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 $\beta \mapsto \left( \frac{1}{n} \sum_{i=1}^{n} (\rho_{\tau} * K_h)(y_i - \langle x_i, \beta \rangle) \right)$ is not only twice-differentiable and convex, but also (provably) strongly convex in a local vicinity of $\beta^\ast$ 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 $\beta \mapsto \tilde{Q}_b^G(\beta) := \left( \frac{1}{n} \sum_{i=1}^{n} (\rho_{\tau} * G_b)(y_i - \langle x_i, \beta \rangle) \right)$. 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 $\|\nabla \widehat{Q}_b^G(\beta)\|_2 \leq \epsilon$, where $\epsilon \geq 0$ is optimization error. However, the equation $\nabla \widehat{Q}_b^G(\beta) = 0$ 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 $\widehat{\beta} \in \text{argmin}_{\beta \in \mathbb{R}^p} \widehat{Q}_h^K(\beta)$, where $\widehat{Q}_h^K(\beta) = (1/n) \sum_{i=1}^n (\rho \ast K_h)(y_i - \langle x_i, \beta \rangle)$. Denote by $\bar{r}_i = y_i - \langle x_i, \beta \rangle$ for $i = 1, \ldots, n$ the fitted residuals. Next, with slight abuse of notation, we define the one-step conquer estimator $\widehat{\beta}$ as a solution to the equation $\nabla^2 \widehat{Q}_b^G(\beta)(\widehat{\beta} - \beta) = -\nabla \widehat{Q}_b^G(\beta)$, or equivalently,

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

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

$$
\widehat{\beta} = \beta - \left( \nabla^2 \widehat{Q}_b^G(\beta) \right)^{-1} \nabla \widehat{Q}_b^G(\beta). \tag{A.2}
$$

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

Theoretical properties of the one-step estimator $\widehat{\beta}$ 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

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

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

$$
\widehat{G}_{2r}(v) = \int_{-\infty}^{v} G_{2r}(u) \, du = \sum_{\ell=0}^{r-1} \frac{(-1)^\ell}{2^\ell \ell!} \phi^{2\ell-1}(v). \tag{2}
$$
In fact, both \( G_2 \) and \( \overline{G}_2 \) have simpler forms \( G_2(u) = p_r(u)\phi(u) \) and \( \overline{G}_2(u) = \Phi(u) + P_r(u)\phi(u) \), 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**

**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 \( \Theta^* = \{ \beta \in \mathbb{R}^p : \beta - \beta^* \in \Theta(\kappa_r^{1/2}h) \} \). Let \( \eta = \sup \{ u | [0, 1] : u(\beta^*_h - \beta^*) \in \Theta(\kappa_r^{1/2}h) \} \) 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

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

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

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

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

\[
\mathbb{E}[K_h(y - \langle x, \beta \rangle)x] = f_{e|x}(0) + R_h(\delta) \tag{B.3} \]
with $R_h(\delta)$ satisfying $|R_h(\delta)| \leq l_0(|\langle x, \delta \rangle| + \kappa_1 h)$. Substituting (B.3) into (B.1) and (B.2) yields

$$
\langle \nabla Q_h(\overline{\beta}^s) - \nabla Q_h(\beta^*) - \beta^* \rangle
\geq ||\overline{\beta}^s - \beta^*||_D^2 - \frac{l_0}{2} \mathbb{E}[\langle x, \overline{\beta}^s - \beta^* \rangle]^3 - l_0 \kappa_1 h \cdot ||\overline{\beta}^s - \beta^*||^2_{\Sigma}
\geq ||\overline{\beta}^s - \beta^*||_D^2 - \frac{l_0}{2} \mu_1 \cdot ||\overline{\beta}^s - \beta^*||^2_{\Sigma} - l_0 \kappa_1 h \cdot ||\overline{\beta}^s - \beta^*||^2_{\Sigma}.
$$

(B.4)

On the other hand, under model (3.1) we have

$$
\langle -\nabla Q_h(\beta^*), \overline{\beta}^s - \beta^* \rangle \leq \|\Sigma^{-1/2} \nabla Q_h(\beta^*)\|_2 \cdot ||\overline{\beta}^s - \beta^*||_{\Sigma},
$$

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

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

from which it follows that $|\mathbb{E}[\overline{\beta}(-\epsilon/h)|x] - \tau| \leq \frac{l_0}{2} \kappa_2 h^2$. Consequently,

$$
\|\Sigma^{-1/2} \nabla Q_h(\beta^*)\|_2 = \sup_{u \in \mathbb{R}^{p-1}} \mathbb{E}[\overline{\beta}(-\epsilon/h) - \tau|\Sigma^{-1/2}x, u] \leq \frac{l_0}{2} \kappa_2 h^2.
$$

(B.5)

Putting together the pieces, we conclude that

$$
\langle -\nabla Q_h(\beta^*), \overline{\beta}^s - \beta^* \rangle \leq \frac{l_0}{2} \kappa_2 h^2 \cdot ||\overline{\beta}^s - \beta^*||_{\Sigma}.
$$

(B.6)

Recall that $f_{\epsilon|x}(0) \geq \underline{f} > 0$ almost surely. Combining (B.1) and (B.4) with (B.6) yields

$$
\underline{f} \cdot ||\overline{\beta}^s - \beta^*||^2_{\Sigma} \leq \underline{f}^{1/2} \cdot ||\overline{\beta}^s - \beta^*||_D \cdot ||\overline{\beta}^s - \beta^*||_{\Sigma}
\leq \left( \frac{\mu_3 + 1}{2} \kappa_2 + \kappa_1 \kappa_2^{1/2} \right) l_0 h^2 \cdot ||\overline{\beta}^s - \beta^*||_{\Sigma}.
$$
Canceling $||\tilde{\beta}^* - \beta^*||_\Sigma$ on both sides we obtain

$$||\tilde{\beta}^* - \beta^*||_\Sigma \leq \left( \frac{\mu_3 + \frac{1}{2}h}{2} + \kappa_1 \frac{h}{L} k_2^{1/2} \right) \frac{l_0 \kappa h}{L} k_2^{1/2} h.$$  

Provided that $h < \frac{L}{(c_k l_0)}$, $\tilde{\beta}^*$ falls in the interior of $\Theta^*$, i.e., $||\tilde{\beta}^* - \beta^*||_\Sigma < \kappa_1^{1/2} h$, enforcing $\eta = 1$ (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

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

Again, by the mean value theorem for vector-valued functions,

$$\Delta = \left\{ D^{-1/2} \int_0^1 \nabla^2 Q_h((1 - t)\beta^* + t \beta_h^*) \, dt \, D^{-1/2} - I_p \right\} D^{1/2}(\beta_h^* - \beta^*). \quad (B.7)$$

Write $\overline{w} = D^{-1/2} x$, we have

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

$$= \left|\left| E \int_0^1 \int_{-\infty}^{\infty} K(u)(f_{\varepsilon|x}(t(x, \beta_h^* - \beta^*) - h u) - f_{\varepsilon|x}(0)) \, du \, dt \, \overline{w} \overline{w} \right|\right|_2$$

$$\leq l_0 \sup_{u \in \mathbb{S}^{p-1}} E \int_0^1 \int_{-\infty}^{\infty} K(u)|\langle t(x, \beta_h^* - \beta^*) + h |u| \rangle| \, du \, dt \, \langle \overline{w}, u \rangle^2$$

$$\leq \frac{l_0}{2} \sup_{u \in \mathbb{S}^{p-1}} E(|\langle x, \beta_h^* - \beta^* \rangle| + \langle \overline{w}, u \rangle)^2 + \frac{l_0}{L} \kappa_1 h$$

$$\leq \frac{l_0}{2} \mu_3 ||\beta_h^* - \beta^*||_\Sigma + \frac{l_0}{L} \kappa_1 h.$$  

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

$$||\Delta||_2 \leq \frac{l_0}{L} (0.5 \mu_3 ||\beta_h^* - \beta^*||_\Sigma + \kappa_1 h) ||\beta_h^* - \beta^*||_D. \quad (B.8)$$
Moreover, applying a second-order Taylor series expansion to \( f_{\hat{e}a} \) yields

\[
\mathbb{E}[\bar{K}(-\varepsilon/h|x) - \tau = \int_{-\infty}^{\infty} K(u) \int_0^{hu} [f_{\hat{e}a}(t) - f_{\hat{e}a}(0)] \, dt \, du \\
= \frac{1}{2} \kappa_2 h^2 \cdot f'_{\hat{e}a}(0) + \int_{-\infty}^{\infty} \int_0^{hu} K(u) [f'_{\hat{e}a}(v) - f'_{\hat{e}a}(0)] \, dv \, dt \, du.
\]

Recalling that \( \nabla Q_h(\beta^*) = \mathbb{E}[\bar{K}(-\varepsilon/h) - \tau|x] \), we get

\[
\left\| D^{-1} \nabla Q_h(\beta^*) - \frac{1}{2} \kappa_2 h^2 \cdot D^{-1} \mathbb{E}[f'_{\hat{e}a}(0)x] \right\|_D \leq \frac{l_1}{6} \ell^{1/2} \kappa_3 h^3. \tag{B.9}
\]

Finally, combining (B.8) and (B.9) proves (3.4).

\[ \square \]

### B.2 Proof of Theorem 3.1

Recall from (2.7) 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 \( h > 0 \), it follows from the optimality of \( \hat{\beta}_h \) that \( \nabla \hat{Q}_h(\hat{\beta}_h) = 0 \) and by convexity, \( \langle \nabla \hat{Q}_h(\hat{\beta}_h) - \nabla \hat{Q}_h(\beta^*), \hat{\beta}_h - \beta^* \rangle \geq 0 \).

Recall from the proof of Proposition 3.1 that \( \Theta(t) = \{ u \in \mathbb{R}^p : \|u\|_\Sigma \leq t \} \) for \( t \geq 0 \). For some \( r > 0 \) to be determined, let \( \eta = \sup\{ u \in [0, 1] : u(\hat{\beta}_h - \beta^*) \in \Theta(r) \} \) and \( \bar{\beta} = \beta^* + \eta(\hat{\beta}_h - \beta^*) \). 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, \( \bar{\beta} \in \beta^* + \partial \Theta(r) \), where \( \partial \Theta(r) \) is the boundary of \( \Theta(r) \). The symmetrized Bregman divergence associated with \( \hat{Q}_h(\cdot) \) for points \( \beta_1, \beta_2 \) is given by

\[
D(\beta_1, \beta_2) = \langle \nabla \hat{Q}_h(\beta_1) - \nabla \hat{Q}_h(\beta_2), \beta_1 - \beta_2 \rangle. \tag{B.10}
\]

By Lemma C.1 in Sun, Zhou and Fan (2020), the three points \( \hat{\beta}_h, \bar{\beta}, \) and \( \beta^* \) satisfy \( D(\bar{\beta}, \beta^*) \leq \eta D(\bar{\beta}, \beta^*) \). Together with the properties that \( \bar{\beta} - \beta^* = \eta(\hat{\beta}_h - \beta^*) \in \Theta(r) \) and \( \nabla \hat{Q}_h(\hat{\beta}_h) = 0 \), we obtain

\[
\left\| \bar{\beta} - \beta^* \right\|_\Sigma^2 - \frac{D(\bar{\beta}, \beta^*)}{\left\| \beta - \beta^* \right\|_\Sigma^2} \leq -\eta \langle \nabla \hat{Q}_h(\beta^*), \hat{\beta}_h - \beta^* \rangle \leq \left\| \Sigma^{-1/2} \nabla \hat{Q}_h(\beta^*) \right\|_2 \cdot \left\| \bar{\beta} - \beta^* \right\|_\Sigma.
\]
Canceling \( \| \tilde{\beta} - \beta^* \|_\Sigma \) on both sides, we obtain
\[
\| \tilde{\beta} - \beta^* \|_\Sigma \leq \frac{\| \Sigma^{-1/2} \nabla \tilde{Q}_h(\beta^*) \|_2}{\inf_{\beta \in \beta^* + \Theta(r)} \{ D(\beta, \beta^*) / \| \beta - \beta^* \|_2^2 \}}.
\] (B.11)

The following two lemmas provide, respectively, upper and lower bounds on \( \| \Sigma^{-1/2} \nabla \tilde{Q}_h(\beta^*) \|_2 \) and \( \inf_{\beta \in \beta^* + \Theta(r)} \{ D(\beta, \beta^*) / \| \beta - \beta^* \|_2^2 \} \).

**Lemma B.1.** Assume that Conditions 3.1–3.4 hold. For any \( t \geq 0 \),
\[
\| \Sigma^{-1/2} \nabla \tilde{Q}_h(\beta^*) \|_2 \leq 1.46 \nu_0 \left\{ C_\tau^{1/2} \sqrt{\frac{4p + 2r}{n}} + 2 \max(1 - \tau, \tau) \frac{2p + t}{n} \right\} + \frac{1}{2} \lambda_0 \kappa_2 h^2
\] (B.12)
with probability at least \( 1 - e^{-t} \), where \( C_\tau = \tau (1 - \tau) + (1 + \tau) \lambda_0 \kappa_2 h^2 \).

Under Condition 3.2, there exist constants \( \bar{f}_h \geq f_h \) such that
\[
f_h = \inf_{|u| \leq h/2} f_{\text{elx}}(u) \leq \sup_{|u| \leq h/2} f_{\text{elx}}(u) \leq \bar{f}_h
\] (B.13)
almost surely. In fact, by the Lipschitz continuity, we can take \( \bar{f}_h = f + l_0 h/2 \) and \( \bar{f}_h = f - l_0 h/2 \).

Throughout the following, we assume (B.13) holds. For every \( \delta \in (0, 1] \), we define \( \eta_\delta \geq 0 \) as
\[
\eta_\delta = \inf\{ \eta > 0 : \mathbb{E}\left[ (\langle \delta, w \rangle)^2 \mathbb{1}(|\langle \delta, w \rangle| > \eta_\delta) \right] \leq \delta \text{ for all } \delta \in \mathbb{R}^{p-1} \},
\] (B.14)
where \( w = \Sigma^{-1/2} x \) is the standardized predictor that satisfies \( \mathbb{E}(ww^\top) = I_p \). 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.** For any \( t > 0 \), \( 0 < h < 2f/l_0 \) and \( 0 < r \leq h/(4 \eta_{1/4}) \) with \( \eta_{1/4} \) defined in (B.14), we have
\[
\inf_{\beta \in \beta^* + \Theta(r)} \{ D(\beta, \beta^*) / \| \beta - \beta^* \|_2^2 \} \geq \frac{3}{4} \bar{L}_h - \bar{f}_h^{1/2} \left( \frac{5}{4} \sqrt{\frac{hp}{r^2 n}} + \sqrt{\frac{ht}{8r^2 n}} \right) - \frac{ht}{3r^2 n}
\] (B.15)
with probability at least \( 1 - e^{-t} \).
In view of (B.11), (B.12), and (B.15), we take \( r = \frac{h}{4} \) so that as long as \((p + t)/n \lesssim h \lesssim 1\),

\[
\|\tilde{\beta} - \beta^*\|_\Sigma < \frac{3\nu_0}{\kappa L} \left\{ C_1^{1/2} \left( \frac{4p + 2t}{n} + 2 \max(1 - \tau, \tau) \frac{2p + t}{n} \right) \right\} + \frac{l_0 \kappa_2}{\kappa I} h^2 \tag{B.16}
\]

with probability at least \( 1 - 2e^{-t} \). With this choice of \( r \), we see that under the constraint \( \sqrt{(p + t)/n} \lesssim h \lesssim 1 \), \( \|\tilde{\beta} - \beta^*\|_\Sigma < r \) with probability at least \( 1 - 2e^{-t} \). In other words, on an event that occurs with high probability, \( \tilde{\beta} \) falls in the interior of \( \beta^* + \Theta(r) \), enforcing \( \eta = 1 \) and \( \hat{\beta}_h = \tilde{\beta} \). The claimed bound for \( \hat{\beta}_h \) then follows immediately. \( \Box \)

### B.2.1 Proof of Lemma B.1

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

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

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

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

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

For each unit vector \( u \in \mathcal{N}_\epsilon \), define centered random variables \( \gamma_{u,i} = \langle u, \Sigma^{-1/2} [\xi_i x_i - \mathbb{E}(\xi_i x_i)] \rangle \).

We first show that \( \mathbb{E}(\gamma_{u,i}^2) \) is bounded. By a change of variable and integration by parts, it can be
shown that
\[
\mathbb{E}[\mathcal{K}^2(-\varepsilon/h|x)] = 2 \int_{-\infty}^{\infty} K(v)\mathcal{K}(v)F_{\gamma|x}(v)dv \\
= 2\tau \int_{-\infty}^{\infty} K(v)\mathcal{K}(v)dv - 2h f_{\gamma|x}(0) \int_{-\infty}^{\infty} vK(v)\mathcal{K}(v)dv \\
+ 2 \int_{-\infty}^{\infty} \int_{-\varepsilon/h}^{\infty} \{f_{\gamma|x}(t) - f_{\gamma|x}(0)\}K(v)\mathcal{K}(v)dvdt \\
\leq \tau + l_0\kappa_2 h^2,
\]

where \(\kappa_2\) and \(l_0\) are constants that appear in Conditions 3.1 and 3.2, respectively. It then follows that
\[
\mathbb{E}[\xi_i^2|x_i] \leq C_\tau := \tau(1 - \tau) + (1 + \tau)l_0\kappa_2 h^2.\]

Moreover, \(|\xi_i| \leq \max(1 - \tau, \tau)\). Hence, for \(k = 2, 3, \ldots\),
\[
\mathbb{E}(\|\mathbf{u}, \Sigma^{-1/2}\xi_i|x_i\|^k) \leq \max(1 - \tau, \tau)^{k-2} \mathbb{E}(\|(\mathbf{u}, \Sigma^{-1/2}\xi_i)\|^k \cdot \mathbb{E}(\xi_i^2|x_i)) \\
\leq C_\tau \max(1 - \tau, \tau)^{k-2} v_0^k \int_0^{\infty} \mathbb{P}(\|(\mathbf{u}, \Sigma^{-1/2}\xi_i)\| \geq v_0) k^{t-1} dt \\
\leq C_\tau \max(1 - \tau, \tau)^{k-2} v_0^k \int_0^{\infty} t^{k-1} e^{-t} dt \\
= C_\tau k! \max(1 - \tau, \tau)^{k-2} v_0^k \\
\leq \frac{k!}{2} \cdot C_\tau v_0^2 \cdot \{2 \max(1 - \tau, \tau) u_0\}^{k-2}.
\]

Consequently, it follows from Bernstein’s inequality that for every \(u \geq 0\),
\[
\frac{1}{n} \sum_{i=1}^{n} \gamma_{ui} \leq u_0 \left( C_\tau^{1/2} \sqrt{\frac{2u}{n}} + \max(1 - \tau, \tau) \frac{2u}{n} \right)
\]
with probability at least \(1 - e^{-u}\).

Finally, applying a union bound over all vectors \(\mathbf{u} \in \mathcal{N}_e\) yields
\[
\|\Sigma^{-1/2}(\nabla \bar{Q}_h(\beta^*) - \nabla Q_h(\beta^*))\|_2 \leq \frac{u_0}{1 - e} \left( C_\tau^{1/2} \sqrt{\frac{2u}{n}} + \max(1 - \tau, \tau) \frac{2u}{n} \right)
\]
with probability at least \(1 - e^{\log((1 + 2)/e)p - u}\). Taking \(\epsilon = 2/(e^2 - 1)\) and \(u = 2p + t (t \geq 0)\) implies the claimed result. \(\square\)
B.2.2 Proof of Lemma B.2

Recall that the empirical loss $\hat{Q}_h(\cdot)$ in (2.6) is convex and twice continuously differentiable with

$$\nabla \hat{Q}_h(\beta) = (1/n) \sum_{i=1}^{n} [K((x_i, \beta) - y_i)/h] - \tau x_i$$

and $\nabla^2 \hat{Q}_h(\beta) = (1/n) \sum_{i=1}^{n} K_h((x_i, \beta) - y_i)x_i x_i^\top$.

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

$$D(\beta, \beta^*) = \frac{1}{n} \sum_{i=1}^{n} \left( K \left( \frac{(x_i, \beta) - y_i}{h} \right) - K \left( \frac{-E_i}{h} \right) \right) \langle x_i, \beta - \beta^* \rangle. \quad (B.17)$$

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

$$D(\beta, \beta^*) \geq \frac{k_f}{nh} \sum_{i=1}^{n} \langle x_i, \beta - \beta^* \rangle^2 1_{\mathcal{E}_i}, \quad (B.18)$$

where $1_{\mathcal{E}_i}$ is the indicator function of $\mathcal{E}_i$ and $k_f = \min_{|u| \leq 1} K(u)$. 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 $\varphi_R(u) = u^2 1(|u| \leq R/2) + |u \text{sign}(u) - R|^2 1(R/2 < |u| \leq R)$, which is $R$-Lipschitz continuous and satisfies

$$u^2 1(|u| \leq R/2) \leq \varphi_R(u) \leq u^2 1(|u| \leq R). \quad (B.19)$$

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 $\delta = \Sigma^{1/2}(\beta - \beta^*)/\|\beta - \beta^*\|_\Sigma$ so that $\delta \in \mathbb{S}^{p-1}$. Together, (B.18) and (B.19) imply

$$\frac{D(\beta, \beta^*)}{\|\beta - \beta^*\|_\Sigma^2} \geq k_f \cdot \frac{1}{nh} \sum_{i=1}^{n} \omega_i : \varphi_R(2r)(\langle w_i, \delta \rangle), \quad (B.20)$$

where $\omega_i := 1(|e_i| \leq h/2)$ 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),

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

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

from which it follows that

$$\inf_{\delta \in \mathbb{S}^{p-1}} \mathbb{E}\{D_0(\delta)\} \geq \int_h h \cdot \left(1 - \sup_{u \in \mathbb{S}^{p-1}} \mathbb{E}[(\langle w, u \rangle)^2 1_{|\langle w, u \rangle| > h/(4r)}]\right), \quad (B.22)$$

By the definition of $\eta_\delta$ in (B.14), we see that as long as $0 < r \leq h/(4\eta_1/4)$,

$$\inf_{\delta \in \mathbb{S}^{p-1}} \mathbb{E}\{D_0(\delta)\} \geq \frac{3}{4} \bar{f}_h. \quad (B.23)$$

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

$$\Delta = \sup_{\delta \in \mathbb{S}^{p-1}} [D^-_0(\delta) - \mathbb{E}[D^-_0(\delta)]]], \quad (B.24)$$

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,

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

This, combined with (B.21), yields

$$\mathbb{E}(\chi_i^2) \leq \frac{\mathbb{E}(\omega_i)}{(4r)^2} \leq \frac{\bar{f}_h h}{(4r)^2}.
$$

With the above preparations, it follows from Theorem 7.3 in Bousquet (2003) that for any $t > 0$,

$$\Delta \leq \mathbb{E}(\Delta) + \sqrt{(\mathbb{E}(\Delta))^2 + \int h t \cdot \mathbb{E}[\varphi_{h/(2r)}(\langle w_i, \delta \rangle)] - \mathbb{E}[\varphi_{h/(2r)}(\langle w_i, \delta \rangle)]} \leq \frac{5}{4} \mathbb{E}(\Delta) + \sqrt{\frac{\bar{f}_h h t}{8r^2 n} + \frac{ht}{3r^2 n}} \quad (B.25)$$

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 $a, b \in \mathbb{R}$. 

11
It then remains to bound the expectation $\mathbb{E}(\Delta)$. Define

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

where $z_i = (w_i, e_i)$ and $\omega_i = \mathbb{1}(|e_i| \leq h/2) \in \{0, 1\}$. By Rademacher symmetrization,

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

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

$$|\mathcal{E}(\delta; z_i) - \mathcal{E}(\delta'; z_i)| \leq \frac{1}{2r} |\langle \omega_i w_i, \delta \rangle - \langle \omega_i w_i, \delta' \rangle|.$$  

(B.26)

Moreover, observe that $\mathcal{E}(\delta; z_i) = 0$ for any $\delta$ such that $\langle \omega_i w_i, \delta \rangle = 0$. 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

$$T = \{ t = (t_1, \ldots, t_n) : t_i = \langle \omega_i w_i, \delta \rangle, i = 1, \ldots, n, \delta \in \mathbb{S}^{p-1} \},$$

and contractions $\phi_i : \mathbb{R} \to \mathbb{R}$ as $\phi_i(t) = (2r/h) \cdot \varphi_{h \omega_i/(2r)}(t)$. By (B.26), $|\phi(t) - \phi(s)| \leq |t - s|$ 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

$$\mathbb{E}(\Delta) \leq 2 \mathbb{E} \left\{ \sup_{\delta \in \mathbb{S}^{p-1}} \frac{1}{n} \sum_{i=1}^{n} e_i \mathcal{E}(\delta; z_i) \right\} = \frac{1}{r} \mathbb{E} \left\{ \sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} e_i \phi_i(t_i) \right\} \leq \frac{1}{r} \mathbb{E} \left\{ \sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} e_i t_i \right\} \leq \frac{1}{r} \sqrt{\frac{8hp}{n}^2} \frac{n}{n} = \frac{1}{r} \sqrt{\frac{h}{12h^2n}} \cdot \frac{n}{n} = \frac{1}{r} \sqrt{\frac{8hp}{3r^2n}},$$

This, together with (B.24) and (B.25), yields

$$\Delta \leq \frac{5}{4} \sqrt{\frac{h}{r^2n}} + \sqrt{\frac{ht}{8r^2n}} + \frac{ht}{3r^2n}. \quad \text{(B.27)}$$

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

Finally, combining (B.17), (B.20), (B.23), and (B.27) proves (B.15). \qed
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 \( \sqrt{(p + t)/n} \leq h \leq 1 \). Define the vector-valued random process

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

where \( D = \mathbb{E}\{f_{x|x}(0)x \Sigma x^T\} \). Since \( \hat{\beta}_h \) falls in a local neighborhood of \( \beta^* \) with high probability, it suffices to bound the local fluctuation \( \sup_{\beta \in [\beta^* + \Theta(r) \cap [\beta^* - 2\beta^*]]} \|\Delta(\beta)\|_2 \). By the triangle inequality,

\[
\sup_{\beta \in [\beta^* + \Theta(r) \cap [\beta^* - 2\beta^*]]} \|\Delta(\beta)\|_2 \leq \sup_{\beta \in [\beta^* + \Theta(r) \cap [\beta^* - 2\beta^*]]} \|\mathbb{E}\Delta(\beta)\|_2 + \sup_{\beta \in [\beta^* + \Theta(r) \cap [\beta^* - 2\beta^*]]} \|\Delta(\beta) - \mathbb{E}\Delta(\beta)\|_2 := I_1 + I_2.
\]

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,

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

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

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

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

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

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

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

$$
\left\| \Sigma^{-1/2} \nabla^2 Q_h((1 - t)\beta^* + t\beta) \Sigma^{-1/2} - D_0 \right\|_2
= \left\| \mathbb{E} \int K(u) \left[ f_{e|x}(t(w, v) - hu) - f_{e|x}(0) \right] \, du \cdot w w^T \right\|_2
\leq l_0 t \sup_{\|u\|_2 = 1} \mathbb{E}((w, u)^2) + l_0 \kappa_1 h \sup_{\|u\|_2 = 1} \mathbb{E}(w, u)^2
\leq l_0 t \left( \sup_{\|u\|_2 = 1} \mathbb{E}((w, u)^3) \right)^{2/3} \left( \frac{1}{3} + l_0 \kappa_1 h \right) \leq l_0 t (\mu_3 r/2 + \kappa_1 h),
$$

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

$$
\sup_{\beta \in (\beta^* + \Theta(r))} \|\mathbb{E}\Delta(\beta)\|_2 \leq l_0 (\mu_3 r/2 + \kappa_1 h) \cdot r. \tag{B.30}
$$

**Upper bound for } I_2:** Next, we provide an upper bound for \( \Delta(\beta) - \mathbb{E}\Delta(\beta) \). Define the centered gradient process \( G(\beta) = \Sigma^{-1/2} \{ \nabla\tilde{Q}_h(\beta) - \nabla Q_h(\beta) \} \), 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

$$
\sup_{\beta \in (\beta^* + \Theta(r))} \|\mathbb{E}\Delta(\beta)\|_2 \leq \sup_{\beta \in (\beta^* + \Theta(r))} \|G(\beta) - G(\beta^*)\|_2
= \sup_{\|v\|_2 \leq r} \left\| G(\beta^* + \Sigma^{-1/2} v) - G(\beta^*) \right\|_2.
$$

We will employ Theorem A.3 in Spokoiny (2013) to bound the supremum \( \sup_{\|v\|_2 \leq r} \|\Delta_0(v)\|_2 \), 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

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

where \( K_i v := K_h(v, v - \epsilon_i) \) satisfies \( 0 \leq K_i v \leq \kappa_h h^{-1} \). For any \( g, h \in \mathbb{S}^{p-1} \) and \( \lambda \in \mathbb{R} \), using the
Given a unit vector $u$, let $\chi = \langle w, g \rangle^2 / (2\nu_1)^2$ so that under Condition 3.5, $\mathbb{P}(\chi \geq u) \leq 2e^{-2u}$ for any $u \geq 0$. It follows that $\mathbb{E}(\chi^k) = 1 + \int_0^\infty u^k \mathbb{P}(\chi \geq u) du \leq 1 + 2 \int_0^\infty u e^{-u} du = 3$, and

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

Taking the supremum over $g \in \mathbb{S}^{p-1}$, we have

$$
\sup_{g \in \mathbb{S}^{p-1}} \mathbb{E}(\langle w, g \rangle^2 / (2\nu_1)^2)^2 \leq 3 \quad \text{and} \quad \sup_{g \in \mathbb{S}^{p-1}} \mathbb{E}(\langle w, g \rangle^4 / (2\nu_1)^2)^2 \leq 8(2\nu_1)^4.
$$
Substituting the above bounds into (B.31) yields that, for any $|\lambda| \leq \min[n^{1/2}h/(4\kappa_n^2), n^{1/2}/\bar{f}]$,

$$
\mathbb{E} \exp \left\{ \lambda n^{1/2} \langle g, \nabla \Delta_0(v) h \rangle \right\}
\leq n \prod_{i=1}^{n} \left[ 1 + \frac{e\lambda^2}{2n} \mathbb{E} \left[ K_{i,v}(w_i,g)(w_i,h) - \mathbb{E}(K_{i,v}(w_i,g)(w_i,h)) \right]^2 e^{\langle w_i,g\rangle(\langle w_i,h \rangle)/(4\nu_i^2)} \right]
\leq n \prod_{i=1}^{n} \left[ 1 + \frac{e\lambda^2}{n} \mathbb{E} (K_{i,v}(w_i,g)(w_i,h))^2 e^{\langle w_i,g\rangle(\langle w_i,h \rangle)/(4\nu_i^2)} + \frac{e\lambda^2}{n} \mathbb{E} [\mathbb{E}(K_{i,v}(w_i,g)(w_i,h))]^2 e^{\langle w_i,g\rangle(\langle w_i,h \rangle)/(4\nu_i^2)} \right]
\leq n \prod_{i=1}^{n} \left[ 1 + C_0^2 \frac{\lambda^2}{2nh} \right] \leq \exp \left\{ C_0^2 \lambda^2/(2h) \right\},
$$

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

$$
\sup_{\|v\|_2 \leq r} \|\Delta_0(v)\|_2 \leq 6C_0 r \sqrt{\frac{4p + 2t}{nh}}
$$
as long as $h \geq 8\kappa_n^2 \sqrt{(2p + t)/n}$ and $n \geq 4\bar{f}^2(2p + t)$.

Together with (B.29) and (B.30), this implies that with probability at least $1 - e^{-t}$,

$$
\sup_{\beta \in \beta^* + \Theta(r)} \|\Delta(\beta)\|_2 \leq 6C_0 r \sqrt{\frac{4p + 2t}{nh}} + l_0(\mu_3 r/2 + \kappa_1 h)r. \tag{B.32}
$$

Recall from the beginning of the proof that $\hat{\beta}_h \in \beta^* + \Theta(r)$ with probability at least $1 - 2e^{-t}$ with $r = r(n, p, t) \approx \sqrt{(p + t)/n + h^2}$. Combined with (B.32), we conclude that with probability at least $1 - 3e^{-t}$, $\|\Delta(\hat{\beta}_h)\|_2 \leq (p + t)/(h^{1/2}n) + h \sqrt{(p + t)/n + h^3}$, as claimed. \hfill \xcirc

### B.4 Proof of Theorem 3.3

Let $a \in \mathbb{R}^p$ 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 $\xi_i = \tau - \tilde{\kappa}(-\epsilon_i/h)$ and $\gamma_i = \langle D^{-1}a, x_i \rangle$. Moreover, write $\delta_n = (p + \log n)/n$. By the Lipschitz continuity of $f_{v|x}$ around 0 and the fundamental theorem of calculus, it can be shown that $|\mathbb{E}(\xi_i|x_i)| \leq 0.5l_0\kappa_2h^2$, from which it follows by the law of iterated expectation that $|\mathbb{E}(\xi_i|x_i)| \leq 0.5l_0\kappa_2\|D^{-1}a\|\Sigma \cdot h^2$. Applying (B.28) and
supplement of Spokoiny and Zhilova (2015), for some constant (B.34) yields

$$\mathbb{E}\left[|n^{1/2}(\alpha, \beta_h - \beta^*) - S_n^0|\right]$$

$$= n^{1/2}\left|\sum_{i=1}^{n} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} \left(\tau - \bar{K}(-\epsilon_i/\tau)|x_i|\right)\right) + |\mathbb{E}(S_n)|\right|$$

$$\leq c_1\|\mathbf{D}^{-1}\alpha\|_\Sigma \cdot n^{1/2}(h^{-1/2}\delta_n + h^2)$$  \hspace{1cm}  \text{(B.33)}$$

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

$$\sup_{x \in \mathbb{R}}|\mathbb{P}(S_n^0 \leq \text{var}(S_n)^{1/2} x) - \Phi(x)| \leq \frac{\mathbb{E}|\gamma \xi - \mathbb{E}(\gamma \xi)|^3}{2[\mathbb{E}(\gamma \xi)^2 - (\mathbb{E}(\gamma \xi))^2]^{3/2} \sqrt{n}}$$  \hspace{1cm}  \text{(B.34)}$$

We have shown that $\mathbb{E}(\gamma \xi) \leq \|\mathbf{D}^{-1}\alpha\|_\Sigma h^2$. 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, $\text{var}(S_n) = \{\tau(1 - \tau) + O(h)\|\mathbf{D}^{-1}\alpha\|^2_\Sigma$ and $\mathbb{E}(\gamma \xi^3) \leq \max(\tau, 1 - \tau)\mathbb{E}(\xi^2|\gamma)\leq \mu_3\{\tau(1 - \tau) + O(h^2)\|\mathbf{D}^{-1}\alpha\|^3_\Sigma$. Substituting these bounds into (B.34) yields

$$\sup_{x \in \mathbb{R}}|\mathbb{P}(S_n^0 \leq \text{var}(S_n)^{1/2} x) - \Phi(x)| \leq c_2n^{-1/2}$$  \hspace{1cm}  \text{(B.35)}$$

for some constant $c_2 > 0$. Set $\sigma^2_x = \tau(1 - \tau)\|\mathbf{D}^{-1}\alpha\|^2_\Sigma$. By an application of Lemma A.7 in the supplement of Spokoiny and Zhilova (2015), for sufficiently small $h$, we have

$$\sup_{x \in \mathbb{R}}|\Phi(x/\text{var}(S_n)^{1/2} - \Phi(x/\sigma_x)| \leq c_3h$$  \hspace{1cm}  \text{(B.36)}$$

Before proceeding, we note that the constants $c_1$–$c_3$ appeared above are all independent of $\alpha$. 

17
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

\[
\mathbb{P}(n^{1/2}(a, \hat{\beta}_h - \beta^*) \leq x) \\
\leq \mathbb{P}(S_n^0 \leq x + c_1\|\mathbf{D}^{-1}a\|_{\Sigma} \cdot n^{1/2}(h^{-1/2}\delta_n + h^2)) + 3n^{-1} \\
\leq \mathbb{P}(\text{var}(S_n)^{1/2}G \leq x + c_1\|\mathbf{D}^{-1}a\|_{\Sigma} \cdot n^{1/2}(h^{-1/2}\delta_n + h^2)) + c_2n^{-1/2} + 3n^{-1} \\
\leq \mathbb{P}(\sigma_x G \leq x + c_1\|\mathbf{D}^{-1}a\|_{\Sigma} \cdot n^{1/2}(h^{-1/2}\delta_n + h^2)) + c_2n^{-1/2} + c_3h + 3n^{-1} \\
\leq \mathbb{P}(\sigma_x G \leq x) + c_1(2\pi)^{-1/2}\|\mathbf{D}^{-1}a\|_{\Sigma} \cdot n^{1/2}(h^{-1/2}\delta_n + h^2)/\sigma_x + c_2n^{-1/2} + c_3h^2 + 3n^{-1},
\]

where the first, second, and third inequalities holds by (B.33), (B.35), and (B.36), respectively, and the last inequality holds by the fact that for any $a \leq b$, $\Phi(b/\sigma_x) - \Phi(a/\sigma_x) \leq (2\pi)^{-1/2}(b - a)/\sigma_x$. 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 $\delta_n^{1/2} \leq h \leq 1$,

\[
\sup_{x \in \mathbb{R}, a \in \mathbb{R}^p} \left| \mathbb{P}(n^{1/2}(a, \hat{\beta}_h - \beta^*) \leq \sigma_x x) - \Phi(x) \right| \leq \frac{p + \log n}{(nh)^{1/2}} + n^{1/2}h^2,
\]

as claimed. \hfill \Box

### 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 $\hat{Q}_h^\circ$ in (2.11) is convex, and thus the first-order condition $\nabla \hat{Q}_h^\circ(\hat{\beta}_h^\circ) = \mathbf{0}$ holds. We use the same localized argument as in the proof of Theorem 3.1. For some $r > 0$ to be determined, define $\hat{\beta}_h^\circ = \beta^* + \eta(\hat{\beta}_h^\circ - \beta^*)$, where $\eta = \sup\{u \in [0, 1] : u(\hat{\beta}_h^\circ - \beta^*) \in \Theta(r)\}$. Similar to (B.11), we have

\[
\|\hat{\beta}_h^\circ - \beta^*\|_{\Sigma} \leq \frac{\|\Sigma^{-1/2}\nabla \hat{Q}_h^\circ(\beta^*)\|_2}{\inf_{\beta \in \beta^* + \Theta(r)} D^\circ(\beta, \beta^*)/\|\beta - \beta^*\|_{\Sigma}^2}, \tag{B.37}
\]

where $D^\circ : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ denotes the symmetrized Bregman divergence associated with $\hat{Q}_h^\circ$. Next, we present two lemmas on the upper and lower bounds of the numerator and denominator of (B.37), respectively.
Lemma B.3. Assume that Conditions 3.1, 3.2, and 3.5 hold. For any $t \geq 0$, there exists some event $E_1(t)$ with $\mathbb{P}[E_1(t)] \geq 1 - 2e^{-t}$ such that, with $\mathbb{P}^*$-probability at least $1 - e^{-t}$ conditioned on $E_1(t)$,

$$\|\Sigma^{-1/2}\nabla \widehat{Q}_h(\beta^*)\|_2 \leq C_1 \left( \sqrt{\frac{p + t}{n}} + h^2 \right) \tag{B.38}$$

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). For any $t \geq 0$, there exists some event $E_2(t)$ with $\mathbb{P}[E_2(t)] \geq 1 - 3e^{-t}$ such that, with $\mathbb{P}^*$-probability at least $1 - e^{-t}$ conditioned on $E_2(t)$,

$$\mathbb{P}^* \left\{ \inf_{\beta \in \beta^* + \Theta(r)} \frac{D_h(\beta, \beta^*)}{\kappa_1\|\beta - \beta^*\|_\Sigma^2} \geq \frac{1}{2} \right\} \geq 1 - e^{-t}. \tag{B.39}$$

as long as $(p + \log n + t)/n \leq h \leq 1$.

Recall from the proof of Theorem 3.1 that the upper bound on $\|\widehat{h} - \beta^*\|_\Sigma$ depends crucially on $\|\Sigma^{-1/2}\nabla \widehat{Q}_h(\beta^*)\|_2$ and $\inf_{\beta \in \beta^* + \Theta(r)} \{D_h(\beta, \beta^*)/\|\beta - \beta^*\|_\Sigma^2\}$. In view of Lemmas B.3 and B.4, we take $r = h/(4\eta_{1/4})$ and for $t \geq 0$, set $E(t) = E_1(t) \cup E_2(t)$ so that $\mathbb{P}[E(t)] \geq 1 - 5e^{-t}$ and

$$\|\widehat{\beta}_h - \beta^*\|_\Sigma \leq C_0 \left( \sqrt{\frac{p + t}{n}} + h^2 \right) \text{ on } E(t) \text{ provided that } \sqrt{\frac{p + t}{n}} \leq h \leq 1,$$

where $C_0 > 0$ depends only on $(\nu_1, \kappa_2, \kappa_l, l_0, \ell)$ and $\tau$. Moreover, by (B.37), with $\mathbb{P}^*$-probability at least $1 - 2e^{-t}$ conditioned on $E(t)$,

$$\|\widehat{\beta}_h - \beta^*\|_\Sigma \leq \frac{2C_1}{\kappa_1\ell} \left( \sqrt{\frac{p + t}{n}} + h^2 \right) < r$$

as long as $\max\{p + \log n + t)/n, \sqrt{(p + t)/n}\} \leq h \leq 1$. Consequently, $\widehat{\beta}_h$ falls in the interior of $\beta^* + \Theta(r)$, thereby implying $\eta = 1$ and $\widehat{\beta}_h = \beta^*_\eta$. This completes the proof. \hfill $\Box$

B.5.1 Proof of Lemma B.3

By the triangle inequality, we have

$$\|\Sigma^{-1/2}\nabla \widehat{Q}_h(\beta^*)\|_2 \leq \|\Sigma^{-1/2}\nabla \widehat{Q}_h(\beta^*) - \Sigma^{-1/2}\nabla \widehat{Q}_h(\beta^*)\|_2 + \|\Sigma^{-1/2}\nabla \widehat{Q}_h(\beta^*)\|_2 := I_1 + I_2. \tag{B.40}$$
The term $I_2$ can be upper bounded by Lemma B.1, and thus it suffices to bound $I_1$. To this end, we will employ Hoeffding’s and Bernstein’s inequalities.

Recall from Section 2.4 that $w_i$ is such that $\mathbb{E}(w_i) = 1$ and $\text{var}(w_i) = 1$. Thus, $\Sigma^{-1/2} \nabla \hat{Q}_h^{\beta}(\beta^*) = (1/n) \sum_{i=1}^n w_i \xi_i w_i$ and $\mathbb{E}^* (\Sigma^{-1/2} \nabla \hat{Q}_h^{\beta}(\beta^*)) = \Sigma^{-1/2} \hat{Q}_h(\beta^*)$, where $w_i = \Sigma^{-1/2} x_i$ and $\xi_i = \tilde{K}(\epsilon_i/h) - \tau$ with $\epsilon_i = y_i - (x_i, \beta^*)$. 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 $|\mathcal{N}_\epsilon| \leq (1 + 2/\epsilon)^p$ such that

$$
\left\| \Sigma^{-1/2} \nabla \hat{Q}_h^{\beta}(\beta^*) - \Sigma^{-1/2} \nabla \hat{Q}_h(\beta^*) \right\|_2 \leq \frac{1}{1 - \epsilon} \max_{u \in \mathcal{N}_\epsilon} \left\{ \frac{1}{n} \sum_{i=1}^n e_i \langle u, w_i \rangle \right\},
$$

where $e_i = w_i - 1$ are i.i.d. Rademacher random variables. By Hoeffding’s inequality, for any $u \geq 0$,

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

For the data-dependent quantity $(1/n) \sum_{i=1}^n \xi_i^2 \langle u, w_i \rangle^2$, as in the proof of Lemma B.1 we have $|\xi| \leq B_\tau := \max(\tau, 1 - \tau)$, $\mathbb{E}(\xi^4 \langle x_i \rangle) \leq C_\tau := \tau(1 - \tau) + (1 + \tau)l_0 \kappa_2 h^2$ almost surely and note that $\mathbb{E} \langle u, w_i \rangle^2 = 1$. Moreover, for $k = 2, 3, \ldots$,

$$
\mathbb{E}(\xi^2 \langle u, w_i \rangle^2)^k \leq C_\tau B_\tau^{2(k-1)} v_1^{2k} \cdot 2k \int_0^\infty \mathbb{P}((\langle u, w_i \rangle) \geq v_1 u) u^{2k-1} \, du \leq C_\tau B_\tau^{2(k-1)} v_1^{2k} \cdot 4k \int_0^\infty u^{2k-1} e^{-u^2/2} \, du = C_\tau B_\tau^{2(k-1)} 2^k v_1^{2k} \cdot 2k \int_0^\infty \nu^{k-1} e^{-\nu} \, dv = C_\tau B_\tau^{2(k-1)} 2^{k+1} v_1^{2k} k!.
$$

In particular, $\mathbb{E}(\xi^4 \langle u, w_i \rangle^4) \leq B_\tau^2 C_\tau + 16 v_1^4$ and $\mathbb{E}(\xi^2 \langle u, w_i \rangle^2)^2 \leq \frac{4}{9} \cdot B_\tau^2 C_\tau + 16 v_1^4 \cdot (2B_\tau^4 v_1^k)^{k-2}$ for $k \geq 3$. Since $\mathbb{E}(\xi_i^2 \langle x_i \rangle) \leq C_\tau$, it then follows from Bernstein’s inequality that, for any $v \geq 0$,

$$
\mathbb{P} \left\{ \frac{1}{n} \sum_{i=1}^n \xi_i^2 \langle u, w_i \rangle^2 \geq C_\tau + 4B_\tau C_\tau^{1/2} v_1^2 \sqrt{\frac{2\nu}{n}} + B_\tau^2 v_1^2 \frac{2\nu}{n} \right\} \leq e^{-v}.
$$

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

$$
\max_{u \in \mathcal{N}_\epsilon} \frac{1}{n} \sum_{i=1}^n \xi_i^2 \langle u, w_i \rangle^2 \leq C_\tau + 4B_\tau C_\tau^{1/2} v_1^2 \sqrt{\frac{4p + 2t}{n}} + 2B_\tau^2 v_1^2 \frac{2p + t}{n} \tag{B.41}
$$
with probability at least $1 - e^{-t}$. Let $\mathcal{E}_t(t)$ be the event that (B.41) holds. Then, with $\mathbb{P}^*$-probability at least $1 - e^{-t}$ conditioned on $\mathcal{E}_t(t)$, and by an application of union bound, we obtain
\[
\|\Sigma^{-1/2} \nabla \hat{Q}_n^*(\beta^*) - \Sigma^{-1/2} \tilde{Q}_n(\beta^*)\|_2 \leq C \sqrt{\frac{p + t}{n}}
\]
as long as $n \geq p + t$, where $C > 0$ depends only on $(v_1, \kappa_2, l_0)$ and $\tau$. This, together with (B.40) and Lemma B.1, proves the claimed bound.

\[\Box\]

### 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
\[
D^i(\beta, \beta^*) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \tilde{K} \left( \frac{(x_i, \beta) - y_i}{h} \right) - \tilde{K} \left( \frac{-\epsilon_i}{h} \right) \right\} w_i(x_i, \beta - \beta^*)
\]
\[
\geq \kappa \|\beta - \beta^\ast\|_{\Sigma}^2 \cdot \frac{1}{nh} \sum_{i=1}^{n} w_i(\omega_i\phi_{h/(2r)}((\omega, \delta)),
\]
where $\delta = \Sigma^{1/2}(\beta - \beta^\ast)/\|\beta - \beta^\ast\|_{\Sigma}$, $\omega_i = 1(|e_i| \leq h/2)$, and $\phi(\cdot)$ is as defined in (B.19). Recall from (B.20) the definition of $D_0(\delta)$. We have
\[
\inf_{\beta \in \beta + \Theta(\tau)} \frac{D^i(\beta, \beta^*)}{\kappa \|\beta - \beta^\ast\|_{\Sigma}^2} \geq \inf_{\delta \in \Theta^{p-1}} D_0(\delta) - \sup_{\delta \in \Theta^{p-1}} \{D_0(\delta) - D^i_0(\delta)\}.
\] (B.42)

We now obtain lower and upper bounds for $D_0(\delta)$ and $\{D_0(\delta) - D^i_0(\delta)\}$, respectively.

Recall that $w_i$ is a random variable that is independent of $x_i$ with $\mathbb{E}(w_i) = 1$. Let $e_i = w_i - 1 \in \{-1, 1\}$. Then, we have
\[
D^i_0(\delta) - D_0(\delta) = \frac{1}{nh} \sum_{i=1}^{n} e_i(\omega_i\phi_{h/(2r)}((\omega, \delta))
\]
Define $\gamma_\ast = \gamma(\epsilon_1, \ldots, \epsilon_n) = \sup_{\delta \in \Theta^{p-1}} [D_0(\delta) - D^i_0(\delta)]$. Recall that $\phi_R(u) \leq (R/2)^2$, we have $\mathbb{E}^\ast(e_i\omega_i\phi_{h/(2r)}((\omega, \delta)))^2 \leq (h/4r)^4 \omega_i$ and $|e_i\omega_i\phi_{h/(2r)}((\omega, \delta))| \leq (h/4r)^2$. Then, by the Talagrand’s
inequality (see Theorem 7.3 in Bousquet (2003)), for every \( t \geq 0 \),
\[
\Gamma_n \leq \mathbb{E}^*(\Gamma_n) + \sqrt{\frac{h^2}{(4r)^4} \sum_{i=1}^n \omega_i \frac{2t}{n} + 4\mathbb{E}^*(\Gamma_n) \frac{h}{(4r)^2} \frac{t}{n} + \frac{h}{(4r)^2} \frac{t}{3n}}
\]
\[
\leq 2\mathbb{E}^*(\Gamma_n) + \frac{h}{(4r)^2} \left\{ \left( \frac{1}{n} \sum_{i=1}^n \omega_i \right)^{1/2} \sqrt{\frac{2t}{n} + \frac{4t}{3n}} \right\}
\]
with probability at least \( 1 - e^{-t} \). Further, by the Lipschitz continuity of \( u \to \varphi_R(u) \) and Talagrand’s contraction principle,
\[
\mathbb{E}^*(\Gamma_n) \leq \frac{1}{2r} \mathbb{E}^* \left\{ \sup_{\delta \in S^{p-1}} \left\| \frac{1}{n} \sum_{i=1}^n e_i \langle \omega_i, \mathbf{w}_i \rangle, \delta \right\| \right\}
\]
\[
\leq \frac{1}{2r} \mathbb{E}^* \left\| \frac{1}{n} \sum_{i=1}^n e_i \langle \omega_i, \mathbf{w}_i \rangle \right\|_2
\]
\[
\leq \frac{1}{2rn} \left( \sum_{i=1}^n \omega_i \| \mathbf{w}_i \|_2 \right)^{1/2}
\]
\[
\leq \frac{\max_{1 \leq i \leq n} \| \mathbf{w}_i \|_2}{2r \sqrt{n}} \left( \frac{1}{n} \sum_{i=1}^n \omega_i \right)^{1/2}.
\]
Together, the last two displays imply
\[
\sup_{\delta \in S^{p-1}} \left\{ D_0(\delta) - D_0^*(\delta) \right\} \leq \left( \frac{1}{n} \sum_{i=1}^n \omega_i \right)^{1/2} \left\{ \frac{\max_{1 \leq i \leq n} \| \mathbf{w}_i \|_2}{r \sqrt{n}} + \frac{h}{(4r)^2} \sqrt{\frac{2t}{n}} + \frac{h}{(4r)^2} \frac{4t}{3n} \right\}
\] (B.43)
with \( \mathbb{P}^* \)-probability at least \( 1 - e^{-t} \).

Next we provide upper bounds for the data-dependent terms \( \max_{1 \leq i \leq n} \| \mathbf{w}_i \|_2 \) 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 \( \mathcal{N}_\epsilon \subseteq S^{p-1} \) with \( |\mathcal{N}_\epsilon| \leq (1 + 2/\epsilon)^p \) such that \( \max_{1 \leq i \leq n} \| \mathbf{w}_i \|_2 \leq (1 - \epsilon)^{-1} \max_{u \in \mathcal{N}_\epsilon} \max_{u \in \mathcal{N}_\epsilon} \langle u, \mathbf{w}_i \rangle \). Given \( 1 \leq i \leq n \) and \( u \in \mathcal{N}_\epsilon \), Condition 3.5 indicates \( \mathbb{P}(\| u, \mathbf{w}_i \| \geq \nu_1 u) \leq 2e^{-u^2/2} \) for any \( u \in \mathbb{R} \). Taking the union bound over \( i \) and \( u \), and setting \( u = \sqrt{2t + 2 \log(2n) + 2p \log(1 + 2/\epsilon)} \) (\( t > 0 \)), we obtain that with probability at least \( 1 - 2n(1 + 2/\epsilon)^p e^{-u^2/2} = 1 - e^{-t} \), \( \max_{1 \leq i \leq n} \| \mathbf{w}_i \|_2 \leq (1 - \epsilon)^{-1} \nu_1 \sqrt{2t + 2 \log(2n) + 2p \log(1 + 2/\epsilon)} \). Minimizing this upper bound with respect to \( \epsilon \in (0, 1) \), we conclude that for any \( t > 0 \),
\[
\mathbb{P} \left[ \max_{1 \leq i \leq n} \| \mathbf{w}_i \|_2 \geq 2\nu_1^2 \left( 3.7p + \log(2n) + t \right) \right] \leq e^{-t}.
\] (B.44)
Moreover, applying Bernstein’s inequality to 
\( \frac{1}{n} \sum_{i=1}^{n} \omega_i = (1/n) \sum_{i=1}^{n} \mathbb{1}(|\varepsilon_i| \leq h/2) \) yields
\[
\frac{1}{n} \sum_{i=1}^{n} \omega_i \leq \mathbb{E}(\omega_i) + \sqrt{2\mathbb{E}(\omega_i) \frac{t}{n}} + \frac{t}{3n} \leq \left( \sqrt{\mathbb{E}(\omega_i)} + \sqrt{\frac{t}{2n}} \right)^2
\]
with probability greater than \( 1 - e^{-t} \). Note that \( \mathbb{E}(\omega_i) \leq \bar{f}h \). Putting together the pieces, we conclude that conditioned on some event that occurs with probability at least \( 1 - 2e^{-t} \),
\[
\sup_{\delta \in \mathbb{S}^{p-1}} \{ D_0(\delta) - D_0^h(\delta) \} \leq \{(\bar{f}h)^{1/2} + (0.5t/n)^{1/2}\} \left( \frac{\nu_1(7.4p + 2 \log(2n) + 2r)^{1/2}}{r^{1/2}} + \frac{h}{(4r)^2} \sqrt{\frac{2r}{n}} \right) + \frac{h}{(4r)^2} \sqrt{\frac{4t}{3n}} \quad (B.45)
\]
with \( \mathbb{P}^* \)-probability greater than \( 1 - e^{-t} \).

Turning to \( D_0(\delta) \), applying Lemma B.2 yields that as long as \( 0 < r \leq h/(4\eta_1/4) \),
\[
\inf_{\delta \in \mathbb{S}^{p-1}} D_0(\delta) \geq \frac{3}{4} \bar{L}_h - \frac{\bar{f}h}{2} \left( \frac{5}{4} \sqrt{\frac{hp}{r^2n}} + \sqrt{\frac{ht}{8r^2n}} \right) - \frac{ht}{3r^2n} \quad (B.46)
\]
with probability at least \( 1 - e^{-t} \). Substituting (B.45) and (B.46) into (B.42), and taking \( r = h/(4\eta_1/4) \) with \( h \geq t/n \), we conclude that conditioned on some event that occurs with probability at least \( 1 - 3e^{-t} \),
\[
\inf_{\beta \in \mathbb{B}^{p+\Theta(r)}} D^h_{\beta}(\beta, \beta^*) \geq \frac{3}{4} \bar{L}_h - C \sqrt{\frac{p + \log n + t}{nh}}
\]
with \( \mathbb{P}^* \)-probability greater than \( 1 - e^{-t} \), where \( C > 0 \) depends only on \( (\nu_1, l_0, \bar{f}) \). This completes the proof of (B.39). \( \square \)

### 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
\[
\Delta^h(\beta) = \Sigma^{-1/2} [\nabla \hat{Q}_h(\beta) - \nabla \hat{Q}_h^*(\beta^*) - D(\beta - \beta^*)], \quad \beta \in \mathbb{R}^p. \quad (B.47)
\]
For a prespecified $r > 0$, a key step is to bound the local fluctuation $\sup_{\beta \in \beta^* + \Theta(r)} \| \Delta^i(\beta) \|_2$. Since $\mathbb{E}(w_i) = 1$, we have $\mathbb{E}^* \{ \nabla \hat{Q}_n^i(\beta) \} = \nabla \hat{Q}_n(\beta)$. Define the (conditionally) centered process

$$G^\delta(\beta) = \Sigma^{-1/2} [ \nabla \hat{Q}_n^i(\beta) - \nabla \hat{Q}_n(\beta) ] = \frac{1}{n} \sum_{i=1}^{n} e_i [ K(\tau) - (\tau \cdot \langle x_i, \beta \rangle) ] w_i,$$

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

$$\Delta^i(\beta) = [ G^\delta(\beta) - G^\delta(\beta^*^*) ] + \Delta(\beta),$$

where $\Delta(\beta)$ is defined in (B.28). By the triangle inequality,

$$\sup_{\beta \in \beta^* + \Theta(r)} \| \Delta^i(\beta) \|_2 \leq \sup_{\beta \in \beta^* + \Theta(r)} \| G^\delta(\beta) - G^\delta(\beta^*^*) \|_2 + \sup_{\beta \in \beta^* + \Theta(r)} \| \Delta(\beta) \|_2. \quad (B.48)$$

It suffices to bound the first term on the right-hand side of (B.48). By a change of variable $v = \Sigma^{1/2}(\beta - \beta^*^*)$, we have $y_i - \langle x_i, \beta \rangle = e_i - \langle w_i, v \rangle$ and

$$\sup_{\beta \in \beta^* + \Theta(r)} \| G^\delta(\beta) - G^\delta(\beta^*^*) \|_2 \leq \sup_{v \in \mathbb{B}^p(1), u \in \mathbb{B}^p(1)} \langle G^\delta(\beta^*^*) + \Sigma^{-1/2} v, u \rangle - \langle G^\delta(\beta^*^*) \rangle \frac{n^{1/2} \langle G^\delta(\beta^*^*) + \Sigma^{-1/2} v, u \rangle - \langle G^\delta(\beta^*^*) \rangle, u \rangle}{\| u \|_2^2}. \quad (B.49)$$

where $\Psi(u, v) = n^{-1/2} \sum_{i=1}^{n} e_i(\langle w_i, u \rangle - \langle w_i, v \rangle - \langle v, u \rangle)$. To characterize the magnitude of $\sup_{u, v \in \mathbb{B}^p(1)} \Psi(u, v)$, define the following classes of measurable functions

$$\mathcal{F}_1 = \{ (w, \epsilon) \mapsto \langle w, u \rangle : u \in \mathbb{B}^p(1) \} \quad \text{and} \quad \mathcal{F}_2 = \{ (w, \epsilon) \mapsto \langle w, v \rangle - \epsilon : v \in \mathbb{B}^p(1) \}. \quad (B.50)$$

Also, define functions $f_0 : (w, \epsilon) \mapsto \bar{K}(\epsilon/h)$ and $\phi : u \mapsto \bar{K}(u/h)$. With this notation, we can write

$$\sup_{u, v \in \mathbb{B}^p(1)} \Psi(u, v) = \sup_{g \in \mathcal{G}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i g(z_i), \quad (B.51)$$

where $z_i = (w_i, e_i) \in \mathbb{R}^p \times \mathbb{R}$ and $\mathcal{G} = \mathcal{F}_1 \cdot (\phi \circ \mathcal{F}_2 - f_0)$ denotes the pointwise product between $\mathcal{F}_1$ and $\phi \circ \mathcal{F}_2 - f_0$. Let $Z = \mathbb{E}^* \{ \sup_{g \in \mathcal{G}} (1/n) \sum_{i=1}^{n} e_i g(z_i) \}$ be the conditional Rademacher average.
By Theorem 13 in Boucheron et al. (2005) and the bound \( \sup_{1 \leq i \leq n, g \in G} g(z_i) \leq r \max_{1 \leq i \leq n} \|w_i\|_2 \), we have

\[
[\mathbb{E}[Z - \mathbb{E}(Z)]^{2k}]^{1/(2k)} \leq 2 \sqrt{\mathbb{E}(Z) \cdot k \kappa r \frac{M_{n,k}}{n} + 2k \kappa r \frac{M_{n,k}}{n}}, \quad \text{valid for any } k \geq 1, \quad (B.52)
\]

where \( \kappa = \sqrt{e}/(2 \sqrt{e} - 2) \) and \( M_{n,k} = (\mathbb{E}(\max_{1 \leq i \leq n} \|w_i\|_2^{2k}))^{1/(2k)} \).

Next we bound \( \mathbb{E}(Z) \) via a maximal inequality specialized to VC type classes. Note that the function class \( \mathcal{F}_1 \) is a \( p \)-dimensional vector space, and \( \mathcal{F}_2 \) is included in a \((p + 1)\)-dimensional vector space. By Lemma 2.6.15 of van der Vaart and Wellner (1996), both \( \mathcal{F}_1 \) and \( \mathcal{F}_2 \) have VC-subgraph of index smaller than or equal to \( p + 2 \). Moreover, since \( \phi(\cdot) = K(\cdot/h) \) is a non-decreasing function, it follows from Lemma 9.9 of Kosorok (2008) that \( \phi \circ \mathcal{F}_2 \) is VC with index less than or equal \( p + 2 \), and so is \( \phi \circ \mathcal{F}_2 - f_0 \). Let \( F_1(w, \varepsilon) = r\|w\|_2 \) and \( F_2(w, \varepsilon) \equiv 1 \) be the envelops of \( \mathcal{F}_1 \) and \( \phi \circ \mathcal{F}_2 - f_0 \), respectively. Hence, by Theorem 9.3 in Kosorok (2008) (or Theorem 2.6.7 in van der Vaart and Wellner (1996)), we obtain the following bounds on the covering numbers of function classes \( \mathcal{F}_1 \) and \( \phi \circ \mathcal{F}_2 - f_0 \): there exists a universal constant \( A_0 > 0 \) such that, for any probability measure \( Q \) with \( \|F_1\|_{Q,2} > 0 \) and any \( \varepsilon \in (0, 1) \),

\[
N(\mathcal{F}_1, L_2(Q), \varepsilon\|F_1\|_{Q,2}) \leq A_0(p + 2)(4\varepsilon)^{p+2} \left( \frac{2}{\varepsilon} \right)^{2(p+1)};
\]

\[
N(\phi \circ \mathcal{F}_2 - f_0, L_2(Q), \varepsilon\|F_2\|_{Q,2}) \leq A_0(p + 2)(4\varepsilon)^{p+2} \left( \frac{2}{\varepsilon} \right)^{2(p+1)}.
\]

For the product class \( G \) with envelop \( G(w, \varepsilon) = F_1(w, \varepsilon)F_2(w, \varepsilon) = r\|w\|_2 \), applying Corollary A.1 in the supplement of Chernozhukov, Chetverikov and Kato (2014) yields that, for any \( \varepsilon \in (0, 1) \),

\[
N(\mathcal{G}, L_2(Q), \varepsilon\|G\|_{Q,2}) \\
\leq N(\mathcal{F}_1, L_2(Q), 2^{-1/2} \varepsilon\|F_1\|_{Q,2})N(\phi \circ \mathcal{F}_2 - f_0, L_2(Q), 2^{-1/2} \varepsilon\|F_2\|_{Q,2}) \\
\leq A_0^2(p + 2)^2(4\varepsilon)^{2(p+2)} \left( \frac{2^{3/2}}{\varepsilon} \right)^{4(p+1)} \leq \left( \frac{A_1}{\varepsilon} \right)^{4(p+2)}
\]

for some \( A_1 > \varepsilon \).

Another key quantity is the variance proxy \( \sup_{g \in G} \mathbb{E}[g^2(w, \varepsilon)] \). Given some \( g = g_{u,v} \in G \) with
\(u, v \in B^p(r)\), we have

\[
\mathbb{E}[g^2(w, \varepsilon)] = \mathbb{E}(\langle w, u \rangle^2 [K(\langle w, v \rangle - \varepsilon)/h] - K(-\varepsilon/h)^2)
\]

\[
= \mathbb{E}(\langle w, u \rangle^2 \int_{-\infty}^{\infty} [K((\langle w, v \rangle - u)/h) - K(-u/h)] f_{\|x\|}(u) du)
\]

\[
= h \mathbb{E}(\langle w, u \rangle^2 \int_{-\infty}^{\infty} [K((\langle w, v \rangle/h + v) - K(v)] f_{\|x\|}(-vh) dv)
\]

\[
\leq f h^{-1} \mathbb{E}(\langle w, u \rangle^2 (w, v)^2 \int_{-\infty}^{\infty} \left( \int_{0}^{1} K(v + s\langle w, v \rangle/h) ds \right)^2 dv)
\]

\[
\leq f h^{-1} \mathbb{E}(\langle w, u \rangle^2 (w, v)^2 \left( \int_{0}^{1} \left( \int_{-\infty}^{\infty} K^2(v + s\langle w, v \rangle/h) dv \right)^{1/2} ds \right)^2)
\]

\[
\leq \kappa_u f h^{-1} \mathbb{E}(\langle w, u \rangle^2 (w, v)^2),
\]

where the second inequality follows from an application of generalized Minkowski's integral inequality. Hence, \(\sup_{g \in G} \mathbb{E}[g^2(w, \varepsilon)] \leq \kappa_u f \mu_4 h^{-1} \). For the envelop \(G\) for \(G\), it can be shown that \(\mathbb{E}[G^2(w, \varepsilon)] = r^2 P, \max_{1 \leq i \leq n} \sup_{g \in G} |g(z_i)| \leq r \max_{1 \leq i \leq n} \|w_i\|_2, \) and \(\mathbb{E} \max_{1 \leq i \leq n} G^2(w, \varepsilon_i) = r^2 M_{n,1}\) with \(M_{n,1}\) given in (B.52). Equipped with the above bounds, we apply Corollary 5.1 in Chernozhukov, Chetverikov and Kato (2014) to obtain

\[
\mathbb{E}(Z) \leq r^2 \sqrt{\frac{p}{nh} \log(A_2ph/r^2) + r M_{n,1} \log(A_2ph/r^2) \frac{p}{n}}, \tag{B.53}
\]

where \(A_2 > 0\) depends on \((\tilde{f}, \kappa_u, \mu_4)\) and \(A_1\). It remains to bound \(M_{n,1}\), which appears in both (B.52) and (B.53). In fact, the exponential tail probability (B.44) can be directly used to bound the expectation. Given a non-negative random variable \(X\) and constants \(A, a > 0\), we have

\[
\mathbb{E}(X) = \int_{0}^{\infty} \mathbb{P}(X \geq t) dt \leq A + \int_{A}^{\infty} \mathbb{P}(X \geq t) dt
\]

\[
= A + \int_{A}^{\infty} \mathbb{P}(X \geq A + t) dt = A + a \int_{0}^{\infty} \mathbb{P}(X \geq A + as) ds.
\]

Taking \(X = \max_{1 \leq i \leq n} \|w_i\|_2^2, A = 2\nu_1^2[3.7p + \log(2n)]\) and \(a = 2\nu_1^2\), we obtain

\[
M_{n,1}^2 = \mathbb{E}(\max_{1 \leq i \leq n} \|w_i\|_2^2) \leq 2\nu_1^4[3.7p + \log(2en)]. \tag{B.54}
\]

Back to the supremum \(\sup_{\beta \in \beta^* + \Theta(r)} \|G^\flat(\beta) - G^\flat(\beta^*)\|_2\), first by (B.49), (B.51), and the definition of
of $Z$ below (B.51), we obtain

$$\sup_{\beta \in \beta^0 + \Theta(r)} \|G^\beta(\beta) - G^{\beta^0}(\beta^*)\|_2 = O_P(r^{-1}Z). \tag{B.55}$$

For the conditional Rademacher average $Z$, it follows from Markov’s inequality that $P\{Z \leq \mathbb{E}(Z) + u\} \leq e^{-u^2 \mathbb{E}(Z - \mathbb{E}(Z))^2}$, valid for any $u > 0$. Applying the bound (B.52) with $k = 1$, (B.53) and (B.54) yields

$$[\mathbb{E}[Z - \mathbb{E}(Z)]^2]^{1/2} \leq \sqrt{\mathbb{E}(Z) \cdot (p + \log n)^{1/2} \frac{r}{n} + (p + \log n)^{1/2} \frac{r}{n}};$$

$$\mathbb{E}(Z) \leq r^2 \sqrt{\frac{p}{nh} \log(A_2 ph/r^2) + r \log(A_2 ph/r^2)(p + \log n)^{1/2} \frac{P}{n}},$$

from which we conclude that

$$r^{-1}Z = O_P \left( r \sqrt{\frac{p}{nh} \log(A_2 ph/r^2) + \log(A_2 ph/r^2)(p + \log n)^{1/2} \frac{P}{n}} \right). \tag{B.56}$$

Turning to the second term on the right-hand side of (B.48), the earlier bound (B.32) implies that there exists some event $G_1(t)$ with $P\{G_1(t)\} \geq 1 - e^{-t}$ such that

$$\sup_{\beta \in \beta^0 + \Theta(r)} \|\Delta(\beta)\|_2 \leq C_1 \left( r \sqrt{\frac{p + t}{nh} + hr + r^2} \right)$$

on $G_1(t)$, \tag{B.57}

where $C_1 > 0$ depends on $(\nu_1, \kappa_1, \kappa_u, l_0, \tilde{f}, f)$.}

With the above preparations, we are ready to prove the claim. Assume that the triplet $(n, p, h)$ satisfies the scaling $\sqrt{(p + \log n)/n} \leq h \leq 1$. By Theorems 3.1 and 3.4, there exist some event $\mathcal{E}_n$ with $P(\mathcal{E}_n) \geq 1 - 5n^{-1}$ and some $r = r(n, p, h) = \sqrt{(p + \log n)/n + h^2} > 0$ such that,

(i) $\|\beta_h^* - \beta^*\|_\Sigma \leq r$ on the event $\mathcal{E}_n$,

(ii) $\|\beta_h^* - \beta^*\|_\Sigma \leq r$ with $P^*$-probability at least $1 - 2n^{-1}$ conditioned on $\mathcal{E}_n$.

Taking $t = \log n$ and $\mathcal{G}_{1n} = \mathcal{G}_1(\log n)$ in (B.57), we have on the event $\mathcal{G}_{1n} \cap \mathcal{E}_n$ that

$$\chi_{1n} := \|\Delta(\beta_h^*)\|_2 \leq \frac{p + \log n}{h^{1/2} n} + h \sqrt{\frac{p + \log n}{n}} + h^3.$$
Moreover, with $\mathbb{P}^*$-probability at least $1 - 2n^{-1}$ conditioned on $E_n$,

$$\left\| \Sigma^{-1/2} \left[ \mathbf{D}(\hat{\beta}_n - \beta^*) - \frac{1}{n} \sum_{i=1}^n w_i \tau - \hat{K}(-\varepsilon_i/h) \right] \right\|_2 \leq \sup_{\beta \in \beta^* + \Theta} \| \Delta^1(\beta) \|_2.$$ 

Let $\chi_{2n} = \mathbb{E}^* \{ \sup_{\beta \in \beta^* + \Theta} \| \Delta^1(\beta) \|_2 \}$, for which it follows from (B.48), (B.56), and (B.57) with $t = \log n$ that

$$\chi_{2n} = O_P \left\{ \frac{\sqrt{(p + \log n)p \log n}}{h^{1/2}n} + h^{3/2} \frac{p \log n}{n} + h \frac{p + \log n}{n} + h^3 + (p + \log n)^{1/2} \frac{p \log n}{n} \right\}.$$ 

With $h = h_n = \{(p + \log n)/n\}^\delta$ for some $\delta \in (1/3, 1/2]$, the first term on the right-hand side dominates the rest except the last one. Then, putting together the pieces establishes the claim. □

## C  Theoretical Properties of One-step Conquer

In this section, we provide theoretical properties of the one-step conquer estimator $\hat{\beta}$, defined in Section A.1. The key message is that, when higher-order kernels are used (and if the conditional density $f_{e|x}(\cdot)$ 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, $\int_{-\infty}^{\infty} u^k G(u) \, du = 0$ for $k = 1, \ldots, \nu - 1$ and $\int_{-\infty}^{\infty} u^\nu G(u) \, du \neq 0$. Moreover, $\kappa^G_k = \int_{-\infty}^{\infty} |u^k G(u)| \, du < \infty$ for $1 \leq k \leq \nu$, $G$ is uniformly bounded with $\kappa^G = \sup_{u \in \mathbb{R}} |G(u)| < \infty$ 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 $f_{e|x}(\cdot)$ of $e$ given $x$ is sufficiently smooth. Therefore, we further impose the following smoothness conditions on $f_{e|x}(\cdot)$.

### Condition 2

Let $\nu \geq 4$ be the integer in Condition 1. The conditional density $f_{e|x}(\cdot)$ is $(\nu - 1)$-times differentiable, and satisfies $|f_{e|x}^{(\nu-2)}(u) - f_{e|x}^{(\nu-2)}(0)| \leq L_{\nu-2} |u|$ 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 $\int_{-\infty}^{\infty} |u^{\nu-1} G(u)| \cdot \sup_{|t| \leq |u|} |f_{e|x}^{(\nu-1)}(t) - f_{e|x}^{(\nu-1)}(0)| \, du \leq C_G$ almost surely.
Notably, we have
\[ \nabla Q^G_b(\beta) = \mathbb{E}[G_b((x, \beta) - y)/b) - \tau] x \quad \text{and} \quad \nabla^2 Q^G_b(\beta) = \mathbb{E}[G_b(y - (x, \beta))xx^T], \tag{C.1} \]
representing the population score and Hessian of \( Q^G_b(\cdot) = \mathbb{E}[\hat{Q}^G_b(\cdot)] \). As \( b \to 0 \), we expect \( \nabla Q^G_b(\beta^*) \) and \( \nabla^2 Q^G_b(\beta^*) \) to converge to 0 (zero vector in \( \mathbb{R}^p \)) and \( D = \mathbb{E}\{f_{\epsilon|x}(0)xx^T\} \), respectively. 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
\[
\left\| \Sigma^{-1/2} \nabla^2 \hat{Q}^G_b(\beta^*) \right\|_2 \leq L_{\nu-2} \kappa^G_b \nu! / \nu!
\]
and
\[
\left\| \Sigma^{-1/2} \nabla^2 \hat{Q}^G_b(\beta^*) \Sigma^{-1/2} - D_0 \right\|_2 \leq C_G b^{\nu-1}/(\nu-1)!,
\]
where \( D_0 = \Sigma^{-1/2} D \Sigma^{-1/2} = \mathbb{E}\{f_{\epsilon|x}(0)ww^T\} \) 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 \( \|\nabla Q^G_b(\beta^*)\|_2 = O(b^\nu) \) and \( \|\nabla^2 Q^G_b(\beta^*) - D\|_2 = O(b^{\nu-1}) \), where \( \nu \geq 4 \) is an even integer. Notably, even if the kernel \( H \) has negative parts, the population Hessian \( \nabla^2 Q^G_b(\beta^*) \) preserves the positive definiteness of \( D \) as long as the bandwidth \( g \) is sufficiently small.

To construct the one-step conquer estimator, two key quantities are the sample Hessian \( \nabla^2 \hat{Q}^G_b(\cdot) \) and sample gradient \( \nabla \hat{Q}^G_b(\cdot) \), both evaluated at \( \beta \), 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, \( \nabla^2 \hat{Q}^G_b(\beta) \) 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
\[
\Theta^*(r) = \{ \beta \in \mathbb{R}^p : \| \beta - \beta^* \|_\Sigma \leq r \}. \tag{C.2}
\]

**Proposition C.2.** Conditions 1, 2 and 3.5 ensure that, with probability at least \( 1 - e^{-t} \),
\[
\sup_{\beta \in \Theta^*(r)} \left\| \Sigma^{-1/2} \left( \nabla^2 \hat{Q}^G_b(\beta) - \nabla^2 \hat{Q}^G_b(\beta) \right) \Sigma^{-1/2} \right\|_2 \leq \sqrt{\frac{p \log n + t}{nb}} + \frac{p \log n + t}{nb} + \frac{(p + t)^{1/2}r}{nb^2},
\]
as long as \( n \gtrsim p + t \).
Proposition C.3. Conditions 1, 2 and 3.5 ensure that, with probability at least $1 - e^{-t}$,

$$
\sup_{\beta \in \Theta(r)} \left\| \Sigma^{-1/2} \left( \nabla \tilde{Q}_b^G(\beta) - \nabla \tilde{Q}_b^G(\beta^*) - D(\beta - \beta^*) \right) \right\|_2 \leq r \left( \sqrt{\frac{p + t}{nb}} + r + b^{y-1} \right) \tag{C.3}
$$

as long as $\sqrt{(p + t)/n} \leq b \leq 1$.

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

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 $n \geq (p \log n)^2 + t$, $\sqrt{(p + t)/n} \leq h \leq ((p + t)/n)^{1/4}$ and $\sqrt{(p + t)/n} \leq b \leq ((p + t)/n)^{1/(2^y)}$. Then, the one-step conquer estimator $\tilde{\beta}$ satisfies the bound

$$
\left\| \tilde{\beta} - \beta^* - D^{-1} \frac{1}{n} \sum_{i=1}^{n} [\tau - G(-e_i/b)] x_i \right\|_2 \leq \left\{ \frac{(p \log n + t)(nb) + b^{y-1}}{n} \right\} \sqrt{\frac{p + t}{n}} \tag{C.4}
$$

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

Theorem C.1 shows that using a higher-order kernel ($y \geq 4$) 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 $(\frac{p \log n}{n})^{1/2} \leq h \leq (\frac{p \log n}{n})^{1/4}$ and $(\frac{p \log n}{n})^{1/2} \leq b \leq (\frac{p \log n}{n})^{1/(2^y)}$. Then,

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

where $\sigma_r^2 = \tau(1 - \tau)\|D^{-1}a\|^2_{\Sigma}$. In particular, with a choice of bandwidth $b \asymp (\frac{p \log n}{n})^{2/(2^y+1)}$,

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

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

C.1 Proof of Proposition C.1

We start from the gradient $\Sigma^{-1/2}\nabla Q_\nu^G(\beta^*) = \mathbb{E}[\mathcal{G}(-\epsilon/b) - \tau]w$ with $w = \Sigma^{-1/2}x$. Let $\mathbb{E}_x$ be the conditional expectation given $x$. By integration by parts,

$$\mathbb{E}_x \mathcal{G}(-\epsilon/b) = \int_{-\infty}^{\infty} \mathcal{G}(t/b) dF_{\epsilon/b}(t) = \int_{-\infty}^{\infty} \mathcal{G}(u) F_{\epsilon/b}(-bu) du.$$  \hspace{1cm} (C.6)

Applying a Taylor expansion with integral remainder on $F_{\epsilon/b}(-bu)$ yields

$$F_{\epsilon/b}(-bu) = F_{\epsilon/b}(0) + \sum_{\ell=0}^{\nu-1} f_{\epsilon/b}(0) \frac{(-bu)\ell}{\ell!} + \frac{(-bu)^{\nu-1}}{(\nu - 2)!} \int_0^1 (1 - w)^{\nu-2} [F_{\epsilon/b}^{(\nu-1)}(-buw) - F_{\epsilon/b}^{(\nu-1)}(0)] dw$$

$$= \tau + \sum_{\ell=0}^{\nu-2} f_{\epsilon/b}(0) \frac{(-bu)^{\ell+1}}{(\ell + 1)!} + \frac{(-bu)^{\nu-1}}{(\nu - 2)!} \int_0^1 (1 - w)^{\nu-2} [f_{\epsilon/b}^{(\nu-1)}(-buw) - f_{\epsilon/b}^{(\nu-1)}(0)] dw.$$

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

$$\mathbb{E}_x \mathcal{G}(-\epsilon/b) = \tau - \frac{b^{\nu-1}}{(\nu - 2)!} \int_{-\infty}^{\infty} \int_0^1 u^{\nu-1} G(u)(1 - w)^{\nu-2} [f_{\epsilon/b}^{(\nu-1)}(-buw) - f_{\epsilon/b}^{(\nu-1)}(0)] dw du.$$

Furthermore, by the Lipschitz continuity of $f_{\epsilon/b}^{(\nu-1)}(\cdot)$ around 0,

$$|\mathbb{E}_x \mathcal{G}(-\epsilon/b) - \tau| \leq \frac{L_{\nu-2} b^\nu}{(\nu - 2)!} \int_{-\infty}^{\infty} \int_0^1 |u^\nu G(u)|(1 - w)^{\nu-2} w dw du$$

$$= B(2, \nu - 1) L_{\nu-2} \kappa_\nu^G b^\nu / (\nu - 2)!,$$

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

$$\|\Sigma^{-1/2}\nabla Q_\nu^G(\beta^*)\|_2 = \sup_{w \in \mathbb{S}^{p-1}} \mathbb{E}_x \mathcal{G}(-\epsilon/b) - \tau |w^\nu u| \leq L_{\nu-2} \kappa_\nu^G b^\nu / \nu!.$$
Turnig to the Hessian, note that

\[ \|\Sigma^{-1/2}\nabla^2 Q_b^G(\beta^*) \Sigma^{-1/2} - D_0\|_2 = \left\| \mathbb{E} \int_{-\infty}^{\infty} G(u) [f_{xi}(-bu) - f_{xi}(0)] du w w^T \right\|_2. \]

Applying a similar Taylor expansion as above, we have

\[ f_{xi}(t) = f_{xi}(0) + \sum_{\ell=1}^{v-1} f^{(\ell)}_{xi}(0) t^\ell / \ell! + \frac{t^{v-1}}{(v-2)!} \int_0^1 (1-w)^{v-2} \{ f^{(v-1)}_{xi}(tw) - f^{(v-1)}_{xi}(0) \} dw. \]  

(C.7)

Under Conditions 1 and 2, it follows that

\[ \|\Sigma^{-1/2}\nabla^2 Q_b^G(\beta^*) \Sigma^{-1/2} - D_0\|_2 \]
\[ \leq \frac{b^{v-1}}{(v-2)!} \sup_{u,\delta \in \mathbb{R}^{v-1}} \mathbb{E} \int_{-\infty}^{\infty} u^{v-1} G(u) (1-w)^{v-2} \{ f^{(v-1)}_{xi}(tw) - f^{(v-1)}_{xi}(0) \} dw du \langle w, u \rangle \langle w, \delta \rangle \]
\[ \leq \frac{b^{v-1}}{(v-1)!} \sup_{u \in \mathbb{R}^{v-1}} \mathbb{E} \langle w, u \rangle^2 \leq \frac{C G}{(v-1)!} b^{v-1}. \]

This completes the proof. \( \square \)

### 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 \( \delta \in \mathbb{B}^p(r) \).

Write \( w_i = \Sigma^{-1/2} x_i \in \mathbb{R}^p \), which are isotropic random vectors, and define

\[ H_n(\delta) = \Sigma^{-1/2} \nabla^2 Q_b^G(\beta) \Sigma^{-1/2} = \frac{1}{n} \sum_{i=1}^{n} G_b(e_i - w_i^T \delta) w_i w_i^T, \quad H(\delta) = \mathbb{E}[H_n(\delta)]. \]  

(C.8)

For any \( \epsilon \in (0, r) \), there exists an \( \epsilon \)-net \( \{ \delta_1, \ldots, \delta_{d_\epsilon} \} \) with \( d_\epsilon \leq (1+2r/\epsilon)^n \) satisfying that, for each \( \delta \in \mathbb{B}^p(r) \), there exists some \( 1 \leq j \leq d_\epsilon \) such that \( \| \delta - \delta_j \|_2 \leq \epsilon \). Hence,

\[ \| H_n(\delta) - H(\delta) \|_2 \]
\[ \leq \| H_n(\delta) - H_n(\delta_j) \|_2 + \| H_n(\delta_j) - H(\delta_j) \|_2 + \| H(\delta_j) - H(\delta) \|_2 \]
\[ =: I_1(\delta) + I_2(\delta_j) + I_3(\delta). \]
For $I_1(\delta)$, note that $G_\theta(u) = (1/b)G(u/b)$ is Lipschitz continuous, i.e. $|G_\theta(u) - G_\theta(v)| \leq L^G b^{-2}|u - v|$ for all $u, v \in \mathbb{R}$. It follows that

$$I_1(\delta) \leq \sup_{u, v \in \mathbb{S}^{p-1}} \frac{1}{n} \sum_{i=1}^{n} |G_\theta(e_i - w_i^\top \delta) - G_\theta(e_i - w_i^\top \delta_j)| \cdot |w_i^\top u \cdot w_i^\top v|$$

$$\leq L^G b^{-2} \sup_{u, v \in \mathbb{S}^{p-1}} \frac{1}{n} \sum_{i=1}^{n} |w_i^\top (\delta - \delta_j) \cdot w_i^\top u \cdot w_i^\top v|$$

$$\leq L^G b^{-2} \epsilon \sup_{u \in \mathbb{S}^{p-1}} \frac{1}{n} \sum_{i=1}^{n} |w_i^\top u|^3. \quad (C.9)$$

Next, we use the standard covering argument to bound $M_{n,3}$. Given $\epsilon_1 \in (0, 1)$, let $N_1$ be an $\epsilon_1$-net of the unit sphere $\mathbb{S}^{p-1}$ with $d_{\epsilon_1} := |N_1| \leq (1 + 2/\epsilon_1)^p$ such that for every $u \in \mathbb{S}^{p-1}$, there exists some $v \in N_1$ satisfying $||u - v||_2 \leq \epsilon_1$. Define the (standardized) design matrix $Z_n = n^{-1/3}(w_1, \ldots, w_n)^\top \in \mathbb{R}^{n \times p}$, so that $M_{n,3} = \sup_{u \in \mathbb{S}^{p-1}} \|Z_n u\|_3^3$. By the triangle inequality,

$$\|Z_n u\|_3 \leq \|Z_n v\|_3 + \|Z_n(u - v)\|_3$$

$$= \|Z_n \delta\|_3 + \left(\frac{1}{n} \sum_{i=1}^{n} |w_i^\top (u - v)|^3\right)^{1/3} \leq \|Z_n v\|_3 + \epsilon_1 M_{n,3}^{1/3}. $$

Taking the maximum over $v \in N_1$, and then taking the supremum over $u \in \mathbb{S}^{p-1}$, we arrive at

$$M_{n,3} \leq (1 - \epsilon_1)^{-3} \tilde{M}_{n,3}, \quad (C.10)$$

where $\tilde{M}_{n,3} = \max_{v \in N_1} (1/n) \sum_{i=1}^{n} |w_i^\top v|^3$.

For every $v \in N_1$, note that $\mathbb{P}(|w_i^\top v|^3 \geq y) \leq 2e^{-y^{2/3}/(2\nu_1)}$ for any $y > 0$. Hence, by inequality (3.6) in Adamczak et al. (2011) with $s = 2/3$, we obtain that for any $\zeta \geq 3$,

$$\frac{1}{n} \sum_{i=1}^{n} |w_i^\top v|^3 \leq \mathbb{E}[|w^\top v|^3] + C_1 \left( \sqrt{\frac{\zeta}{n}} + \frac{\zeta^{3/2}}{n} \right)$$

with probability at least $1 - e^{3-\zeta}$, where $C_1 > 0$ is a constant depending only on $\nu_1$. Taking the union bound over all vectors $v$ in $N_1$ yields that, with probability at least $1 - d_{\epsilon_1} e^{3-\zeta} \geq 1 - e^{3+\log(1+2/\epsilon_1)-\zeta},$

$$\tilde{M}_{n,3} \leq \mu_3 + C_1 \left( \sqrt{\frac{\zeta}{n}} + \frac{\zeta^{3/2}}{n} \right)$$
where \( \mu_3 = \sup_{u \in \mathbb{S}^{p-1}} \mathbb{E}[w^7 u^3] \). Reorganizing the terms, we have

\[
\tilde{M}_{n,3} \leq \mu_3 + C_1 \left[ \sqrt{\frac{p \log(1 + 2/\epsilon_1) + 4 + t}{n}} + \frac{[p \log(1 + 2/\epsilon_1) + 4 + t]^{3/2}}{n} \right] \tag{C.11}
\]

with probability at least \( 1 - e^{-t}/2 \). Hence, taking \( \epsilon = 1/8 \) in (C.10) and (C.11) implies

\[
M_{n,3} \leq 1.5\mu_3 + 1.5C_1 \left\{ \sqrt{\frac{3p + 4 + t}{n}} + \frac{(3p + 4 + t)^{3/2}}{n} \right\}.
\]

Plugging the above bound into (C.9) yields that as long as \( n \geq p + t \),

\[
\sup_{\delta \in \mathbb{B}^p(r)} I_1(\delta) \leq (p + t)^{1/2} b^{-2} \epsilon \tag{C.12}
\]

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

\[
I_3(\delta) \leq L^G b^{-2} \sup_{u, \tilde{w} \in \mathbb{S}^{p-1}} \mathbb{E}[\tilde{w}^7 (\delta - j) \cdot w^7 u \cdot w^7 v] \leq L^G \mu_3 b^{-2} \epsilon \tag{C.13}
\]

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

Turning to \( I_2(\delta_j) \), note that \( H_u(\delta_j) - H(\delta_j) = (1/n) \sum_{i=1}^n (1 - \mathbb{E})\phi_{ij_i} w_i w_j^T \), where \( \phi_{ij} = G_b(\epsilon_i - w_i^T \delta_i) \) satisfy \( |\phi_{ij}| \leq \kappa_u^G b^{-1} \) and

\[
\mathbb{E}(\phi_{ij}^2 | x_i) = \frac{1}{b^2} \int_{-\infty}^{\infty} G^2 \left( \frac{(w_i, \delta_i - j)}{b} \right) f_{x_i}(t) dt = \frac{1}{b^2} \int_{-\infty}^{\infty} G^2(u) f_{x_i}(w_j^T \delta - bu) du \leq \frac{\tilde{m}_G^2}{b}
\]

almost surely, where \( m_G^2 := \int_{-\infty}^{\infty} G^2(u) du < \infty \). Given \( \delta_2 \in (0, 1/2) \), there exits an \( \delta_2 \)-net \( \mathcal{M} \) of the sphere \( \mathbb{S}^{p-1} \) with \( |\mathcal{M}| \leq (1 + 2/\epsilon_2)^p \) such that \( \|H_u(\delta_j) - H(\delta_j)\|_2 \leq (1 - 2\epsilon_2)^{-1} \max_{u \in \mathcal{M}} |w^7 (H_u(\delta_j) - H(\delta_j))|u| \). Given \( u \in \mathcal{M} \) and \( k = 2, 3, \ldots \), we bound the higher order moments of \( \phi_{ij}(w_i^T u)^2 \) by

\[
\mathbb{E}[\phi_{ij}(w_i^T u)^{2k}] \leq \tilde{m}_G b^{-1} \cdot (\kappa_u^G b^{-1})^{k-2} 2^k \cdot 2k \int_0^{\infty} \mathbb{P}(|w^T u| \geq v_1 u) u^{2k-1} du
\]

\[
\leq \tilde{m}_G b^{-1} \cdot (\kappa_u^G b^{-1})^{k-2} 2^k \cdot 4k \int_0^{\infty} u^{2k-1} e^{-u^2/2} du
\]

\[
\leq \tilde{m}_G b^{-1} \cdot (\kappa_u^G b^{-1})^{k-2} 2^k \cdot 2^{k+1} k!.
\]

In particular, \( \mathbb{E}[\phi_{ij}(w_i^T u)^4] \leq 16v_1^4 \tilde{m}_G b^{-1} \), and \( \mathbb{E}[\phi_{ij}(w_i^T u)^2] \leq \frac{16}{2} \cdot 16v_1^4 \tilde{m}_G b^{-1} \cdot (2v_1^2 \kappa_u^G b^{-1})^{k-2} \)
for $k \geq 3$. Applying Bernstein’s inequality and the union bound, we find that for any $u \geq 0$, 

$$
\|H_n(\delta_j) - H(\delta_j)\|_2 
\leq \frac{1}{1 - 2\epsilon_2} \max_{u \in M} \left| \frac{1}{n} \sum_{i=1}^{n} (1 - E)\phi_{ij}(w_i^Tu)^2 \right| 
\leq \frac{2\nu_i^2}{1 - 2\epsilon_2} \left( 2 \sqrt{2 \bar{m}_{G;u} + k_{G;u}^G u / nb} \right)
$$

with probability at least $1 - 2(1 + 2/\epsilon_2)e^{-u} = 1 - (1/2)e^{\log(4)/p log(1 + 2/\epsilon_2) - u}$. Setting $\epsilon_2 = 2/(e^3 - 1)$ 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 $j = 1, \ldots, d_\epsilon$ and setting $v = p(1 + 2r/\epsilon) + t$, we obtain that with probability at least $1 - d_\epsilon e^{-v} \geq 1 - e^{-v/2}$,

$$
\max_{1 \leq j \leq d_\epsilon} I_2(\delta_j) \leq \sqrt{\frac{p \log(3er/\epsilon) + t}{nb}} + \frac{p \log(3er/\epsilon) + t}{nb}.
$$

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

$$
\sup_{\beta \in \Theta^*(r)} \left\| \Sigma^{-1/2} [\nabla^2 \tilde{Q}_G^b(\beta) - \nabla^2 \tilde{Q}_G^b(\beta^*)] \Sigma^{-1/2} \right\|_2 \leq \sqrt{\frac{p \log n + t}{nb}} + \frac{p \log n + t}{nb} + \frac{(p + t)^{1/2}r}{nb^2}
$$

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

\[\square\]

C.3 Proof of Proposition C.3

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

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

35
Recall that $D_0 = \Sigma^{-1/2}D\Sigma^{-1/2} = \mathbb{E}[f_{\text{edge}}(0)ww^T]$. For the first term on the right-hand side, using the mean value theorem for vector-valued functions yields

$$\mathbb{E} \Delta_b(\beta) = \left\{ \Sigma^{-1/2} \int_0^1 \nabla^2 Q^G_b((1-s)\beta^* + s\beta) \, ds \, \Sigma^{-1/2} - D_0 \right\} \Sigma^{1/2}(\beta - \beta^*).$$

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

$$\nabla^2 Q^G_b((1-s)\beta^* + s\beta) = \mathbb{E} \int_{-\infty}^\infty G(u)f_{\text{edge}}(sw^\delta \cdot bu) \, du \cdot xx^T.$$ 

For every $s \in [0, 1]$ and $u \in \mathbb{R}$, it ensures from $f_{\text{edge}}(\cdot)$ being Lipschitz that

$$|f_{\text{edge}}(sw^\delta - bu) - f_{\text{edge}}(-bu)| \leq l_0 s \cdot |w^\delta|.$$ 

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

$$f_{\text{edge}}(-bu) = f_{\text{edge}}(0) + \sum_{\ell=1}^{\nu-1} \frac{f^{(\ell)}_{\text{edge}}(0)(-bu)^\ell}{\ell!} + \frac{(-bu)^{\nu-1}}{(\nu-2)!} \int_0^1 (1-w)^{\nu-2}[f^{(\nu-1)}_{\text{edge}}(-bu) - f^{(\nu-1)}_{\text{edge}}(0)] \, dw.$$ 

Consequently,

$$\left\| \Sigma^{-1/2} \int_0^1 \nabla^2 Q^G_b((1-s)\beta^* + s\beta) \, ds \, \Sigma^{-1/2} - D_0 \right\|_2 \leq \left\| \frac{b^{\nu-1}}{(\nu-2)!} \mathbb{E} \int_{-\infty}^\infty \int_0^1 w^{\nu-1} G(u)(1-w)^{\nu-2}[f^{(\nu-1)}_{\text{edge}}(-bu) - f^{(\nu-1)}_{\text{edge}}(0)] \, dwdu \cdot ww^T \right\|_2$$

$$+ l_0 \frac{b^{\nu-1}}{2} \left\| \mathbb{E} |w^\delta| \cdot ww^T \right\|_2 \leq \frac{C_G b^{\nu-1}}{(\nu-1)!} \sup_{u \in \mathbb{R}^{p-1}} \mathbb{E} \langle w, u \rangle^2 + \frac{l_0}{2} \sup_{u \in \mathbb{R}^{p-1}} \mathbb{E} |w^\delta| \langle w, u \rangle^2$$

$$\leq \left( \frac{C_G b^{\nu-1}}{(\nu-1)!} b^{\nu-1} + 0.5 l_0 \mu_3 \|\delta\|_2 \right).$$

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

$$\sup_{\beta \in \Theta^*(r)} \|\mathbb{E} \Delta_b(\beta)\| \leq \left( \frac{C_G b^{\nu-1}}{(\nu-1)!} b^{\nu-1} + 0.5 l_0 \mu_3 r \right) r \leq (b^{\nu-1} + r) r.$$ 

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 $1 - e^{-t}$,

$$\sup_{\beta \in \Theta^*(r)} ||\Delta_\beta(\beta) - E\Delta_\beta(\beta)||_2 \leq r \sqrt{\frac{p + t}{nb}}$$

as long as $\sqrt{(p + t)/n} \leq b \leq 1$.

Combining the last two displays completes the proof of the claim (C.3). \hfill \qed

### C.4 Proof of Theorem C.1

**Step 1** (Consistency of the initial estimate). First, note that the consistency of the initial estimator $\overline{\beta}$—namely, $\overline{\beta}$ 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 ensuing estimator $\overline{\beta}$ satisfies

$$||\overline{\beta} - \beta^*||_\Sigma \leq r_n \leq \sqrt{\frac{p + t}{n}}$$

(C.15)

with probability at least $1 - 2e^{-t}$ as long as $(\frac{p+t}{n})^{1/2} \leq h \leq (\frac{p+t}{n})^{1/4}$. Provided that the sample Hessian $\nabla^2 Q^G(\overline{\beta})$ is invertible, by the definition of $\overline{\beta}$ we obtain

$$\Sigma^{1/2}(\overline{\beta} - \beta) = -[\Sigma^{-1/2}\nabla^2 Q^G(\overline{\beta})\Sigma^{-1/2}]^{-1}\Sigma^{-1/2}\nabla Q^G(\overline{\beta})$$

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

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

or equivalently,

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

$$- \overline{D}_0^{-1}\Sigma^{-1/2}[\nabla Q^G(\overline{\beta}) - \nabla Q^G(\beta^*) - D(\overline{\beta} - \beta^*)],$$
where $D_0 = E(f_{t\xi}(w)w^T) = \Sigma^{-1/2}D\Sigma^{-1/2}$ and $\tilde{D}_0 := \Sigma^{-1/2}\tilde{Q}_b^G(\tilde{\beta})\Sigma^{-1/2}$. It follows that

$$
\left\| \Sigma^{1/2}(\tilde{\beta} - \beta^*) + \Sigma^{1/2}D^{-1}\tilde{Q}_b^G(\beta^*) \right\|_2 = \left\| \Sigma^{1/2}(\tilde{\beta} - \beta^*) + D_0^{1/2}\tilde{Q}_b^G(\beta^*) \right\|_2 
\leq \left\| I_p - \tilde{D}_0^{-1}D_0 \right\|_2 \left\| \tilde{\beta} - \beta^* \right\|_2 + \left\| \Sigma^{1/2}\tilde{Q}_b^G(\beta^*) \right\|_2
$$

(C.16)

According to the above bound, we need to further control the following three terms:

$$
\left\| I_p - \tilde{D}_0^{-1}D_0 \right\|_2, \left\| \Sigma^{-1/2}\tilde{Q}_b^G(\beta^*) \right\|_2 \text{ and } \left\| \Sigma^{1/2}[\tilde{Q}_b^G(\tilde{\beta}) - \tilde{Q}_b^G(\beta^*) - D(\tilde{\beta} - \beta^*)] \right\|_2.
$$

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

$$
\left\| \tilde{D}_0 - D_0 \right\|_2 \leq \left\| \Sigma^{-1/2}[\nabla^2\tilde{Q}_b^G(\tilde{\beta}) - \nabla^2\tilde{Q}_b^G(\beta^*)] \Sigma^{-1/2} \right\|_2 + \left\| \Sigma^{-1/2}\nabla^2\tilde{Q}_b^G(\beta^*)\Sigma^{-1/2} - D_0 \right\|_2.
$$

Let the bandwidth $b$ satisfy $\max\{\frac{p\log n + t}{n}, n^{-1}\} \leq b \leq 1$. Then, applying Propositions C.1 and C.2 with $r = r_n$ yields that with probability $1 - 3e^{-t}$,

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

Note that, under Condition 3.2, $0 < \lambda_{\min}(D_0) \leq \lambda_{\max}(D_0) \leq \tilde{f}$. For sufficiently large $n$, this implies $\left\| D_0^{-1}\tilde{D}_0 - I_p \right\|_2 \leq \tilde{f}^{-1}\delta_n < 1$, and hence

$$
\left\| D_0^{-1}\tilde{D}_0 - I_p \right\|_2 \leq \delta_n \frac{\delta_n}{\tilde{f} - \delta_n} \text{ and } \left\| \tilde{D}_0^{-1} \right\|_2 \leq \frac{1}{\tilde{f} - \delta_n}.
$$

**Step 3** (Controlling the score). In view of (C.16), it remains to control $\left\| D^{-1/2}\tilde{Q}_b^G(\beta^*) \right\|_2$ and $\left\| \Sigma^{-1/2}[\tilde{Q}_b^G(\tilde{\beta}) - \tilde{Q}_b^G(\beta^*)] - \Sigma^{-1/2}D(\tilde{\beta} - \beta^*) \right\|_2$. For the latter, applying the concentration bound (C.15) and Proposition C.3 yields that, with probability at least $1 - 3e^{-t}$,

$$
\left\| \Sigma^{-1/2}[\tilde{Q}_b^G(\tilde{\beta}) - \tilde{Q}_b^G(\beta^*)] - \Sigma^{-1/2}D(\tilde{\beta} - \beta^*) \right\|_2 \leq \frac{p + t}{n b^{1/2}} + b^{r-1} \sqrt{\frac{p + t}{n}}
$$

(C.17)
as long as \((\frac{p+t}{n})^{1/2} \leq b \leq 1\). Turning to \(\|D^{-1/2}\nabla \hat{Q}_b^G(\beta^*)\|_2\), it follows from Lemma B.1 and Proposition C.1 that with probability at least 1 \(- e^{-t}\),

\[
\|\Sigma^{-1/2}\nabla \hat{Q}_b^G(\beta^*)\|_2 \leq \sqrt{\frac{p + t}{n}} + b'.
\] (C.18)

Finally, combining the bounds (C.15)–(C.18), we conclude that with probability at least 1 \(- 5e^{-t}\),

\[
\|\Sigma^{-1/2}[\nabla \hat{Q}_b^G(\bar{\beta}) - \nabla \hat{Q}_b^G(\beta^*)] - \Sigma^{-1/2}D(\bar{\beta} - \beta^*)\|_2 \leq \left(\sqrt{\frac{p \log n + t}{nb'}} + b'^{-1}\right)\left(\sqrt{\frac{p + t}{n}} + b'\right),
\]

provided that \(\max\{\frac{p \log n + t}{n}, (\frac{p + t}{n})^{1/2}\} \leq b \leq 1\). Under the sample size requirement \(n \geq p (\log n)^2 + t\) and additional constraint \(b \leq (\frac{p + t}{n})^{1/(2\nu)}\), this leads to the claimed bound (C.4). \(\square\)

\section{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 \( \tau \) 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.