2}\delta_n + h^2)/\sigma_r + c_2 n^{-1/2} + c_3 h + 3n^{-1},$$

where the first, second, and third inequalities hold by (B.32), (B.34), and (B.35), respectively, and the last inequality follows from the fact that for any $a \leq b$, $\Phi(b/\sigma_r) - \Phi(a/\sigma_r) \leq (2\pi)^{-1/2}(b-a)/\sigma_r$. A similar argument leads to a series of reverse inequalities. Note the above bounds are independent of $x$ and $a$, and therefore hold uniformly over all $x$ and $a$. Putting together the pieces, we conclude that under the bandwidth requirement $\delta_n^{1/2} \leq h \leq 1$,

$$\sup_{x \in \mathbb{R}, a \in \mathbb{R}^p} |\mathbb{P}(n^{1/2}\langle a, \tilde{\beta}_h - \beta^* \rangle \leq \sigma_r x) - \Phi(x)| \leq \frac{p + \log n}{(nh)^{1/2}} + n^{1/2}h^2,$$

as claimed. \(\square\)

### B.5 Proof of Theorem 3.4

Keep the notation used in the proof of Theorem 3.1. With nonnegative weights $w_i$, the corresponding weighted loss $\tilde{Q}_h^w$ in (2.9) is convex, and thus the first-order condition $\nabla \tilde{Q}_h^w(\tilde{\beta}_h^w) = 0$ holds. We use the same localized argument as in the proof of Theorem 3.1. For some $r > 0$ to be determined,
Lemma B.3. Assume that Conditions 3.1, 3.2, and 3.5 hold. For any $t \geq 0$, there exists some event $E_{l}(t)$ with $P[E_{l}(t)] \geq 1 - 2e^{-t}$ such that conditioned on $E_{l}(t)$, the bound $1/2(\beta^{*})^{2}_{\Sigma}$ holds, and with $P^{*}$-probability at least $1 - e^{-t}$,

$$
||\nabla \hat{Q}^{\beta}_{h}(\beta^{*})||\Omega \leq C_{1}\left(\sqrt{\frac{p + t}{n}} + \hat{h}^{2}\right)
$$

as long as $n \geq p + t$, where $C_{1} > 0$ is a constant depending only on $(v_{1}, \kappa_{2}, l_{0})$ and $\tau$.

Lemma B.4. Let $r = h/(4\eta_{1/4})$ with $\eta_{1/4}$ defined in (B.14), and $t \geq 0$. Suppose the sample size and bandwidth satisfy $(p + \log n + t)/n \leq h \leq 1$. Then, there exists some event $E_{l}(t)$ with $P[E_{l}(t)] \geq 1 - 3e^{-t}$ such that, conditioning on $E_{l}(t)$,

$$
\inf_{\beta \in \beta^{*} + \Theta(r)} \frac{D(\beta, \beta^{*})}{||\beta - \beta^{*}||^{2}_{\Sigma}} \geq 0.5k_{t}L
$$

and $P^{*}\left\{\inf_{\beta \in \beta^{*} + \Theta(r)} \frac{D(\beta, \beta^{*})}{||\beta - \beta^{*}||^{2}_{\Sigma}} \geq 0.5k_{t}L\right\} \geq 1 - e^{-t}. \quad \text{(B.38)}$

Recall from the proof of Theorem 3.1 that the upper bound on $||\hat{\beta}^{h}_{h} - \beta^{*}||\Sigma$ depends on $||\nabla \hat{Q}^{\beta}_{h}(\beta^{*})||\Omega$ and $\inf_{\beta \in \beta^{*} + \Theta(r)} [D(\beta, \beta^{*})/||\beta - \beta^{*}||^{2}_{\Sigma}]$. In view of Lemmas B.3 and B.4, the bound (3.6) holds whenever the event $E_{l}(t) := E_{l}(t) \cap E_{2}(t)$ occurs. Putting the bounds (B.36), (B.37) and (B.38) together, we conclude that with $P^{*}$-probability at least $1 - 2e^{-t}$ conditioned on $E_{l}(t)$,

$$
||\hat{\beta}^{h}_{h} - \beta^{*}||\Sigma \leq 2(k_{t}L)^{-1}C_{1}\left(\sqrt{\frac{p + t}{n}} + \hat{h}^{2}\right) < r = h/(4\eta_{1/4})
$$

as long as $\max\{(p + \log n + t)/n, \sqrt{(p + t)/n}\} \leq \hat{h} \leq 1$. Consequently, $\hat{\beta}^{h}_{h}$ falls into the interior of $\beta^{*} + \Theta(r)$, thereby implying $\eta = 1$ and $\hat{\beta}^{h}_{h} = \hat{\beta}^{h}_{h}$. This completes the proof. \qed
B.5.1 Proof of Lemma B.3

By the triangle inequality,

$$
\|\nabla \widehat{Q}_n^h(\beta^*)\|_\Omega \leq \|\nabla \widehat{Q}_n^h(\beta^*) - \nabla \widehat{Q}_h(\beta^*)\|_\Omega + \|\nabla \widehat{Q}_h(\beta^*)\|_\Omega := I_1 + I_2. 
$$

(B.39)

Lemma B.1 provides an upper bound for $I_2$ with high probability, and thus it suffices to bound $I_1$. To this end, we will employ Hoeffding’s and Bernstein’s inequalities.

Note that $\Sigma^{-1/2}\nabla \widehat{Q}_n^h(\beta^*) = (1/n) \sum_{i=1}^n w_i \cdot \xi_i w_i$ and $\mathbb{E}[\Sigma^{-1/2}\nabla \widehat{Q}_n^h(\beta^*)] = \Sigma^{-1/2}\widehat{Q}_h(\beta^*)$, where $w_i = \Sigma^{-1/2}x_i$ and $\xi_i = \bar{K}(-\xi_i/h) - \tau$ with $\xi_i = y_i - \langle x_i, \beta^* \rangle$. Using a similar covering argument as in the proof of Lemma B.1, for any $\epsilon \in (0, 1)$, there exists an $\epsilon$-net $\mathcal{N}_\epsilon$ of the unit sphere $\mathbb{S}^{p-1}$ with cardinality $|\mathcal{N}_\epsilon| \leq (1 + 2/\epsilon)^p$ such that

$$
\|\nabla \widehat{Q}_n^h(\beta^*) - \nabla \widehat{Q}_h(\beta^*)\|_\Omega \leq \frac{1}{1 - \epsilon} \max_{u \in \mathcal{N}_\epsilon} \frac{1}{n} \sum_{i=1}^n e_i \cdot \xi_i \langle u, w_i \rangle,
$$

where $e_i = w_i - 1$ are independent Rademacher random variables. By Hoeffding’s inequality,

$$
\mathbb{P}^* \left\{ \frac{1}{n} \sum_{i=1}^n e_i \xi_i \langle u, w_i \rangle \geq \left( \frac{1}{n} \sum_{i=1}^n \xi_i^2 \langle u, w_i \rangle^2 \right)^{1/2} \sqrt{2u/n} \right\} \leq e^{-u} \text{ for any } u \geq 0.
$$

For the data-dependent quantity $(1/n) \sum_{i=1}^n \xi_i^2 \langle u, w_i \rangle^2$, recall from the proof of Lemma B.1 that $|\xi_i| \leq \bar{\tau} = \max(\tau, 1 - \tau)$, $\mathbb{E}(\xi_i^2 | w_i) \leq \bar{\tau}^2 = (1 - \tau) + (1 + \tau)\log h^2$ and hence $\mathbb{E}(\xi_i^2 \langle u, w_i \rangle^2) \leq \bar{\tau}^2$. Combined with the sub-Gaussianity of $w_i$, we obtain that, for $k = 2, 3, \ldots,$

$$
\mathbb{E}(\xi_i^2 \langle u, w_i \rangle^2)^k \leq \bar{\tau}^2 e^{2(1-\epsilon)} u_1^2 \cdot 2k \int_0^{\infty} \mathbb{P}(\|u\| \geq u_1 u) u^{2k-1} du \\
= \bar{\tau}^2 e^{2(1-\epsilon)} u_1^2 \cdot 4k \int_0^{\infty} u^{2k-1} e^{-u^2/2} du \\
= \bar{\tau}^2 e^{2(1-\epsilon)k} u_1^{2k} \cdot 2k \int_0^{\infty} v^{k-1} e^{-v} dv = 2^{k+1} k! \cdot \bar{\tau}^2 e^{2(1-\epsilon)k} u_1^{2k}.
$$

In particular, $\mathbb{E}(\xi_i^4 \langle u, w_i \rangle^4) \leq (4\bar{\tau}^2)^2 u_1^4$ and $\mathbb{E}(\xi_i^2 \langle u, w_i \rangle^2)^k \leq \frac{\bar{\tau}^2}{2} \cdot (4\bar{\tau}^2)^2 u_1^4 \cdot (2\bar{\tau}^2 u_1^2)^{k-2}$ for $k \geq 3$.

With the above calculations, it follows from Bernstein’s inequality that

$$
\mathbb{P} \left\{ \frac{1}{n} \sum_{i=1}^n \xi_i^2 \langle u, w_i \rangle^2 \geq \bar{\tau}^2 + 4\bar{\tau}^2 v \sqrt{\frac{2v}{n} + 2u_1^2 \bar{\tau}^2} \right\} \leq e^{-v} \text{ for any } v \geq 0.
$$

In the above analysis, we set $\epsilon = 2/(\epsilon^2 - 1)$ so that $(1 + 2/\epsilon)^p = e^{2p}$, and take $u = v = 2p + t$. 

18
Applying the union bound, we conclude that

$$\max_{u \in \mathcal{N}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}^2_i \langle u, w_i \rangle^2 \leq \nu^2 + 4
\nu_t \sqrt{\frac{4p + 2t}{n}} + 2\nu_t^2 \mathbb{P} \left( \frac{p + t}{n} \right)$$  \hspace{1cm} (B.40)

with probability at least $1 - e^{-t}$. Let $E_1(t)$ be the event that (B.40) and (B.12) hold, so that $P[E_1(t)] \geq 1 - 2e^{-t}$. Using the union bound again, we obtain that with $\mathbb{P}^\ast$-probability at least $1 - e^{-t}$ conditioned on $E_1(t)$,

$$\|\nabla \hat{Q}_h(\beta^\ast) - \nabla \hat{Q}_h(\beta^\ast)\|_\Omega \leq C \sqrt{\frac{p + t}{n}}$$

as long as $n \geq p + t$, where $C > 0$ depends only on $(\nu_1, \kappa_2, l_0)$ and $T$. This, together with (B.39) and (B.12), proves the claimed bound. \hfill \square

### B.5.2 Proof of Lemma B.4

Using arguments similar to (B.17), (B.18), and (B.20) in the proof of Lemma B.2, we obtain

$$D^0(\beta, \beta^\ast) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \hat{K} \left( \frac{(x_i, \beta) - y_i}{h} \right) - \hat{K} \left( \frac{-\bar{e}_i}{h} \right) \right\} w_i \langle x_i, \beta - \beta^\ast \rangle$$

$$\geq \kappa \|\beta - \beta^\ast\|_\Sigma^2 \cdot \frac{1}{nh} \sum_{i=1}^{n} (1 + e_i) \omega_i \cdot \varphi_{h/(2r)}((w_i, \delta)),$$

where $\delta = \Sigma^{1/2}(\beta - \beta^\ast)/\|\beta - \beta^\ast\|_\Sigma$, $\omega_i = 1(|e_i| \leq h/2)$, and $\varphi(\cdot)$ is as in (B.19). Recall from (B.20) the definition of $D_0(\delta)$. We have

$$\inf_{\beta \in \beta + B(\nu)} \frac{D^0(\beta, \beta^\ast)}{\kappa \|\beta - \beta^\ast\|_\Sigma^2} \geq \inf_{\delta \in \mathbb{S}^{p-1}} D_0(\delta) - \sup_{\delta \in \mathbb{S}^{p-1}} \{D_0(\delta) - D_0^0(\delta)\}. \hspace{1cm} (B.41)$$

In what to follow, we obtain lower and upper bounds for $D_0(\delta)$ and $\sup_{\delta \in \mathbb{S}^{p-1}} \{D_0(\delta) - D_0^0(\delta)\}$, respectively. For the latter, we write

$$\Gamma_\delta = \Gamma_\delta(e_1, \ldots, e_n) = \sup_{\delta \in \mathbb{S}^{p-1}} (D_0(\delta) - D_0^0(\delta)) = \sup_{\delta \in \mathbb{S}^{p-1}} \frac{1}{nh} \sum_{i=1}^{n} e_i \omega_i \cdot \varphi_{h/(2r)}((w_i, \delta)).$$

Since $\varphi_\delta(u) \leq (R/2)^2$ and $e_i \in \{-1, 1\}$, we have $\mathbb{E}^\ast[|e_i \omega_i \cdot \varphi_{h/(2r)}((w_i, \delta))|^2] \leq (h/4r)^4 \omega_i$ and $|e_i \omega_i \cdot \varphi_{h/(2r)}((w_i, \delta))| \leq (h/4r)^2$. Then, by Bousquet’s version of Talagrand’s inequality with explicit
constants (see Theorem 7.3 in Bousquet (2003)), for every $t \geq 0$,

$$\Gamma_n \leq \mathbb{E}^*(\Gamma_n) + \sqrt{\frac{h^2}{(4r)^4} \frac{1}{n} \sum_{i=1}^{n} \omega_i \cdot \frac{2r}{n} + 4\mathbb{E}^*(\Gamma_n) \frac{h}{(4r)^2} \frac{t}{n} + \frac{h}{(4r)^2} \frac{t}{3n}}$$

$$\leq 2\mathbb{E}^*(\Gamma_n) + \frac{h}{(4r)^2} \left( \frac{1}{n} \sum_{i=1}^{n} \omega_i \right)^{1/2} \sqrt{\frac{2r}{n} + 4t}$$

with probability at least $1 - e^{-t}$. Further, by the Lipschitz continuity of $u \to \varphi_E(u)$ and Talagrand’s contraction principle,

$$\mathbb{E}^*(\Gamma_n) \leq \frac{1}{2r} \mathbb{E}^* \left\{ \sup_{\delta \in \mathbb{S}^{p-1}} \frac{1}{n} \sum_{i=1}^{n} e_i(\omega_i w_i, \delta) \right\}$$

$$\leq \frac{1}{2r} \mathbb{E}^* \left\| \frac{1}{n} \sum_{i=1}^{n} e_i(\omega_i w_i) \right\|_2$$

$$\leq \frac{1}{2r} \left( \sum_{i=1}^{n} \omega_i \|w_i\|_2 \right)^{1/2}$$

$$\leq \max_{1 \leq i \leq n} \|w_i\|_2 \cdot \left( \frac{1}{n} \sum_{i=1}^{n} \omega_i \right)^{1/2} \cdot \frac{1}{2r n^{1/2}}.$$

Together, the last two displays imply

$$\Gamma_n \leq \left( \frac{1}{n} \sum_{i=1}^{n} \omega_i \right)^{1/2} \left( \max_{1 \leq i \leq n} \|w_i\|_2 \cdot \frac{1}{rn^{1/2}} + \frac{h}{(4r)^2} \frac{2r}{n} + \frac{h}{(4r)^2} \frac{4t}{3n} \right) \quad (B.42)$$

with $\mathbb{P}^*$-probability at least $1 - e^{-t}$.

Next we provide upper bounds for the data-dependent quantities $\max_{1 \leq i \leq n} \|w_i\|_2$ and $(1/n) \sum_{i=1}^{n} \omega_i$.

As in the proof of Lemma B.3, for any $\epsilon \in (0, 1)$, there exists a subset $\mathcal{N}_\epsilon \subseteq \mathbb{S}^{p-1}$ with $|\mathcal{N}_\epsilon| \leq (1 + 2/e)^p$ such that $\max_{1 \leq i \leq n} \|w_i\|_2 \leq (1 - \epsilon)^{-1} \max_{u \in \mathcal{N}_\epsilon} \langle u, w_i \rangle$. Given $1 \leq i \leq n$ and $u \in \mathcal{N}_\epsilon$, Condition 3.5 indicates $\mathbb{P}(\langle (u, w_i) \rangle \geq \nu_1 u) \leq 2e^{-\nu_1^2/2}$ for any $u \in \mathbb{R}$. Taking the union bound over $i$ and $u$, and setting $u = \sqrt{2t + 2 \log(2n) + 2p \log(1 + 2/e)} \ (t > 0)$, we obtain that with probability at least $1 - 2n(1 + 2/e)^p e^{-\nu_1^2/2} = 1 - e^{-t}$, $\max_{1 \leq i \leq n} \|w_i\|_2 \leq (1 - \epsilon)^{-1} \nu_1 \sqrt{2t + 2 \log(2n) + 2p \log(1 + 2/e)}$.

Minimizing this upper bound with respect to $\epsilon \in (0, 1)$, we conclude that for any $t > 0$, the event

$$\mathcal{E}_{\max}(t) := \left\{ \max_{1 \leq i \leq n} \|w_i\|_2^2 \leq 2\nu_1^2 \left[ 3.7p + \log(2n) + t \right] \right\} \quad (B.43)$$

occurs with probability at least $1 - e^{-t}$. For $(1/n) \sum_{i=1}^{n} \omega_i = (1/n) \sum_{i=1}^{n} \mathbb{I}(|x_i| \leq h/2)$, note that
\[ \mathbb{E} \omega_i \leq f_i h, \] and by Bernstein’s inequality,
\[
\frac{1}{n} \sum_{i=1}^{n} \omega_i \leq \mathbb{E} \omega_i + \sqrt{\mathbb{E} \omega_i \cdot \frac{2t}{n} + \frac{t}{3n}} \leq \left( \sqrt{\mathbb{E} \omega_i} + \sqrt{\frac{t}{2n}} \right)^2
\]
with probability greater than 1 - \( e^{-t} \). Let be \( E_{\text{smo}}(t) \) the event that this bound holds, where “smo” stands for “smoothing”. Putting together the pieces, we conclude that conditioned on \( E_{\text{max}}(t) \cap E_{\text{smo}}(t) \),

\[
\sup_{\delta \in \mathcal{B} \cap \mathcal{D}} \{ D_0(\delta) - D_0(\delta) \} 
\leq \left( f_i h \right)^{1/2} + (0.5t/n)^{1/2} \left\{ \frac{\nu_1(7.4p + 2 \log(2n) + 2r)^{1/2}}{rn^{1/2}} + \frac{h}{(4r)^2} \sqrt{\frac{2r}{n}} \right\} + \frac{h}{(4r)^2} \sqrt{\frac{4t}{3n}}
\] (B.44)

holds with \( \mathbb{P}^\ast \)-probability greater than 1 - \( e^{-t} \).

Turning to \( D_0(\delta) \), let \( E_{\text{lsc}}(t) \) be the event that the bound (B.15) with \( r = h/(4\eta_1/4) \) holds, implying local strong convexity. By Lemma B.2, \( \mathbb{P}[E_{\text{lsc}}(t)] \geq 1 - e^{-t} \). Provided that \( (p + \log n + t)/n \leq h \leq 1 \), substituting (B.44) and (B.15) into (B.41) yields

\[
\inf_{\beta \in (\beta^* + \Theta(r))} \frac{D^\beta(\beta, \beta^*)}{E[||\beta - \beta^*||^2]} \geq \frac{3}{4}(f - 0.5l_0 h) - C \sqrt{\frac{p + \log n + t}{nh}} \geq \frac{1}{2} f
\]

with \( \mathbb{P}^\ast \)-probability greater than 1 - \( e^{-t} \) conditioned on \( E_{\text{max}}(t) \cap E_{\text{smo}}(t) \cap E_{\text{lsc}}(t) \). This completes the proof of (B.38) by setting \( E_2(t) = E_{\text{max}}(t) \cap E_{\text{smo}}(t) \cap E_{\text{lsc}}(t) \). \( \square \)

**B.6 Proof of Theorem 3.5**

The proof is based on an argument similar to that used in the proof of Theorem 3.2. To begin with, define the random process

\[
\Delta^\beta(\beta) = \Sigma^{-1/2} [\nabla \hat{Q}^\beta_h(\beta) - \nabla \hat{Q}^\beta_h(\beta^*) - D(\beta - \beta^*)], \ \beta \in \mathbb{R}^p. \quad (B.45)
\]

For a prespecified \( r > 0 \), a key step is to bound the local fluctuation \( \sup_{\beta \in (\beta^* + \Theta(r))} ||\Delta^\beta(\beta)||_2 \). Since \( \mathbb{E}(w_i) = 1 \), we have \( \mathbb{E}[\nabla \hat{Q}^\beta_h(\beta)] = \nabla \hat{Q}_h(\beta) \). Define the (conditionally) centered process

\[
G^\beta(\beta) = \Sigma^{-1/2} [\nabla \hat{Q}^\beta_h(\beta) - \nabla \hat{Q}_h(\beta)] = \frac{1}{n} \sum_{i=1}^{n} [\hat{K}(x_i^T \beta - y_i)/h - \tau] e_i w_i,
\]

21
so that $\Delta^h(\beta)$ be be written as

$$\Delta^h(\beta) = \{ G^h(\beta) - G^h(\beta^*) \} + \Delta(\beta),$$

where $\Delta(\beta)$ is defined in (B.27). By the triangle inequality,

$$\sup_{\beta \in \beta^* + \Theta(r)} \|\Delta^h(\beta)\|_2 \leq \sup_{\beta \in \beta^* + \Theta(r)} \|G^h(\beta) - G^h(\beta^*)\|_2 + \sup_{\beta \in \beta^* + \Theta(r)} \|\Delta(\beta)\|_2. \tag{B.46}$$

Let $E_3(t)$ denote the event that the bound (B.31) holds. It suffices to deal with the first term on the right-hand side of (B.46). Using a change of variable $\delta = \Sigma^{1/2}(\beta - \beta^*) \in \mathbb{B}^p(r)$ for $\beta \in \beta^* + \Theta(r)$, we have $y_i - x_i^T \beta = \varepsilon_i - w_i^T \delta$ and

$$\sup_{\beta \in \beta^* + \Theta(r)} \|G^h(\beta) - G^h(\beta^*)\|_2 \leq \sup_{\varepsilon \in \mathbb{B}^p(r)} \|G^h(\beta^* + \Sigma^{-1/2} \delta) - G^h(\beta^*)\|_2, \tag{B.47}$$

where $\Delta^h_0(\delta) = (1/n) \sum_{i=1}^n \varepsilon_i w_i \{ \bar{K}(w_i^T \delta - \varepsilon_i) / h) - \bar{K}(-\varepsilon_i / h) \}$ satisfies $\Delta^h_0(0) = 0$ and $\mathbb{E}^*[\Delta^h_0(\delta)] = 0$. Note that $\nabla \Delta^h_0(\delta) = (1/n) \sum_{i=1}^n \varepsilon_i K_i, \delta w_i w_i^T$, where $K_i, \delta = K_h(\varepsilon_i^T \delta - \varepsilon_i)$. For any $\lambda \in \mathbb{R}$ and $u, v \in \mathbb{S}^{p-1}$, we have

$$\mathbb{E}^* \exp[\lambda n^{1/2} u^T \Delta^h_0(\delta) v] = \prod_{i=1}^n \mathbb{E}^* \exp[\lambda n^{-1/2} \varepsilon_i K_i, \delta w_i w_i^T u \cdot w_i^T v] \leq \prod_{i=1}^n \exp \left\{ \frac{\lambda^2}{2n} K_i^2(\varepsilon_i^T u \cdot w_i^T v)^2 \right\} = \exp \left\{ \frac{\lambda^2}{2n} \sum_{i=1}^n K_i^2(\varepsilon_i^T u \cdot w_i^T v)^2 \right\}.$$

Note that $K_i, \delta \leq \kappa u h^{-1}$, and by Hölder’s inequality,

$$\frac{1}{n} \sum_{i=1}^n K_i^2(\varepsilon_i^T u \cdot w_i^T v)^2 \leq \frac{\kappa u}{h} \left( \frac{1}{n} \sum_{i=1}^n K_i, \delta(w_i^T u)^4 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^n K_i, \delta(w_i^T v)^4 \right)^{1/2}.$$

Define the function $\Lambda_{r,h}(\cdot, \cdot) : \mathbb{R}^p \times \mathbb{R}^p \mapsto [0, \infty)$

$$\Lambda_{r,h}(u, v) = \frac{1}{n} \sum_{i=1}^n K_h(r w_i^T v - \varepsilon_i)(w_i^T u)^4 \text{ for } u, v \in \mathbb{S}^{p-1}, \tag{B.48}$$

and write $\|\Lambda_{r,h}\|_\infty = \sup_{u, v \in \mathbb{S}^{p-1}} \Lambda_{r,h}(u, v)$. With this notation, it follows that for any $\lambda \in \mathbb{R}$ and
\( u, v \in S^{p-1}, \)

\[
\sup_{\delta \in \mathbb{B}^p(r)} \mathbb{E}^\ast \exp \{ \lambda n^{1/2} u^\top \Delta_0^b(\delta) v \} \leq \exp \left\{ \frac{4^2}{2} \kappa_h - \epsilon \| \Lambda_{r,h} \|_\infty \right\}.
\]

Thus, applying a conditional version of Theorem A.3 in Spokoiny (2013) yields

\[
\sup_{\delta \in \mathbb{B}^p(r)} \| \Delta_0^b(\delta) \|_2 \leq 6 \kappa_u 2^{1/2} \| \Lambda_{r,h} \|_\infty^{1/2} \cdot r \sqrt{\frac{4p + 2t}{nh}} \tag{B.49}
\]

with \( \mathbb{P}^\ast \)-probability at least \( 1 - e^{-t} \).

Next, we bound the data-dependent quantity \( \| \Lambda_{r,h} \|_\infty \). For any \( \epsilon_1, \epsilon_2 \in (0, 1) \), there exist \( \epsilon_1 \)- and \( \epsilon_2 \)-nets \( \{ u_1, \ldots, u_{d_1} \} \) and \( \{ v_1, \ldots, v_{d_2} \} \) of \( S^{p-1} \) with \( d_1 \leq (1 + 2/\epsilon_1)^p \) and \( d_2 \leq (1 + 2/\epsilon_2)^p \). Given \( u, v \in S^{p-1} \), there exist some \( 1 \leq j \leq d_1 \) and \( 1 \leq k \leq d_2 \) such that \( \| u - u_j \|_2 \leq \epsilon_1 \) and \( \| v - v_k \|_2 \leq \epsilon_2 \).

At \( (u, v) \), consider the decomposition

\[
\Lambda_{r,h}(u, v) = \Lambda_{r,h}(u, v) - \Lambda_{r,h}(u, v) + \Lambda_{r,h}(u, v).
\]

For \( \Lambda_{r,h}(u, v) - \Lambda_{r,h}(u, v) \), the Lipschitz continuity of \( K(\cdot) \) ensures that

\[
| \Lambda_{r,h}(u, v) - \Lambda_{r,h}(u, v) | \leq \frac{kr}{n h^2} \sum_{i=1}^n | w_i^\top (v - v_k) (w_i^\top u) |^4 \leq \frac{kr \epsilon_2}{h^2} \cdot \max_{1 \leq i \leq n} \| w_i \|_2 \cdot \frac{1}{n} \sum_{i=1}^n (w_i^\top u)^4 \tag{B.50}
\]

For \( \Lambda_{r,h}(u, v) \), by the triangle inequality for the \( L^4 \)-norm we have

\[
\Lambda_{r,h}^{1/4}(u, v_k) = \left\{ \frac{1}{n} \sum_{i=1}^n K_h(r w_i^\top v_k - \epsilon_i)(w_i^\top u) \right\}^{1/4} \leq \left\{ \frac{1}{n} \sum_{i=1}^n K_h(r w_i^\top v_k - \epsilon_i)(w_i^\top u) \right\}^{1/4} + \left\{ \frac{1}{n} \sum_{i=1}^n K_h(r w_i^\top v_k - \epsilon_i)(w_i^\top u) \right\}^{1/4} \leq \Lambda_{r,h}(u_j, v_k) + \epsilon_1 \cdot \sup_{u \in S^{p-1}} \Lambda_{r,h}^{1/4}(u, v_k),
\]

which in turn implies

\[
\sup_{u \in S^{p-1}} \Lambda_{r,h}(u, v_k) \leq (1 - \epsilon_1)^{-4} \max_{1 \leq j \leq d_1} \Lambda_{r,h}(u_j, v_k). \tag{B.51}
\]

In view of (B.50) and (B.51), it suffices to bound \( \max_{1 \leq i \leq n} \| w_i \|_2, \sup_{u \in S^{p-1}} (1/n) \sum_{i=1}^n (w_i^\top u)^4 \) and
\[
\max_{(j,k) \in [d_1] \times [d_2]} \Lambda_{r,h}(u_j, v_k). \]

By (B.43), the bound
\[
\max_{1 \leq i \leq n} \|w_i\|_2 \leq (p + \log n + t)^{1/2} \text{ holds on } \mathcal{E}_{\max}(t).
\]

For the supremum \(\sup_{u \in \mathbb{B}^{p-1}} (1/n) \sum_{i=1}^{n} (w_i^4 u_i)^d\), similarly to (B.51) it can be shown that
\[
\sup_{u \in \mathbb{B}^{p-1}} \frac{1}{n} \sum_{i=1}^{n} (w_i^4 u_i)^d \leq (1 - \epsilon_1)^{-d} \max_{1 \leq i \leq d_1} \frac{1}{n} \sum_{i=1}^{n} (w_i^4 u_i)^d.
\]

Fix \(j \) and \(k \), Condition 3.5 implies
\[
\mathbb{E} e^{[\|w_j^4 u_j\|^2 / (36u_1^4)]^{1/2}} = \mathbb{E} e^{[\|w_j^4 u_j\|^2 / (6u_1^2)]} = 1 + \int_0^{\infty} e^{tu} \mathbb{P} [|w_j^4 u_j| \geq u_1(6u)^{1/2}] du \leq 1 + 2 \int_0^{\infty} e^{u-3u} du = 1 + 1 = 2.
\]

Therefore, \(\|(w_j^4 u_j)^d\|_{\psi_{1/2}} \leq 36u_1^4\), where \(\| \cdot \|_{\psi_r}\) denotes the \(\psi_r\)-norm \((r > 0)\); see Definition 2.1 in Adamczak et al. (2011). Since \(0 \leq K_h(rw_j^4 v_k - \epsilon_i) \leq \kappa_h^{-1}\), it is easy to see that \(\|K_h(rw_j^4 v_k - \epsilon_i)(w_j^4 u_j)^d\|_{\psi_{1/2}} \leq 36\kappa_h u_1^4 h^{-1}\). Moreover, note that \(\mathbb{E}(w_j^4 u_j)^d \leq m_d\) and
\[
\mathbb{E}[K_h(rw_j^4 v_k - \epsilon_i)(w_j^4 u_j)^d] = \mathbb{E}[\mathbb{E}[K_h(rw_j^4 v_k - \epsilon_i)|x_i](w_j^4 u_j)^d] \leq \bar{f} m_d.
\]

Hence, for any \(u, v \geq 3\), applying inequality (3.6) (and those above it) with \(s = 1/2\) in Adamczak et al. (2011) and the union bound, we obtain that
\[
P \left\{ \max_{1 \leq i \leq d_1} \frac{1}{n} \sum_{i=1}^{n} (w_i^4 u_i)^d \geq m_d + C_1 u_1^4 \left( \sqrt{\frac{u}{n}} + \frac{u^2}{n} \right) \right\} \leq d_1 e^{-u}
\]
and
\[
P \left\{ \max_{(j,k) \in [d_1] \times [d_2]} \Lambda(u_j, v_k) \geq \bar{f} m_d + C_1 \kappa_h u_1^4 \left( \sqrt{\frac{v}{nh^2}} + \frac{v^2}{nh} \right) \right\} \leq d_1 d_2 e^{-v}.
\]

Taking \(\epsilon_1 = 1 - 2^{-1/4}, \epsilon_2 = n^{-2}, u = p \log(1 + 2/\epsilon_1) + t\) and \(v = p \log(1 + 2/\epsilon_1)(1 + 2/\epsilon_2) + t\) in the above bounds, it follows that with probability at least \(1 - 2e^{-t}\),
\[
\max_{1 \leq i \leq d_1} \frac{1}{n} \sum_{i=1}^{n} (w_i^4 u_i)^d \leq m_d + C_2 u_1^4 \left( \sqrt{\frac{p + t}{n}} + \frac{(p + t)^2}{n} \right)
\]
Denote by $E_4(t)$ the event that the above two bounds hold. For any $r \leq h$, conditioned on $E_{\max}(t) \cap E_4(t)$ with $t = \log n$, we have

$$
\|A_{r,h}\|_\infty \leq 2 \max_{(j,k) \in [d_1] \times [d_2]} A(u_j, v_k) + \frac{2\sqrt{t}}{n^2} \cdot \max_{1 \leq i \leq n} \|w_i\|_2 \cdot \sup_{1 \leq j \leq d_1} \frac{1}{n} \sum_{i=1}^{n} (w_i^j u_j)^4
$$

\begin{equation}
\leq 2 \tilde{f} m_4 + C_4 h^{-1} \left\{ \sqrt{\frac{p \log n}{n}} + \frac{(p \log n)^2}{n} \right\}. \tag{B.52}
\end{equation}

With the above preparations, we are ready to prove the Bahadur representation for the bootstrap estimate. Let $E_1(t)$ and $E_2(t) = E_{\max}(t) \cap E_{\text{smo}}(t) \cap E_{\text{dic}}(t)$ be the events defined in the proof of Lemmas B.3 and B.4. In the rest of the proof, we take $t = \log n$ and set the bandwidth $h \asymp (q/n)^{2/5}$ with $q = p + \log n$. Then, conditioned on $E_1(t) \cap E_2(t)$, $\|\hat{\beta}_h - \beta^*\|_{\Sigma} \leq \rho_{\text{est}} = \sqrt{q/n}$, and $\|\hat{\beta}_h - \beta^*\|_{\Sigma} \leq \rho_{\text{est}} = \sqrt{q/n}$ with $\mathbb{P}^*$-probability at least $1 - 2n^{-1}$. Conditioned further on $E_3(t) \cap E_4(t)$,

$$
\|D(\hat{\beta}_h - \beta^*) + \nabla \hat{Q}_h(\beta^*)\|_{\Omega} = \|D(\hat{\beta}_h)\|_2 \leq \sup_{\beta \in \beta^* + \Theta_{\text{est}}} \|\Delta(\beta)\|_2 \leq \left(\frac{q}{n}\right)^{4/5},
$$

and with $\mathbb{P}^*$-probability at least $1 - 3n^{-1}$,

$$
\|D(\hat{\beta}_h - \beta^*) + \nabla \hat{Q}_h(\beta^*)\|_{\Omega} = \|D(\hat{\beta}_h)\|_2 \leq \sup_{\beta \in \beta^* + \Theta_{\text{est}}} \|\Delta(\beta)\|_2 \leq \left(\frac{q}{n}\right)^{4/5} \sqrt{\left(\frac{q}{n}\right)^{3/5} \left(\frac{p \log n}{n}\right)^{1/4}} + \left(\frac{q}{n}\right)^{3/5} \frac{p \log n}{n^{1/2}}.
$$

Together, the above two bounds proves the claimed result. \qed

## C Theoretical Properties of One-step Conquer

In this section, we provide theoretical properties of the one-step conquer estimator $\tilde{\beta}$, defined in Section A.1. The key message is that, when higher-order kernels are used (and if the conditional density $f_{\text{smo}}(\cdot)$ has enough derivatives), the asymptotic normality of the one-step estimator holds under weaker growth conditions on $p$. For example, the scaling condition $p = o(n^{3/8})$ that is required for the conquer estimator can be reduced to roughly $p = o(n^{7/16})$ for the one-step conquer estimator using a kernel of order 4.
**Condition 1.** Let $G(\cdot)$ be a symmetric kernel of order $\nu > 2$, that is, $\int_{-\infty}^{\infty} t^k G(u) \, du = 0$ for $k = 1, \ldots, \nu - 1$ and $\int_{-\infty}^{\infty} u G(u) \, du \neq 0$. Moreover, $\kappa_{G,k} = \int_{-\infty}^{\infty} |u|^k G(u) \, du < \infty$ for $1 \leq k \leq \nu$, $G$ is uniformly bounded with $\kappa_{G,u} = \sup_{u \in \mathbb{R}} |G(u)| < \infty$ and is $l_G$-Lipschitz continuous for some $l_G > 0$.

The use of a higher-order kernel does not necessarily reduce bias unless the conditional density $f_{\varepsilon|x}(\cdot)$ of $\varepsilon$ given $x$ is sufficiently smooth. Therefore, we further impose the following smoothness conditions on $f_{\varepsilon|x}(\cdot)$.

**Condition 2.** Let $\nu \geq 4$ be the integer in Condition 1. The conditional density $f_{\varepsilon|x}(\cdot)$ is $(\nu - 1)$-times differentiable, and satisfies $|f_{\varepsilon|x}^{(\nu-2)}(u) - f_{\varepsilon|x}^{(\nu-2)}(0)| \leq l_{\nu-2} |u|$ for all $u \in \mathbb{R}$ almost surely (over the random vector $x$), where $l_{\nu-2} > 0$ is a constant. Also, there exists some constant $C_G > 0$ such that $\int_{-\infty}^{\infty} |u|^{\nu-2} G(u) \cdot \sup_{|t| \leq |u|} |f_{\varepsilon|x}^{(\nu-1)}(t) - f_{\varepsilon|x}^{(\nu-1)}(0)| \, du \leq C_G$ almost surely.

Notably, we have

$$\nabla Q_b^G(\beta) = \mathbb{E}[\nabla G((x, \beta) - y)/b - \tau] x \quad \text{and} \quad \nabla^2 Q_b^G(\beta) = \mathbb{E}[G_y(y - (x, \beta)) : x : x^\top], \quad (C.1)$$

representing the population score and Hessian of $Q_b^G(\cdot)$. As $b \to 0$, we expect $\nabla Q_b^G(\beta^*)$ and $\nabla^2 Q_b^G(\beta^*)$ to converge to $0$ (zero vector in $\mathbb{R}^p$) and $\mathbf{D} = \mathbb{E}[f_{\varepsilon|x}(0) x : x^\top]$, respectively. Recall that $\Omega = \Sigma^{-1}$, and $\|u\|_\Omega = \|\Sigma^{-1/2} u\|_2$ for $u \in \mathbb{R}^p$. For any symmetric matrix $A \in \mathbb{R}^{p \times p}$, with slight abuse of notation we use $\| \cdot \|_\Omega$ to denote the relative operator norm, defined as $\|A\|_\Omega = \|\Sigma^{-1/2} A \Sigma^{-1/2}\|_2$.

The following proposition validates this claim by providing explicit error bounds.

**Proposition C.1.** Let $b \in (0, 1)$ be a bandwidth. Under Conditions 1 and 2, we have

$$\|\nabla Q_b^G(\beta^*)\|_\Omega \leq l_{\nu-2} \kappa_{G,v} b^{\nu}/\nu! \quad \text{and} \quad \|\nabla^2 Q_b^G(\beta^*) - \mathbf{D}_0\|_\Omega \leq C_G b^{\nu-1}/(\nu - 1)!,$$

where $\mathbf{D}_0 = \Sigma^{-1/2} \mathbf{D} \Sigma^{-1/2} = \mathbb{E}[f_{\varepsilon|x}(0) w : w^\top]$ with $w = \Sigma^{-1/2} x$.

Proposition C.2 shows that when a higher-order kernel is used, the bias is significantly reduced in the sense that $\|\nabla Q_b^G(\beta^*)\|_2 = O(b^{\nu})$ and $\|\nabla^2 Q_b^G(\beta^*) - \mathbf{D}\|_2 = O(b^{\nu-1})$, where $\nu \geq 4$ is an even integer. Notably, even if the kernel $G$ has negative parts, the population Hessian $\nabla^2 Q_b^G(\beta^*)$ preserves the positive definiteness of $\mathbf{D}$ as long as the bandwidth $b$ is sufficiently small.

To construct the one-step conquer estimator, two key quantities are the sample Hessian $\nabla^2 \tilde{Q}_b^G(\cdot)$ and sample gradient $\nabla \tilde{Q}_b^G(\cdot)$, both evaluated at $\mathbf{\bar{\beta}}$, a consistent initial estimate. In the next two
propositions, we establish uniform convergence results of the Hessian and gradient of the empirical smoothed loss to their population counterparts. As a direct consequence, $\nabla^2 \tilde{Q}_h^G(\beta)$ is positive definite with high probability, provided that $\beta$ is consistent (i.e., in a local vicinity of $\beta^*$). To be more specific, for $r > 0$, we define the local neighborhood

$$\Theta^*(r) = \{ \beta \in \mathbb{R}^p : \|\beta - \beta^*\|_{\Sigma} \leq r \}. \quad (C.2)$$

**Proposition C.2.** Conditions 1, 2 and 3.5 ensure that, with probability at least $1 - e^{-t}$,

$$\sup_{\beta \in \Theta^*(r)} \| \nabla^2 \tilde{Q}_h^G(\beta) - \nabla^2 Q_h^G(\beta) \|_{\Omega} \lesssim \sqrt{p \log n + t} \frac{p \log n + t}{nb} + \frac{(p + t)^{1/2} r}{nb^2}$$

as long as $n \gtrsim p + t$.

**Proposition C.3.** Conditions 1, 2 and 3.5 ensure that, with probability at least $1 - e^{-t}$,

$$\sup_{\beta \in \Theta^*(r)} \| \nabla \tilde{Q}_h^G(\beta) - \nabla Q_h^G(\beta^*) - \nabla \hat{D}(\beta - \beta^*) \|_{\Omega} \lesssim r \left( \sqrt{\frac{p + t}{nb}} + r + b^{\nu - 1} \right) \quad (C.3)$$

as long as $\sqrt{(p + t)/n} \lesssim b$.

With the above preparations, we are ready to present the Bahadur representation for the one-step conquer estimator $\hat{\beta}$.

**Theorem C.1.** Assume Conditions 3.1, 3.2 and 3.5 in the main text and Conditions 1 and 2 hold. For any $t > 0$, let the sample size $n$, dimension $p$ and the bandwidths $h, b > 0$ satisfy $n \gtrsim p (\log n)^2 + t, \sqrt{(p + t)/n} \lesssim h \lesssim ((p + t)/n)^{1/4}$ and $\sqrt{(p + t)/n} \lesssim b \lesssim ((p + t)/n)^{1/(2\nu)}$. Then, the one-step conquer estimator $\tilde{\beta}$ satisfies the bound

$$\left\| \hat{D}(\tilde{\beta} - \beta^*) - \frac{1}{n} \sum_{i=1}^n [\tau - \hat{G}(\varepsilon_i/h)]_+ \right\|_{\Omega} \lesssim \left\{ \sqrt{\frac{(p \log n + t)/(nb)}{n}} + \frac{b^{\nu - 1}}{\text{bias term}} \right\} \left[ \frac{p + t}{n} \right] \quad (C.4)$$

with probability at least $1 - 5e^{-t}$, where $\hat{G}(u) = \int_{-\infty}^u \hat{G}(v) \, dv$.

Theorem C.1 shows that using a higher-order kernel ($\nu \geq 4$) allows one to choose larger bandwidth, thereby reducing the “variance” and the total Bahadur linearization error. Similarly to Theorem 3.3 in the main text, the following asymptotic normal approximation result for linear projections of one-step conquer is a direct consequence of Theorem C.1.
Theorem C.2. Assume Conditions 3.1, 3.2 and 3.5 in the main text and Conditions 1 and 2 hold. Let the bandwidths satisfy \((q/n)^{1/2} \leq h \leq (q/n)^{1/4}\) and \((q/n)^{1/2} \leq b \leq (q/n)^{1/(2\nu)}\), where \(q := p + \log n\). Then,

\[
\sup_{x \in \mathbb{R}, a \in \mathbb{R}^p} \left| \mathbb{P}(n^{1/2}(a, \tilde{\beta} - \beta^*) \leq \sigma_x x) - \Phi(x) \right| \leq \sqrt{\frac{(p + \log n)\log n}{nb}} + n^{1/2}b^*, \quad (C.5)
\]

where \(\sigma_x^2 = \tau(1 - \tau)||D^{-1}a||^2_{\Sigma}\). In particular, with a choice of bandwidth \(b \asymp (q/n)^{2/(2\nu+1)}\),

\[
\sup_{x \in \mathbb{R}, a \in \mathbb{R}^p} \left| \mathbb{P}(n^{1/2}(a, \tilde{\beta} - \beta^*) \leq \sigma_x x) - \Phi(x) \right| \to 0
\]

as \(n, p \to \infty\) under the scaling \(p^{4\nu/(2\nu-1)(\log n)^{2(\nu+1)/(2\nu-1)}} = o(n)\).

Let \(G(\cdot)\) be a kernel of order \(\nu = 4\). In view of Theorem C.2, we take \(h \asymp (p + \log n)/n^{2/5}\) as in the main text and \(b = (p + \log n)/n^{2/9}\), thereby obtaining that \(n^{1/2}(a, \tilde{\beta} - \beta)\), for an arbitrary \(a \in \mathbb{R}^p\), is asymptotically normally distributed as long as \(p\log n)^{9/16} = o(n^{7/16})\) as \(n \to \infty\).

C.1 Proof of Proposition C.1

We start from the gradient \(\Sigma^{-1/2}\nabla Q^G_\beta(\beta^*) = \mathbb{E}[\tilde{G}(-\varepsilon/b) - \tau|w = \Sigma^{-1/2}x]\). Let \(\mathbb{E}_x\) be the conditional expectation given \(x\). By integration by parts,

\[
\mathbb{E}_x \tilde{G}(-\varepsilon/b) = \int_{-\infty}^{\infty} \tilde{G}(-t/b) dF_{\varepsilon|x}(t) = \int_{-\infty}^{\infty} G(u)F_{\varepsilon|x}(-bu) du. \quad (C.6)
\]

Applying a Taylor expansion with integral remainder on \(F_{\varepsilon|x}(-bu)\) yields

\[
F_{\varepsilon|x}(-bu) = F_{\varepsilon|x}(0) + \sum_{\ell=1}^{v-1} \frac{(-bu)^\ell}{\ell!} \int_0^1 (1-w)^{v-2}\left\{ F_{\varepsilon|x}^{(v-2)}(-buw) - F_{\varepsilon|x}^{(v-2)}(0) \right\} dw
\]

\[
= \tau + \sum_{\ell=0}^{v-1} \frac{(-bu)^{\ell+1}}{(\ell+1)!} \int_0^1 (1-w)^{v-2}\left\{ F_{\varepsilon|x}^{(v-2)}(-buw) - f_{\varepsilon|x}^{(v-2)}(0) \right\} dw.
\]

Recall that \(G\) is a kernel of order \(\nu \geq 4\) (an even integer) and \(\kappa_{G, \nu} = \int_{-\infty}^{\infty} |u^\nu G(u)| du < \infty\). Substituting the above expansion into (C.6), we obtain

\[
\mathbb{E}_x \tilde{G}(-\varepsilon/b) = \tau - \frac{b^{v-1}}{(v-2)!} \int_{-\infty}^{\infty} \int_0^1 u^{v-1} G(u)(1-w)^{v-2}\left\{ f_{\varepsilon|x}^{(v-2)}(-buw) - f_{\varepsilon|x}^{(v-2)}(0) \right\} dw du.
\]
Furthermore, by the Lipschitz continuity of $f^{(\nu - 2)}(\cdot)$ around 0,

$$
\|E_x \mathcal{G}(-\varepsilon/b) - \tau\| \leq \frac{l_{\nu-2}b^\nu}{(\nu - 2)!} \int_{-\infty}^{\infty} \int_{0}^{1} |u'|G(u)|(1 - w)^{\nu - 2} w \, dw \, du = B(2, \nu - 1)l_{\nu-2}\kappa_{G,\nu} b^\nu/(\nu - 2)!,
$$

where $B(x, y) := \int_{0}^{1} t^{x-1}(1 - t)^{y-1} \, dt$ denotes the beta function. In particular, $B(2, \nu - 1) = \Gamma(2)\Gamma(\nu - 1)/\Gamma(\nu + 1) = (\nu - 2)!/\nu!$. Putting together the pieces yields

$$
\|\nabla^2 Q^G_b(\beta^*)\|_\Omega = \sup_{u \in \mathbb{S}^{p-1}} E E_x [\mathcal{G}(-\varepsilon/b) - \tau] w^T u \leq l_{\nu-2}\kappa_{G,\nu} b^\nu/\nu!.
$$

Turning to the Hessian, note that

$$
\|\nabla^2 Q^G_b(\beta^*) - D_0\|_\Omega = \left\| \mathbb{E} \int_{-\infty}^{\infty} G(u)[f_{\varepsilon|x}(bu) - f_{\varepsilon|x}(0)] \, du \, w w^T \right\|_2.
$$

Applying a similar Taylor expansion as above, we have

$$
f_{\varepsilon|x}(t) = f_{\varepsilon|x}(0) + \sum_{\ell=1}^{\nu-1} f^{(\ell)}_{\varepsilon|x}(0) t^\ell/\ell! + \int_{0}^{1} (1 - w)^{\nu - 2} f^{(\nu - 1)}_{\varepsilon|x}(tw) - f^{(\nu - 1)}_{\varepsilon|x}(0) \, dw. \quad (C.7)
$$

Under Conditions 1 and 2, it follows that

$$
\|\nabla^2 Q^G_b(\beta^*) - D_0\|_\Omega
\leq \frac{b^{\nu-1}}{(\nu - 2)!} \sup_{u, \delta \in \mathbb{S}^{p-1}} \mathbb{E} \int_{-\infty}^{\infty} |u'^{-1}G(u)(1 - w)^{\nu - 2} f^{(\nu - 1)}_{\varepsilon|x}(buw) - f^{(\nu - 1)}_{\varepsilon|x}(0)] \, dw \, du \langle w, u \rangle \langle w, \delta \rangle
\leq \frac{b^{\nu-1}}{(\nu - 1)!} \sup_{u, \delta \in \mathbb{S}^{p-1}} \mathbb{E} \int_{-\infty}^{\infty} |u'^{-1}G(u)| \sup_{|t| \leq |u|} \left| f^{(\nu - 1)}_{\varepsilon|x}(t) - f^{(\nu - 1)}_{\varepsilon|x}(0) \right| \, du \cdot \langle |w, u| \rangle \langle w, \delta \rangle
\leq \frac{C_G b^{\nu-1}}{(\nu - 1)!} \sup_{u \in \mathbb{S}^{p-1}} \mathbb{E} \langle w, u \rangle^2 = \frac{C_G}{(\nu - 1)!} b^{\nu-1}.
$$

This completes the proof. \qed
C.2 Proof of Proposition C.2

Consider the change of variable \( \delta = \Sigma^{1/2}(\beta - \beta^*) \), so that \( \beta \in \Theta^*(r) \) is equivalent to \( \delta \in \mathbb{B}^r(r) \). Write \( w_i = \Sigma^{-1/2}x_i \in \mathbb{R}^p \), which are isotropic random vectors, and define

\[
\mathbf{H}_n(\delta) = \Sigma^{-1/2} \nabla^2 \tilde{Q}_b^G(\beta) \Sigma^{-1/2} = \frac{1}{n} \sum_{i=1}^{n} G_b(\varepsilon_i - w_i^* \delta) w_i w_i^*, \quad \mathbf{H}(\delta) = \mathbb{E}[\mathbf{H}_n(\delta)]. \tag{C.8}
\]

For any \( \epsilon_1 \in (0, r) \), there exists an \( \epsilon \)-net \( \mathcal{N}_1 := \{ \delta_1, \ldots, \delta_d \} \) with \( d_1 \leq (1 + 2r/\epsilon_1)^p \) satisfying that, for every \( \delta \in \mathbb{B}^r(r) \), there exists some \( 1 \leq j \leq d_1 \) such that \( \| \delta - \delta_j \|_2 \leq \epsilon_1 \). Hence,

\[
\| \mathbf{H}_n(\delta) - \mathbf{H}(\delta) \|_2 \\
\leq \| \mathbf{H}_n(\delta) - \mathbf{H}_n(\delta_j) \|_2 + \| \mathbf{H}_n(\delta_j) - \mathbf{H}(\delta_j) \|_2 + \| \mathbf{H}(\delta_j) - \mathbf{H}(\delta) \|_2 \\
=: I_1(\delta) + I_2(\delta_j) + I_3(\delta).
\]

For \( I_1(\delta) \), note that \( G_b(u) = (1/b)G(u/b) \) is Lipschitz continuous, i.e. \( |G_b(u) - G_b(v)| \leq l_b b^{-2}|u - v| \) for all \( u, v \in \mathbb{R} \). It follows that

\[
I_1(\delta) \leq \sup_{u, v \in \mathbb{S}^{p-1}} \frac{1}{n} \sum_{i=1}^{n} \left| G_b(\varepsilon_i - w_i^* \delta) - G_b(\varepsilon_i - w_i^* \delta_j) \right| \cdot |w_i^* u \cdot w_i^* v| \\
\leq l_b b^{-2} \sup_{u, v \in \mathbb{S}^{p-1}} \frac{1}{n} \sum_{i=1}^{n} |w_i^* (\delta - \delta_j) \cdot w_i^* u \cdot w_i^* v| \\
\leq l_b b^{-2} \epsilon_1 \sup_{u \in \mathbb{S}^{p-1}} \frac{1}{n} \sum_{i=1}^{n} |w_i^* u|^3. \tag{C.9}
\]

Next, we use the standard covering argument to bound \( M_{n,3} \). Given \( \epsilon_2 \in (0, 1) \), let \( \mathcal{N}_2 \) be an \( \epsilon_2 \)-net of the unit sphere \( \mathbb{S}^{p-1} \) with \( d_2 := |\mathcal{N}_2| \leq (1 + 2/\epsilon_2)^p \) such that for every \( u \in \mathbb{S}^{p-1} \), there exists some \( v \in \mathcal{N}_2 \) satisfying \( \| u - v \|_2 \leq \epsilon_2 \). Define the (standardized) design matrix \( \mathbf{W}_n = n^{-1/3}(w_1, \ldots, w_n)^T \in \mathbb{R}^{n \times p} \), so that \( M_{n,3} = \sup_{u \in \mathbb{S}^{p-1}} \| \mathbf{W}_n u \|_2^3 \). By the triangle inequality,

\[
\| \mathbf{W}_n u \|_3 \leq \| \mathbf{W}_n v \|_3 + \| \mathbf{W}_n(u - v) \|_3 \\
= \| \mathbf{W}_n \delta \|_3 + \left( \frac{1}{n} \sum_{i=1}^{n} |w_i^*(u - v)|^3 \right)^{1/3} \leq \| \mathbf{W}_n v \|_3 + \epsilon_2 \cdot M_{n,3}^{1/3}.
\]

30
Taking the maximum over $v \in \mathcal{N}_2$, and then taking the supremum over $u \in \mathbb{S}^{p-1}$, we arrive at

$$M_{n,3} \leq (1 - \epsilon_2)^{-3} \cdot N_{n,3} := (1 - \epsilon_2)^{-3} \cdot \max_{v \in \mathcal{N}_2} \frac{1}{n} \sum_{i=1}^{n} |w_i^T v|^3. \quad (C.10)$$

For every $v \in \mathcal{N}_2$, note that

$$\mathbb{E}||w_i^T v||^3/(6^{3/2} \nu_3)^{3/2} = 1 + \int_0^\infty e^u P(|w_i^T v| \geq \nu_1 (6u)^{1/2}) du \leq 1 + 2 \int_0^\infty e^{-2u} du = 2,$$

implying $||w_i^T v||_3 \leq 6^{3/2} \nu_3$. Hence, by inequality (3.6) in Adamczak et al. (2011) with $s = 2/3$, we obtain that for any $z \geq 3$,

$$\frac{1}{n} \sum_{i=1}^{n} |w_i^T v|^3 \leq \mathbb{E}|w_i^T v|^3 + C_1 \left( \sqrt{\frac{z}{n}} + \frac{z^{3/2}}{n} \right)$$

with probability at least $1 - e^{-z}$, where $C_1 > 0$ is a constant depending only on $\nu_1$. Taking the union bound over all vectors $v$ in $\mathcal{N}_2$ yields that, with probability at least $1 - d_2 e^{-z} \geq 1 - e^{p \log(1 + 2/\epsilon_2) - z}$,

$$N_{n,3} \leq m_3 + C_1 \left( \sqrt{\frac{z}{n}} + \frac{z^{3/2}}{n} \right)$$

where $m_3 = \sup_{u \in \mathbb{S}^{p-1}} \mathbb{E}|w^T u|^3$. Reorganizing the terms, we get

$$N_{n,3} \leq m_3 + C_1 \left( \sqrt{\frac{p \log(1 + 2/\epsilon_2) + \log 2 + t}{n}} + \frac{p \log(1 + 2/\epsilon_2) + \log 2 + t^{3/2}}{n} \right) \quad (C.11)$$

with probability at least $1 - e^{-t}/2$. Taking $\epsilon_2 = 1/8$ in (C.10) and (C.11) implies

$$M_{n,3} \leq 1.5m_3 + 1.5C_1 \left( \sqrt{\frac{3p + 1 + t}{n}} + \frac{(3p + 1 + t)^{3/2}}{n} \right).$$

Under the sample size scaling $n \gtrsim p + t$, plugging the above bound into (C.9) yields

$$\sup_{\delta \in \mathbb{B}^p(r)} I_1(\delta) \lesssim (p + t)^{1/2} b^{-2} \epsilon_1 \quad (C.12)$$

with probability at least $1 - e^{-t}/2$. For $I_3(\delta)$, it can be similarly obtained that

$$I_3(\delta) \leq lGb^{-2} \sup_{u, v \in \mathbb{S}^{p-1}} \mathbb{E}|w_i^T (\delta - \delta_j) \cdot w_i^T u \cdot w_i^T v| \leq lGb^{-2} \epsilon_1 \quad (C.13)$$

uniformly over all $\delta \in \mathbb{B}^p(r)$. 

31
Turning to $I_2(\delta_j)$, note that $H_u(\delta_j) - H(\delta_j) = (1/n) \sum_{i=1}^n (1 - E) \phi_{ij} w_i w_i^T$, where $\phi_{ij} = G_b(e_i - u_i^T \delta_j)$ satisfy $|\phi_{ij}| \leq \kappa_{G,b} b^{-1}$ and

$$\mathbb{E}(\phi_{ij}^2 | x_i) = \frac{1}{b^2} \int_{-\infty}^{\infty} G^2\left(\frac{w_i \delta - t}{b}\right) f_{e_i|x_i}(t) \, dt = \frac{1}{b} \int_{-\infty}^{\infty} G^2(u) f_{e_i|x_i}(u_i^T \delta - bu) \, du \leq \frac{\tilde{f} \kappa_{G^2,0}}{b}$$

almost surely, where $\kappa_{G^2,0} := \int_{-\infty}^{\infty} G^2(u) \, du < \infty$. Given $\epsilon_3 \in (0, 1/2)$, there exists an $\epsilon_3$-net $N_3$ of the sphere $S^{p-1}$ with $|N_3| \leq (1 + 2/\epsilon_3)^p$ such that $\|H_u(\delta_j) - H(\delta_j)\|_2 \leq (1 - 2\epsilon_3)^{-1} \max_{u \in M} \|u^T(H_u(\delta_j) - H(\delta_j))u\|$. Given $u \in N_3$ and $k = 2, 3, \ldots$, we bound the higher order moments of $\phi_{ij}(w_i^T u)^2$ by

$$\mathbb{E}|\phi_{ij}(w_i^T u)^2|^k \leq \tilde{f} \kappa_{G^2,0} b^{-1} \cdot (\kappa_{G,b} b^{-1})^{k-2} v_1^{2k} \cdot 2k \int_{0}^{\infty} \mathbb{P}(\|u_i^T u\| \geq \nu_i) u_i^{2k-1} \, du$$

$$\leq \tilde{f} \kappa_{G^2,0} b^{-1} \cdot (\kappa_{G,b} b^{-1})^{k-2} v_1^{2k} \cdot 4k \int_{0}^{\infty} u^{2k-1} e^{-u^2/2} \, du$$

$$\leq \tilde{f} \kappa_{G^2,0} b^{-1} \cdot (\kappa_{G,b} b^{-1})^{k-2} v_1^{2k} \cdot 2^{k+1} k!.$$

In particular, $\mathbb{E}\phi_{ij}^2(w_i^T u)^4 \leq (4v_1^2)^2 \tilde{f} \kappa_{G^2,0} b^{-1}$, and $\mathbb{E}\phi_{ij}(w_i^T u)^2 \leq \frac{b^2}{2} (4v_1^2)^2 \tilde{f} \kappa_{G^2,0} b^{-1} \cdot (2v_1^2 \kappa_{G,b} b^{-1})^{k-2}$ for $k \geq 3$. Applying Bernstein’s inequality and the union bound, we find that for any $u \geq 0$,

$$\|H_u(\delta_j) - H(\delta_j)\|_2$$

$$\leq \frac{1}{1 - 2\epsilon_3} \max_{u \in M} \left|\sum_{i=1}^n (1 - E) \phi_{ij}(w_i^T u)^2\right| \leq \frac{2\nu_1}{1 - 2\epsilon_3} \left(2 \sqrt{2 \tilde{f} \kappa_{G^2,0} \frac{u}{nb}} + \kappa_{G,b} \frac{u}{nb}\right)$$

with probability at least $1 - 2(1 + 2/\epsilon_3)^p e^{-u} = 1 - (1/2)^{p \log(4) + p \log(1+2/\epsilon_3) - u}$. Setting $\epsilon_3 = 2/(e^3 - 1)$ and $u = \log(4) + 3p + v$, it follows that with probability at least $1 - e^{-v}/2$,

$$I_2(\delta_j) \leq \sqrt{\frac{p + v}{nb} + \frac{p + v}{nb}}.$$

Once again, taking the union bound over $j = 1, \ldots, d_1$ and setting $v = p \log(1 + 2r/\epsilon_i) + t$, we obtain that with probability at least $1 - d_1 e^{-v} \geq 1 - e^{-t}/2$,

$$\max_{1 \leq j \leq d_1} I_2(\delta_j) \leq \sqrt{\frac{p \log(3er/\epsilon_1) + t}{nb} + \frac{p \log(3er/\epsilon_1) + t}{nb}}. \quad (C.14)$$

Finally, combining (C.12), (C.13) and (C.14), and taking $\epsilon_1 = r/n \in (0, r)$ in the beginning of
the proof, we conclude that with probability at least $1 - e^{-t}$,

$$\sup_{\beta \in \Theta^*(r)} \| \nabla^2 \tilde{Q}_G^b(\beta) - \nabla^2 Q_G^b(\beta) \|_\Omega \lesssim \sqrt{p \log n + t \frac{1}{nb}} + \frac{p \log n + t}{nb} + \frac{(p + t)^{1/2}}{nb^2}$$

as long as $n \gtrsim p + t$. This completes the proof. □

C.3 Proof of Proposition C.3

Define the stochastic process $\Delta_b(\beta) = \Sigma^{-1/2} (\nabla^2 \tilde{Q}_G^b(\beta) - \nabla^2 Q_G^b(\beta^*).$ By the triangle inequality,

$$\sup_{\beta \in \Theta^*(r)} \| \Delta_b(\beta) \|_2 \leq \sup_{\beta \in \Theta^*(r)} \| \mathbb{E} \Delta_b(\beta) \|_2 + \sup_{\beta \in \Theta^*(r)} \| \Delta_b(\beta) - \mathbb{E} \Delta_b(\beta) \|_2$$

Recall that $D_0 = \Sigma^{-1/2} \mathbf{D} \Sigma^{-1/2} = \mathbb{E} [\hat{f}_{el}(0) x w^T]$. For the first term on the right-hand side, using the mean value theorem for vector-valued functions yields

$$\mathbb{E} \Delta_b(\beta) = \left\{ \Sigma^{-1/2} \int_0^1 \nabla^2 Q_G^b((1 - s)\beta^* + s\beta) \mathbf{d}s \Sigma^{-1/2} - D_0 \right\} \Sigma^{1/2}(\beta - \beta^*).$$

By a change of variable $\delta = \Sigma^{1/2}(\beta - \beta^*)$,

$$\nabla^2 Q_G^b((1 - s)\beta^* + s\beta) = \mathbb{E} \int_{-\infty}^{\infty} G(u)\hat{f}_{el}(sw^T \delta - bu) \mathbf{d}u \cdot xx^T.$$

For every $s \in [0, 1]$ and $u \in \mathbb{R}$, it ensures from $\hat{f}_{el}(\cdot)$ being Lipschitz that

$$|\hat{f}_{el}(sw^T \delta - bu) - \hat{f}_{el}(wu)| \leq l_0 s \cdot |w^T \delta|.$$

Moreover, by the Taylor expansion (C.7),

$$\hat{f}_{el}(wu) = \hat{f}_{el}(0) + \sum_{l=1}^{\nu-1} \frac{\hat{f}_{el}^{(l)}(0)(-bu)^l}{l!} + \frac{(-bu)^{\nu-1}}{(\nu - 2)!} \int_0^1 (1 - w)^{\nu-2} \left( f_{el}^{(\nu-1)}(-bu) - f_{el}^{(\nu-1)}(0) \right) \mathbf{d}w.$$
Consequently,

\[
\left\| \Sigma^{-1/2} \int_0^1 \nabla^2 Q^G_b((1-s)\beta^* + s\beta) \, ds \Sigma^{-1/2} - D_0 \right\|_2 \\
\leq \left\| \frac{b^{v-1}}{(v-2)!} \int_{-\infty}^\infty \int_0^1 u^{v-1} G(u)(1-w)^{v-2} \{ f^{(v-1)}(-buw) - f^{(v-1)}(0) \} \, du \right\|_2 \\
+ 0.5l_0 \cdot \left\| \mathbb{E} \right\|_{w^T \delta \cdot w^T} \\
\leq \frac{C_G b^{v-1}}{(v-1)!} \sup_{u \in \mathbb{S}^{p-1}} \mathbb{E}(w, u)^2 + \frac{l_0}{2} \sup_{u \in \mathbb{S}^{p-1}} \mathbb{E}(w^T \delta)(w, u)^2 \\
\leq \left( \frac{C_G}{(v-1)!} b^{v-1} + 0.5l_0 \|\delta\|_2 \right).
\]

Taking the supremum over \( \beta \in \Theta^*(r) \), or equivalently \( \delta \in \mathbb{B}(r) \), yields

\[
\sup_{\beta \in \Theta^*(r)} \left\| \mathbb{E} \Delta_b(\beta) \right\| \leq \left( \frac{C_G}{(v-1)!} b^{v-1} + 0.5l_0 m_3 \|\delta\|_2 \right) r \leq (b^{v-1} + r)r.
\]

For the stochastic term \( \sup_{\beta \in \Theta^*(r)} \|\Delta_b(\beta) - \mathbb{E}\Delta_b(\beta)\|_2 \), following the proof of Theorem 3.2, it can be similarly shown that with probability at least \( 1 - e^{-t} \),

\[
\sup_{\beta \in \Theta^*(r)} \|\Delta_b(\beta) - \mathbb{E}\Delta_b(\beta)\|_2 \leq r \sqrt{\frac{p + t}{nb}}
\]

as long as \( \sqrt{(p + t)/n} \leq b \).

Combining the last two displays completes the proof of (C.3). \( \square \)

**C.4 Proof of Theorem C.1**

**Step 1** (Consistency of the initial estimate). First, note that the consistency of the initial estimator \( \beta^* \)—namely, \( \beta^* \) lies in a local neighborhood of \( \beta^* \) with high probability, is a direct consequence of Theorem 3.1. Given a non-negative kernel \( K(\cdot) \) and for any \( t > 0 \), the initial estimator \( \beta^* \) satisfies

\[
\|\beta^* - \beta^*\|_\Sigma \leq r_n \leq \sqrt{\frac{p + t}{n}} \quad (C.15)
\]
with probability at least $1 - 2e^{-t}$ as long as $(\frac{p+1}{n})^{1/2} \leq h \leq (\frac{p+1}{n})^{1/4}$. Let $\mathcal{E}_\text{min}(t)$ be the event that (C.15) holds. Provided that the sample Hessian $\nabla^2 \Q_b^G(\beta)$ is invertible, we have

$$D(\tilde{\beta} - \beta) = -D[\nabla^2 \Q_b^G(\beta)]^{-1} \nabla \Q_b^G(\beta)$$

$$= -D[\nabla^2 \Q_b^G(\beta)]^{-1} \Sigma^{1/2} \cdot \Sigma^{-1/2} \{\nabla \Q_b^G(\beta) - \nabla \Q_b^G(\beta^*) - D(\beta - \beta^*)\}$$

$$- D[\nabla^2 \Q_b^G(\beta)]^{-1} \Sigma^{1/2} \cdot \Sigma^{-1/2} \{D(\beta - \beta^*) + \nabla \Q_b^G(\beta^*)\},$$

or equivalently,

$$\Sigma^{-1/2} D(\beta - \beta^*) = -D_0 \tilde{D}_0^{-1} \cdot \Sigma^{-1/2} \{\nabla \Q_b^G(\beta) - \nabla \Q_b^G(\beta^*) - D(\beta - \beta^*)\}$$

$$+ (I_p - D_0 \tilde{D}_0^{-1}) D_0 \cdot \Sigma^{1/2} \{\beta - \beta^*\} - D_0 \tilde{D}_0^{-1} \cdot \Sigma^{-1/2} \nabla \Q_b^G(\beta^*),$$

where $D_0 = E(f_d(x)(\omega \omega^T)) = \Sigma^{-1/2} D \Sigma^{-1/2}$ and $\tilde{D}_0 := \Sigma^{-1/2} \nabla^2 \Q_b^G(\beta) \Sigma^{-1/2}$. It follows that

$$\|D(\tilde{\beta} - \beta^*) + \nabla \Q_b^G(\beta^*)\|_\Omega$$

$$\leq \|I_p - D_0 \tilde{D}_0^{-1} D_0\|_2 \cdot \|\tilde{\beta} - \beta^*\|_\Sigma + \|I_p - D_0 \tilde{D}_0^{-1}\|_2 \cdot \|\nabla \Q_b^G(\beta^*)\|_\Omega$$

$$+ \|D_0 \tilde{D}_0^{-1}\|_2 \cdot \|\nabla \Q_b^G(\beta) - \nabla \Q_b^G(\beta^*) - D(\beta - \beta^*)\|_\Omega. \quad \text{(C.16)}$$

In view of (C.16), it remains to bound the following three quantities:

$$\|I_p - D_0 \tilde{D}_0^{-1}\|_2, \|\nabla \Q_b^G(\beta^*)\|_\Omega \text{ and } \|\nabla \Q_b^G(\beta) - \nabla \Q_b^G(\beta^*) - D(\beta - \beta^*)\|_\Omega.$$

**Step 2** (Consistency of the sample Hessian $\nabla^2 \Q_b^G(\beta)$). Recall that $\tilde{D}_0 = \Sigma^{-1/2} \nabla^2 \Q_b^G(\beta) \Sigma^{-1/2}$ and $D_0 = \Sigma^{-1/2} D \Sigma^{-1/2}$. By the triangle inequality,

$$\|\tilde{D}_0 - D_0\|_2 \leq \|\nabla^2 \Q_b^G(\beta) - \nabla^2 \Q_b^G(\beta^*)\|_\Omega + \|\nabla^2 \Q_b^G(\beta^*) - D\|_\Omega.$$

Let the bandwidth $b$ satisfy $\max\{\frac{p \log n + t}{n}, n^{-1/2}\} \leq b \leq 1$. Conditioned on $\mathcal{E}_\text{min}(t)$, applying Propositions C.1 and C.2 with $r = r_n$ yields that, with probability $1 - e^{-t}$,

$$\|\tilde{D}_0 - D_0\|_2 \leq \delta_n \times \sqrt{\frac{p \log n + t}{nb}} + b^{r-1}.$$

Under Condition 3.2, $0 < f \leq \lambda_{\text{min}}(D_0) \leq \lambda_{\text{max}}(D_0) \leq \bar{f}$, so that $\|D_0^{-1}\|_2 \leq \bar{f}^{-1}$. For sufficiently
large $n$ and small $b$, this implies $\|\tilde{D}_0 D_0^{-1} - I_p\|_2 \leq \frac{\delta_n}{\ell - \delta_n} < 1$, and hence

$$\|D_0 \tilde{D}_0^{-1} - I_p\|_2 \leq \frac{\delta_n}{\ell - \delta_n}, \quad \|D_0 \tilde{D}_0^{-1}\|_2 \leq \frac{\ell}{\ell - \delta_n}. \quad (C.17)$$

**Step 3 (Controlling the scores).** For $\|\nabla \hat{Q}_b^G (\beta^*)\|_\Omega$, it follows from Lemma B.1 and Proposition C.1 that with probability at least $1 - e^{-t}$,

$$\|\nabla \hat{Q}_b^G (\beta^*)\|_\Omega \lesssim \sqrt{\frac{p + t}{n} + b^\nu}. \quad (C.18)$$

Turning to $\|\nabla \hat{Q}_b^G (\beta) - \nabla \hat{Q}_b^G (\beta^*) - D(\beta - \beta^*)\|_\Omega$, applying the concentration bound (C.15) and Proposition C.3 we obtain that, with probability at least $1 - e^{-t}$ conditioned on $\mathcal{E}_{\text{init}} (t)$,

$$\|\nabla \hat{Q}_b^G (\beta) - \nabla \hat{Q}_b^G (\beta^*) - D(\beta - \beta^*)\|_\Omega \lesssim \frac{p + t}{nb^{1/2}} + b^{\nu - 1} \sqrt{\frac{p + t}{n}} \quad (C.19)$$

as long as $\left(\frac{p + t}{n}\right)^{1/2} \leq b \leq 1$.

Finally, combining the bounds (C.15)–(C.19), we conclude that with probability at least $1 - 5e^{-t}$,

$$\|D(\hat{\beta} - \beta^*) + \nabla \hat{Q}_b^G (\beta^*)\|_\Omega \lesssim \left(\sqrt{\frac{p \log n + t}{nb}} + b^{\nu - 1}\right) \left(\sqrt{\frac{p + t}{n}} + b^\nu\right),$$

provided that max$\left[\frac{p \log n + t}{n}, \left(\frac{p + t}{n}\right)^{1/2}\right] \leq b \leq 1$. Under the sample size scaling $n \gtrsim p (\log n)^2 + t$ and bandwidth constraint $b \lesssim \left(\frac{p + t}{n}\right)^{1/(2\nu)}$, this leads to the claimed bound (C.4). \hfill \Box

**D Additional Simulation Results**

In this section, we present additional results for the numerical studies as described in Section 5 in Figures D.1–D.5.
References


Figure D.1: Results under models (5.1)–(5.3) in Section 5 with \( \tau = [0.1, 0.3, 0.5, 0.7] \) and \( t_2 \) noise, averaged over 100 simulations. This figure extends the last row of Figure 3 to other quantile levels.
Figure D.2: Elapsed time of standard QR, Horowitz’s smoothing and conquer under $t_2$-noise setting when $\tau \in \{0.1, 0.3, 0.5, 0.7\}$. This figure extends the last row of Figure 4 to other quantile levels.
Figure D.3: Empirical coverage, confidence interval width, and elapsed time of six methods with $N(0,4)$ errors under $\tau = 0.9$. Other details are as in Figure 6.
Figure D.4: Empirical coverage, confidence interval width, and elapsed time of six methods with $t_2$ errors under $\tau = 0.5$. Other details are as in Figure 6.
Figure D.5: Empirical coverage, confidence interval width, and elapsed time of six methods with $\mathcal{N}(0, 4)$ errors under $\tau = 0.5$. Other details are as in Figure 6.