Supplement to “Pairwise Difference Estimation of High Dimensional Partially Linear Model”

Fang Han∗, Zhao Ren† and Yuxin Zhu‡

This supplementary material provides notation introduction, additional results, technical challenges of the analysis, and all the technical proofs. For almost all proof subsections in Section A4, we first restate the target theorem or lemma with more explicit dependence among all relevant constants, and then provide the details of its proof.

A1 Notation

Throughout the paper, we define $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{Z}^+$ to be sets of real numbers, integers, and positive integers. For $n \in \mathbb{Z}^+$, write $[n] = \{1, \ldots, n\}$. Let $1_{(\cdot)}$ stand for the indicator function. For arbitrary vectors $v, v' \in \mathbb{R}^p$ and $0 < q < \infty$, we define $\|v\|_0 = \sum_{j=1}^p 1_{(v_j \neq 0)}$, $\|v\|_q^2 = \sum_{j=1}^p |v_j|^q$, and $\langle v, v' \rangle = \sum_{j=1}^p v_j v'_j$. For an arbitrary matrix $\Omega = (\Omega_{ij}) \in \mathbb{R}^{p \times q}$, we define $\|\Omega\|_2 = \max_{i \in [p]} \sum_{j=1}^q |\Omega_{ij}|$. For a symmetric real matrix $\Omega$, let $\lambda_{\min}(\Omega)$ denote its smallest eigenvalue. For a set $S$, we denote $|S|$ to be its cardinality and $S^c$ to be its complement. For a vector $v \in \mathbb{R}^p$ and an index set $\mathcal{S}$, we write $v_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|}$ to be the sub-vector of $v$ of components indexed by $\mathcal{S}$. For a real function $f : \mathcal{X} \to \mathbb{R}$, let $\|f\|_\infty = \sup_{x \in \mathcal{X}} f(x)$. For an arbitrary function $f : \mathbb{R}^k \to \mathbb{R}$, we use $\nabla f = (\nabla_1 f, \ldots, \nabla_k f)^T$ to denote its gradient. For some absolutely continuous random vector $X \in \mathbb{R}^p$, let $f_X$ denote its density function, $F_X$ denote its distribution function, and $\Sigma_X$ denote its covariance matrix. For some joint continuous random vector $(X^T, W)^T \in \mathbb{R}^{p+1}$ and some measurable function $\psi(\cdot) : \mathbb{R}^p \to \mathbb{R}^m$, let $f_{W|\psi(X)}(w, z)$ denote the value of the conditional density of $W = w$ given $\psi(X) = z$. For any two numbers $a, b \in \mathbb{R}$, we define $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. For any two real sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$, or equivalently $b_n \gtrsim a_n$, if there exists an absolute constant $C$ such that $|a_n| \leq C|b_n|$ for any large enough $n$. We write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. We denote $I_p$ to be the $p \times p$ identity matrix for $p \in \mathbb{Z}^+$. Let $c, c', C, C' > 0$ be generic constants, whose actual values may vary from place to place.

In addition, we write $\mathcal{B}_2^p = \{x \in \mathbb{R}^p : \|x\|_2 \leq 1\}$ and $\mathcal{S}_2^{p-1} = \{x \in \mathbb{R}^p : \|x\|_2 = 1\}$. Let $e_j \in \mathbb{R}^p$ be a vector that has 1 at the $j$-th position, and 0 elsewhere.

∗Department of Statistics, University of Washington, Seattle, WA 98195, USA; e-mail: faghan@uw.edu
†Department of Statistics, University of Pittsburgh, Pittsburgh, PA 15260, USA; email: zren@pitt.edu
‡Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA; e-mail: yuzhu@jhsph.edu
A2 Additional Results

A2.1 Examples satisfying Assumption 5

Example A2.1. Suppose function $g : \mathbb{R} \to \mathbb{R}$ is piecewise $(M_L, \alpha)$-Hölder for some $\alpha \in (0, 1]$, and have discontinuity points $a_1, \ldots, a_J$ with jump size bounded in absolute value by $C_g$, for positive absolute constants $M_L$ and $C_g$. Also suppose $|f_W(w)| \leq M$ for some positive absolute constant $M$. Consider set

$$A = \bigcup_{j=1}^{J} \left\{ (-\infty, a_j] \times [a_j, +\infty) \cup [a_j, +\infty) \times (-\infty, a_j] \right\}$$

and consider box kernel function $K(w) = \mathbb{I}(|w| \leq 1/2)$. Then

$$|g(w_1) - g(w_2)| \leq (J + 1)M_L \cdot |w_1 - w_2|^\alpha + JC_g \mathbb{I}\{(w_1, w_2) \in A\},$$

for any $w_1, w_2 \in \mathbb{R}$, and that

$$\mathbb{E} \left[ \frac{1}{h} K \left( \frac{W_{ij}}{h} \right) \mathbb{I}\{(W_i, W_j) \in A\} \right] = \frac{2}{h} \sum_{j=1}^{J} \int_{a_j}^{a_j} \int_{-\infty}^{+\infty} \mathbb{I}\{|w_1 - w_2| \leq h/2\} f_W(w_1)f_W(w_2) dw_1 dw_2 \leq \frac{JM^2}{4} h.$$ 

Thus we have verified two equations in Assumption 5 with $M_g = (J + 1)M_L$, $M_d = JC_g$, and $M_a = JM^2/4$.

Example A2.2. Suppose $W \sim \text{Unif}[0, 1]$, kernel function $K(w) = \mathbb{I}(w \in [-1/2, 1/2])$, and

$$g(w) = \begin{cases} w, & w \in [0, 1/2), \\ w + 1, & w \in [1/2, 1]. \end{cases}$$

Suppose $h \leq 1/2$ and consider a slightly different set $A = \{[1/4, 1/2] \times [1/2, 3/4] \cup [1/2, 3/4] \times [1/4, 1/2] \}$ than that in Example A2.1. One can easily check that $|g(w_1) - g(w_2)| \leq 3|w_1 - w_2| + \mathbb{I}\{(w_1, w_2) \in A\}$, and

$$\mathbb{E} \left[ \frac{1}{h} K \left( \frac{W_{ij}}{h} \right) \mathbb{I}\{(W_i, W_j) \in A\} \right] = h/4.$$ 

Thus we have verified two equations in Assumption 5 with $M_g = 3$, $M_d = 1$, and $M_a = 1/4$.

A2.2 Extending results to heavy-tailed noise

Corollary A2.1. Assume that there exist some absolute constants $K_1, C_0 > 0$ and $1/(2 + \epsilon) < \xi < 3/4$, such that

$$h_n \in [K_1(\log p/n)^{1/2}, C_0] \quad \text{and} \quad n \geq C\{(\log p)^{5/(3-4\xi)} + (\log p)^3 \vee q^{4/3}(\log p)^{1/3} \vee q(\log p)^2\},$$

where $K_1(\log p/n)^{1/2} < C_0$, and the quantity $q$ and the dependence of constant $C$ will be specified case by case below. Denote $\eta_n = \|\mathbb{E}[\hat{X}\hat{X}^T | \hat{W} = 0]\|_\infty$. We then have, replacing Assumption 12 with Assumption 17 in corresponding results, the following assertions are still true. Also all
three positive constants $C', c, c'$ that have different values in specific cases, but only depend on $M, M_K, C_0, \kappa_x, \kappa_\ell, M_\ell, \xi, C$.

(1) Analogue of Theorem 3.1: Assume Assumption 14 holds with $\gamma = 1$. Set $q = s$. Assume further that $\lambda_n \geq C\{h_n + (\log p/n)^{1/2}\}$, where $C$ only depends on $M, M_K, C_0, \kappa_x, M_u, \kappa_\ell, M_\ell, \xi, \epsilon, \zeta, K_1$. Then under Assumptions 6-11, 14, 15, and 17, we have

$$\mathbb{P}(\|\hat{\beta}_h - \beta^*\|^2 \leq C's\lambda_n^2) \geq 1 - c \exp(-c' \log p) - c \exp(-c' \log n) - \epsilon_n.$$

(2) Analogue of Theorem 3.2: Assume Assumption 14 holds with a general $\gamma \in (0,1]$. Set $q = s$. Assume further that $\lambda_n \geq C\{h_n + (\log p/n)^{1/2}\}$, where $C$ only depends on $M, M_K, C_0, \kappa_x, M_u, \kappa_\ell, M_\ell, \xi, \epsilon, \zeta, \gamma, K_1$. Then under Assumptions 6-11, 14, 15, and 17, we have

$$\mathbb{P}(\|\hat{\beta}_h - \beta^*\|^2 \leq C's\lambda_n^2) \geq 1 - c \exp(-c' \log p) - c \exp(-c' \log n) - \epsilon_n.$$

(3) Analogue of Theorem 3.3: Assume Assumption 14 holds with a general $\gamma \in [1/4, 1]$. Set $q = s + nh_n^{2\gamma}/\log p$. Assume further that $\lambda_n \geq C\{h_n + (\log p/n)^{1/2}\}$, where $C$ only depends on $M, M_K, C_0, \kappa_x, M_u, \kappa_\ell, M_\ell, \xi, \epsilon, \gamma, K_1$. Then under Assumptions 6-8, 10-11, 14-16, and 17, we have

$$\mathbb{P}(\|\hat{\beta}_h - \beta^*\|^2 \leq C'(s\lambda_n^2 + s \log p/n + n\lambda_n^2h_n^{2\gamma}/\log p)) \geq 1 - c \exp(-c' \log p) - c \exp(-c' \log n) - \epsilon_n.$$

(4) Analogue of Theorem 2.3:

a. Assume that $g(\cdot)$ is $\alpha$-Hölder for $\alpha \geq 1$, and $g(\cdot)$ has compact support when $\alpha > 1$. Set $q = s$. Assume further that $\lambda_n \geq C\{h_n + (\log p/n)^{1/2}\}$ and $n \geq (\log p)^4$, where $C$ only depends on $M, M_K, C_0, \kappa_x, M_u, \kappa_\ell, M_\ell, \xi, \epsilon, \zeta, K_1$, and Hölder parameters of $g(\cdot)$. Then under Assumptions 6-8, 9', 10-11, 13 and 17, we have

$$\mathbb{P}(\|\hat{\beta}_h - \beta^*\|^2 \leq C's\lambda_n^2) \geq 1 - c \exp(-c' \log p) - c \exp(-c' \log n).$$

b. Assumption 5 holds with $\alpha \in (0,1]$. Set $q = s$. Assume further that $\lambda_n \geq C\{h_n + (\log p/n)^{1/2}\}$ and $n \geq C(\log p)^4$, where $C$ only depends on $M, M_K, C_0, \kappa_x, M_u, \kappa_\ell, M_\ell, \xi, \epsilon, K_1, M_g, M_d, M_a$, and $\gamma = \alpha$ if $M_dM_a = 0$, $\gamma = \alpha \wedge 1/2$ otherwise. Then under Assumptions 6-8, 9', 10-11, 13 and 17, we have

$$\mathbb{P}(\|\hat{\beta}_h - \beta^*\|^2 \leq C's\lambda_n^2) \geq 1 - c \exp(-c' \log p) - c \exp(-c' \log n).$$

c. Assume Assumption 5 holds with $\alpha \in [1/4, 1]$. Set $q = s + nh_n^{2\gamma}/\log p$. Assume further that $\lambda_n \geq C\{h_n + (\log p/n)^{1/2}\}$ and $n \geq C(\log p)^4$, where $C$ only depends on $M, M_K, C_0, \kappa_x, M_u, \kappa_\ell, M_\ell, \xi, \epsilon, K_1, M_g, M_d, M_a$ and $\gamma = \alpha$ if $M_dM_a = 0$, $\gamma = \alpha \wedge 1/2$ if otherwise. Then under Assumptions 6-8, 9', 10-11, 13 and 17,

$$\mathbb{P}(\|\hat{\beta}_h - \beta^*\|^2 \leq C'(s\lambda_n^2 + s \log p/n + n\lambda_n^2h_n^{2\gamma}/\log p)) \geq 1 - c \exp(-c' \log p) - c \exp(-c' \log n).$$

(5) Analogue of Theorem 2.2: Set $q = s$. Assume that $\lambda_n \geq C\{h_n + (\log p/n)^{1/2}\}$ and $n \geq C(\log p)^4$, where $C$ depends only on $M, M_K, C_0, \kappa_x, M_u, \kappa_\ell, M_\ell, \xi, \epsilon, K_1, M_g$. Then under
Assumptions 6-11, 4, and 17, we have
\[ \mathbb{P}(\|\hat{\beta}_{hn} - \beta^*\|_2^2 \leq C's\lambda_n^2) \geq 1 - c\exp(-c' \log p) - c\exp(-c' \log n). \]

A3 Technical challenges of the analysis

The main results of the paper, including Theorems 3.1, 3.2, 3.3, 2.2, as well as Theorem 2.3, are all based on the general framework introduced in Section 2.1. For this, one major object of interest is to verify the empirical RE condition (Assumption 3 in Section 2.1) based on the population RE conditions such as Assumption 9 and its variant Assumption 16. This result is formally stated in Corollary A3.1 at the end of this section. The proof follows the standard reduction principle in Rudelson and Zhou (2013) applied to Theorem A3.1, the proof of which rests on several advanced U-statistics exponential inequalities (Giné et al., 2000; Houdré and Reynaud-Bouret, 2003) and nonasymptotic random matrix analysis tools specifically tailored for U-matrices (Vershynin, 2012; Mitra and Zhang, 2014), and thus deserves a discussion.

We start with a definition of the restricted spectral norm (Han and Liu, 2016). For an arbitrary \( p \times p \) real matrix \( M \) and an integer \( q \in [p] \), the \( q \)-restricted spectral norm \( \|M\|_{2,q} \) of \( M \) is defined to be
\[ \|M\|_{2,q} := \max_{v \in \mathbb{R}^p, \|v\|_0 \leq q} \left| v^T M v / v^T v \right|. \]

As pointed in the seminal paper Rudelson and Zhou (2013), the empirical RE condition, i.e., Assumption 3, is closely related to the \( q \)-restricted spectral norm of Hessian matrix for the loss function regarding a special choice of \( q \). Our proof relies on a study of this \( q \)-restricted spectral norm.

In Assumption 3, letting \( \hat{\Gamma}_n(\theta, h_n) = \hat{L}_n(\beta, h_n) \), simple algebra yields
\[ \delta \hat{L}_n(\Delta, h_n) = \Delta^T \left\{ \binom{n}{2}^{-1} \sum_{i<j} \frac{1}{h_n} K(\tilde{W}_{ij}/h_n) \tilde{X}_{ij} \tilde{X}_{ij}^T \right\} \Delta = \Delta^T \hat{T}_n \Delta. \]

Note that \( \hat{T}_n \) is a random U-matrix, namely, a random matrix formulated as a matrix-valued U-statistic. As was discussed in the previous sections, \( h_n \) is usually picked to be of the order \( (\log p/n)^{1/2} \), rendering a large bump as \( \tilde{W}_{ij} \) is close to zero. Consequently, when \( h_n \) is set in the regime of interest, the variance of the kernel \( g_\Delta(D_i, D_j) = h_n^{-1} K(\tilde{W}_{ij}/h_n)(\tilde{X}_{ij} \Delta)^2 \) will explode at the rate of \( (n/\log p)^{1/2} \), leading to a loose and sub-optimal bound when using Bernstein inequality for non-degenerate U-statistics (see, e.g., Proposition 2.3(a) in Arcones and Gine (1993)). Thus a more careful study of this random U-matrix \( \hat{T}_n \) is need.

The next theorem gives a concentration inequality for \( \hat{T}_n \) under the \( q \)-restricted spectral norm.

**Theorem A3.1.** For some \( q \in [p] \), suppose there exists some absolute constant \( C > 0 \) such that
\[ n \geq C \cdot \left[ \left\{ q^{4/3}(\log p)^{1/3} \vee q(\log p)^2 \right\} + \log(1/\alpha) \right]. \]

Then under Assumptions 7, 8, and 11, with probability at least \( 1 - \alpha \),
\[ \|\hat{T}_n - \mathbb{E}\hat{T}_n\|_{2,q} \leq C' \cdot \left[ \frac{q(\log p)^{1/4}}{n^{3/4}} + \frac{q(\log p)^2}{n} + \frac{\log(1/\alpha)}{n} \right]^{1/2}, \]
where $C'$ is a positive constant only depending on $M,M_K,C_0,\kappa_x,C$.

The proof of Theorem A3.1 follows the celebrated Hoeffding’s decomposition. However, there are two major challenges. On one hand, different from most existing investigations on nonasymptotic random matrix theory, the first order term of $\delta\hat{L}_n(\Delta, h_n)$, after decomposition, does not have a natural product structure, namely, it cannot be written as $n^{-\frac{1}{2}}\sum_{i=1}^{n}U_iU_i^\top$ for some independent random vectors $\{U_i \in \mathbb{R}^p, i \in [n]\}$. Hence, we cannot directly follow those well-established arguments based on a natural product structure, but have to resort to properties of the kernel. To this end, we state the following two auxiliary lemmas, which are repeatedly used in the proofs, and can be regarded as extensions to the classic results in, for example, Robinson (1988).

**Lemma A3.2.** Assume random variables $W \in \mathbb{R}$ and $Z \in \mathcal{Z}$, such that

$$
\left| \frac{\partial f_W|Z(w,z)}{\partial w} \right| \leq M_1,
$$

for some positive constant $M_1$ with any $z$ in the range of $Z$ and any $w$ in the range of $W$. Also, let $K(\cdot)$ be a kernel function such that $\int_{-\infty}^{+\infty} |w|K(w)\,dw \leq M_2$ for some constant $M_2 > 0$. Then we have for any $h > 0$,

$$
\left| \mathbb{E}\left[ \frac{1}{h} K\left( \frac{W}{h} \right) Z \right] - \mathbb{E}[Z|W = 0]f_W(0) \right| \leq M_1 M_2 \mathbb{E}[\|Z\|]h.
$$

**Lemma A3.3.** Let $(W_1, Z_1), (W_2, Z_2) \in \mathbb{R} \times \mathcal{Z}$ be i.i.d.. Assume

$$
\left| \frac{\partial f_W|Z_1(w,z)}{\partial w} \right| \leq M_1
$$

holds for some positive constant $M_1$ with any $z$ in the range of $Z_1$ and any $w$ in the range of $W$. Let $K(\cdot)$ be a kernel function such that $\int_{-\infty}^{+\infty} |w|K(w)\,dw \leq M_2$ for some constant $M_2 > 0$. Let $\varphi: \mathcal{Z}^2 \to \mathbb{R}$ be a measurable function. Then we have for any $h > 0$,

$$
\left| \mathbb{E}\left[ \frac{\varphi(Z_1,Z_2)}{h} K\left( \frac{W_1-W_2}{h} \right)|W_2, Z_2 \right] - \mathbb{E}[\varphi(Z_1,Z_2)|W_1=W_2,W_2,Z_2] f_{W_1}(W_2) \right|
\leq M_1 M_2 \mathbb{E}[\|\varphi(Z_1,Z_2)\|Z_2]\,h.
$$

On the other hand, the second order term of $\delta\hat{L}_n(\Delta, h_n)$, after decomposition, forms a degenerate U-statistic, and requires further study. To control this term, one might consider using the two-term Bernstein inequality for degenerate U-statistics (see, e.g., Proposition 2.3(c) in Arcones and Giné (1993) or Theorem 4.1.2 in de la Peña and Giné (2012)). But it will add an additional polynomial $\log p$ multiplicity term in the upper bound. Instead, we adopt the sharpest four-term Bernstein inequality discovered by Giné et al. (2000), get rid of several inexplicit terms (e.g., the $\ell_2 \to \ell_2$ norm), and formulate it into the following user-friendly tail inequality. We state this result in the following auxiliary lemma. The constants here are able to be explicitly calculated thanks to Houdré and Reynaud-Bouret (2003).

**Lemma A3.4.** Let $Z_1, \ldots, Z_n, Z \in \mathcal{Z}$ be i.i.d., and $g: \mathcal{Z}^2 \to \mathbb{R}$ be a symmetric measurable function with $\mathbb{E}[g(Z_1,Z_2)] < \infty$. Write $U_n(g) = \sum_{i<j} g(Z_i,Z_j)$ and $f(z) = \mathbb{E}[g(Z,z)]$. Let

$$
B_g = \|g\|_{\infty}, B_f = \sup_{Z_2} \mathbb{E}\left[\|g(Z_1,Z_2)\|Z_2\right], \text{ and } \sigma^2 = \mathbb{E}[g(Z_1,Z_2)^2].
$$
In addition, denote \( B^2 = n \sup_{Z_2} \mathbb{E}[g(Z_1, Z_2)^2 | Z_2] \). We then have
\[
\mathbb{P}( |U_n(g) - \mathbb{E}[U_n(g)]| \geq t + C_1 n \sigma u^{1/2} + C_2 B_f u + C_3 B u^{3/2} + C_4 B g u^2 ) 
\leq 2 \exp \left( \frac{-t^2/n^2}{8n \mathbb{E}[f(Z_2)^2] + 4B_f \cdot t/n} \right) + C_5 e^{-u},
\] (A3.1)
where we take positive absolute constants
\[
C_1 = 2(1 + \epsilon)^{3/2},
\]
(A3.2)
\[
C_2 = 8\sqrt{2}(2 + \epsilon + \epsilon^{-1}),
\]
\[
C_3 = \epsilon(1 + \epsilon^{-1})^2(5/2 + 32 \epsilon^{-1}) + [\{2\sqrt{2}(2 + \epsilon + \epsilon^{-1}) \} \vee (1 + \epsilon)^2/\sqrt{2}],
\]
\[
C_4 = \{ 4\epsilon(1 + \epsilon^{-1})^2(5/2 + 32 \epsilon^{-1}) \} \vee 4(1 + \epsilon)^2/3,
\]
\[
C_5 = 2.77,
\]
for any \( \epsilon > 0 \). For cases that \( f(z) = 0 \) (corresponding to the degenerate case), \( t \) can be set as zero and the first term on the second line of (A3.1) can be eliminated.

Combining Theorem A3.1 with Theorem 10 and the follow-up arguments in Rudelson and Zhou (2013), we immediately have the following corollary, which verifies the desired empirical RE condition corresponding to different situations. Note that Assumption 9’ is stronger than both Assumption 9 and its variant Assumption 16. Thus the results below still hold when Assumption 9’ is imposed in Section 2.2.2.

**Corollary A3.1.** Suppose Assumptions 6-8 and 10-11 are satisfied.

1. Assume Assumption 9 holds, and that
   \[
n \geq C \{ s^{4/3}(\log p)^{1/3} \vee s(\log p)^2 \},
   \]
   for some constant \( C > 0 \) only depending on \( M, M_K, C_0, \kappa_x, \kappa_\ell, M_\ell \). Then we have
   \[
   \mathbb{P} \left[ \delta \hat{L}_n(\Delta, h_n) \geq \frac{\kappa_\ell M_\ell}{4} \| \Delta \|_2^2 \text{ for all } \Delta \in \{ \Delta' \in \mathbb{R}^p : \| \Delta S^c \|_1 \leq 3 \| \Delta S \|_1 \} \right] 
   \geq 1 - c \exp(-c' \log p) - c \exp(-c'n),
   \]
   where \( c, c' \) are positive constants only depending on \( M, M_K, C_0, \kappa_x, \kappa_\ell, M_\ell, C \).

2. Assume Assumption 16 holds, and that
   \[
n \geq C \{ s + nh_n^{2\gamma}/(\log p)^{4/3}(\log p)^{1/3} \vee s + nh_n^{2\gamma}/(\log p)(\log p)^2 \},
   \]
   for some constant \( C > 0 \) only depending on \( M, M_K, C_0, \kappa_x, \kappa_\ell, M_\ell, \zeta, \gamma \). Then we have
   \[
   \mathbb{P} \left\{ \delta \hat{L}_n(\Delta, h_n) \geq \frac{\kappa_\ell M_\ell}{4} \| \Delta \|_2^2 \text{ for all } \Delta \in \mathbb{C}_{\hat{S}_n} \right\} 
   \geq 1 - c \exp(-c' \log p) - c \exp(-c'n),
   \]
   where \( \mathbb{C}_{\hat{S}_n} := \{ v \in \mathbb{R}^p : \| v_{\mathcal{J}} \|_1 \leq 3 \| v_{\mathcal{J}} \|_1 \text{ for some } \mathcal{J} \subset [p] \text{ and } |\mathcal{J}| \leq s + \zeta^2 nh_n^{2\gamma}/(\log p) \} \), and \( c, c' \) are positive constants only depending on \( M, M_K, C_0, \kappa_x, \kappa_\ell, M_\ell, C \).
A4 Technical proofs

A4.1 Proof of Theorem 2.1

Proof. By (2.3), we have
\[ \|\hat{\theta}^*_n - \theta^*\|^2 \leq \rho_n^2. \]
So it suffices to show that
\[ \|\hat{\theta}_n - \hat{\theta}^*_n\|^2 \leq 9\tilde{s}_n\lambda_n^2/\kappa_1^2 \]
holds with probability at least \(1 - \epsilon_1, n - \epsilon_2, n\) whenever \(\lambda_n \leq \kappa_1 r/3\tilde{s}_n^{1/2}\). We split the rest of the proof into two main steps.

**Step I.** Denote \(\tilde{\Delta} = \hat{\theta}_n - \hat{\theta}^*_n\). Recall definition of sets \(\hat{S}_n\) and \(C_{\hat{S}_n}\), and further define function \(\mathcal{F}(\Delta) = \hat{\Gamma}_n(\hat{\theta}^*_n + \Delta, h_n) - \hat{\Gamma}_n(\hat{\theta}^*_n, h_n) + \lambda_n(\|\hat{\theta}^*_n + \Delta\|_1 - \|\hat{\theta}^*_n\|_1).\)

For the first step, we show that if \(\mathcal{F}(\Delta) > 0\) for all \(\Delta \in C_{\hat{S}_n} \cap \{\Delta' \in \mathbb{R}^p : \|\Delta'\|_2 = \eta\}\), then \(\|\tilde{\Delta}\|_2 \leq \eta\). To this end, we first show that
\[ \tilde{\Delta} = \mathcal{F}(\Delta) = \mathcal{S}_{\hat{S}_n}. \] (A4.1)
Applying triangle inequality and some algebra, we obtain
\[ \|\hat{\theta}^*_n + \Delta\|_1 - \|\hat{\theta}^*_n\|_1 \geq \|\mathcal{S}_{\hat{S}_n}\|_1 - \|\mathcal{S}_{\hat{S}_n}\|_1. \] (A4.2)
We also have, with probability at least \(1 - \epsilon_1, n\),
\[ \hat{\Gamma}_n(\hat{\theta}^*_n + \Delta, h_n) - \hat{\Gamma}_n(\hat{\theta}^*_n, h_n) \geq (\nabla \hat{\Gamma}_n(\hat{\theta}^*_n, h_n), \Delta) \]
\[ \geq -\|\nabla \hat{\Gamma}_n(\hat{\theta}^*_n, h_n)\|_\infty \cdot \|\Delta\|_1 \]
\[ \geq -\frac{\lambda_n}{2}(\|\mathcal{S}_{\hat{S}_n}\|_1 + \|\mathcal{S}_{\hat{S}_n}\|_1), \] (A4.3)
where the first inequality is by convexity of \(\hat{\Gamma}_n(\theta, h)\) in \(\theta\) as assumed in Assumption 3, the second is by Hölder’s inequality, and the last is by Assumption 2. Combining (A4.2) and (A4.3), and using the fact that \(\mathcal{F}(\tilde{\Delta}) \leq 0\), we have
\[ 0 \geq \frac{\lambda_n}{2}(\|\mathcal{S}_{\hat{S}_n}\|_1 - 3\|\mathcal{S}_{\hat{S}_n}\|_1), \]
thus proving (A4.1).

Next, we assume that \(\|\tilde{\Delta}\|_2 > \eta\). Then, because \(\tilde{\Delta} \in C_{\hat{S}_n}\) and \(C_{\hat{S}_n}\) is star-shaped, there exists some \(t \in (0, 1)\), such that \(t\tilde{\Delta} \in C_{\hat{S}_n} \cap \{\Delta' \in \mathbb{R}^p : \|\Delta'\|_2 = \eta\}\). However, by convexity of \(\mathcal{F}(\cdot)\),
\[ \mathcal{F}(t\tilde{\Delta}) \leq t\mathcal{F}(\tilde{\Delta}) + (1 - t)\mathcal{F}(0) = t\mathcal{F}(\tilde{\Delta}) \leq 0. \]
By contradiction, we complete the proof of the first step.

**Step II.** For the second step, we show that under Assumptions 1-3, we have \(\mathcal{F}(\Delta) > 0\) for all \(\Delta \in C_{\hat{S}_n} \cap \{\Delta' \in \mathbb{R}^p : \|\Delta'\|_2 = \eta\}\), for some appropriately chosen \(\eta\), and then complete the proof.

Combining Assumptions 2, 3, and (A4.2), for any \(\Delta \in C_{\hat{S}_n} \cap \{\Delta' \in \mathbb{R}^p : \|\Delta'\|_2 = \eta\}\), where we take \(\eta = 3\tilde{s}_n^{1/2}\lambda_n/\kappa_1\), and \(\lambda_n \leq \kappa_1 r/(3\tilde{s}_n^{1/2})\) so that \(\eta \leq r\), we have that with probability at least
\[ F(\Delta) \geq \langle \nabla \hat{f}_n(\hat{\theta}_{\hat{h}_n}, h), \Delta \rangle + \kappa_1 \| \Delta \|_2^2 + \lambda_n(\| \hat{\theta}_{\hat{h}_n} + \Delta \|_1 - \| \hat{\theta}_{\hat{h}_n} \|_1) \]
\[ \geq -\| \nabla \hat{f}_n(\hat{\theta}_{\hat{h}_n}, h_n) \|_\infty \cdot \| \Delta \|_1 + \kappa_1 \| \Delta \|_2^2 + \lambda_n(\| \Delta S_\circ \|_1 - \| \Delta \|_1) \]
\[ \geq -\lambda_n \| \Delta \|_1/2 + \kappa_1 \| \Delta \|_2^2 + \lambda_n(\| \Delta S_\circ \|_1 - \| \Delta \|_1) \]
\[ \geq \kappa_1 \| \Delta \|_2^2 - 3\lambda_n \tilde{s}_n^{1/2} \| \Delta \|_2/2, \]
where the first inequality is by Assumption 3, the second is by H"older's inequality and (A4.2), the third is by Assumption 2, and the last is due to the fact that \( \| \Delta S_\circ \|_1 \leq \tilde{s}_n^{1/2} \| \Delta \|_2 \leq \tilde{s}_n^{1/2} \| \Delta \|_2. \)

Then we have
\[ F(\Delta) \geq \kappa_1 \eta^2 - 3\tilde{s}_n^{1/2} \lambda_n \eta/2 = 9\tilde{s}_n \lambda_n^2/(2\kappa_1) > 0, \]
which, using result from Step I, implies that \( \| \hat{\Delta} \|_2^2 \leq \eta^2 = 9\tilde{s}_n \lambda_n^2/\kappa_1^2. \)

Combining with Assumption 2, we have
\[ \| \hat{\theta}_{\hat{h}_n} - \theta^* \|_2^2 \leq \frac{18\tilde{s}_n \lambda_n^2}{\kappa_1^2} + 2\rho_n^2, \]
with probability at least \( 1 - \epsilon_{1,n} - \epsilon_{2,n}. \) This completes the proof of Theorem 2.1. \( \square \)

A4.2 Proof of Theorem 3.1

In the sequel, with a slight abuse of notation, we use an equivalent representation of Assumption 15 for writing
\[ P\{ |U_k - E[U_k]| \leq A(\log(np)/n)^{1/2}, \text{ for all } k \in [p] \} \geq 1 - \epsilon_n \]
to replace (3.3), noting that we assume \( p > n. \) Hereafter we also slight abuse of notation and do not distinguish \( \log(np)/n \) from \( \log p/n. \)

**Theorem A4.1 (Theorem 3.1).** Assume Assumption 14 holds with \( \gamma = 1. \) Further assume \( h_n \geq K_1(\log(np)/n)^{1/2} \) for positive absolute constant \( K_1, \) and assume \( h_n \leq C_0 \) for positive constant \( C_0. \)

We also take \( \lambda_n \geq 4(A + A')(\log(np)/n)^{1/2} + 8\kappa_2 M_\varphi h_n, \) where
\[ A' = \{16\sqrt{3}(1 + c)^{1/2} M_1 + 4\sqrt{3}C_1(1 + c)^{1/2} M_1^{1/2} K_1^{-1/2} + 8C_2(1 + c) + 8C_3(1 + c) M_1^{1/2} M_1^{1/2} K_1^{-1/2} + 8C_4(1 + c)^2 M_1 K_1^{-1} + 8M_1(c + 2)\} \kappa_1 \kappa_2, \]
for positive absolute constant \( c, M_1 = M + MM_1 C_0, \) and \( C_1, \ldots, C_4 \) as defined in (A3.2). Suppose
we have 
\[ n > \max \left\{ 64(c + 2)^2(c + 1)(\log(np))^3/3, 3, \right\} \]

\[ \frac{48\sqrt{6}M_K\kappa^2q}{K_1p\{\log(np)\}^{1/2}} \left( \frac{2^{10} \cdot 6 \cdot \sqrt{6}M_f\kappa^2q}{\kappa\mu p} \right)^{2/3} \left( \frac{144\kappa^4}{K_1^2p^2\log(np)} \right)^{1/3}, \]

\[ \left[ \frac{2^{11} \cdot 6 \cdot \sqrt{3}(2 + c)^{1/2}C_1M_K^{1/2}M_f^{1/2}\kappa^2}{K_1^{1/2}\kappa_\ell M_f} \right]^{4/3} \cdot q^{4/3}\{\log(np)\}^{1/3}, \]

\[ \left[ \frac{2^8 \cdot 6 \cdot (20 + 7.5c)(2 + c)C_2M_f^{3/2}}{\kappa_\ell M_f} \right]^{1/2} \cdot q^{1/2}\log(np), \]

\[ \left[ \frac{2^8 \cdot 6(c + 2)^{3/2}C_3\{144(2 + c)^2M_KM_f\kappa^4\kappa_\ell^{-1} + 192M_f^{4\kappa^4} + 8M_f^{4\kappa^4}\}^{1/2}}{\kappa_\ell M_f} \right]^{4/3} q^{4/3}\{\log(np)\}^{2/3}, \]

\[ \left[ \frac{2^{10} \cdot 6 \cdot \sqrt{6}(2 + c)^3C_4\kappa^2}{K_1^{1/2}\kappa_\ell M_f} \right]^{2/3} q^{2/3}\{\log(np)\}^{5/3}, \]

\[ \left[ \frac{2^{11} \cdot 6 \cdot (20 + 7.5c)(2 + c)M_f^{3/2}}{\kappa_\ell M_f} \right] q\{\log(np)\}^{2}, \]

\[ \left[ \frac{2^{20}\{(3M_f^2\kappa^2 + 2M_f^{2C_0^2}\kappa^2) \vee 2M\}^{\kappa^2}}{(\kappa_\ell M_f)^2 \wedge (16\kappa_\ell M_f)^2} \right] q \log \left( \frac{6ep}{q} \right), \]

\[ \frac{2^{24}K_1^{2}M_f^{2M_f^{2\kappa^2}\log(np)}}{(\kappa_\ell M_f)^2} \],

where \( q = 2305s \). Then under Assumptions 6-12, 14-15, we have

\[ \| \beta_{h_n} - \beta^* \|_2^2 \leq \frac{288s\lambda_p}{M_f^2\kappa_f^2}, \]

with probability at least \( 1 - 12.54\exp(-c\log p) - 2\exp(-c'\kappa) - \epsilon_n \cdot p \), where \( c' = (\kappa_\ell^2 M_f^2 \wedge 64\kappa_\ell M_f)^{1/2} \), \( \lambda_p \geq 4(A + A') \{\log(np)/n\}^{1/2} + 8\kappa^2 M_f \lambda p \}

\[ \right[ \left[ \frac{2^{16}\{(3M_f^2\kappa^2 + 2M_f^{2C_0^2}\kappa^2) \vee 2M\}^{\kappa^2}}{(\kappa_\ell M_f)^2 \wedge (16\kappa_\ell M_f)^2} \right] q \log \left( \frac{6ep}{q} \right), \]

Proof. See Proof of Theorem 3.2. \( \square \)

### A4.3 Proof of Theorem 3.2

**Theorem A4.2** (Theorem 3.2). Assume Assumption 14 holds with a general \( \gamma \in (0, 1] \). Further assume \( h_n \geq K_1\{\log(np)/n\}^{1/2} \) for positive absolute constant \( K_1 \), and assume \( h_n \leq C_0 \) for positive constant \( C_0 \). We also take \( \lambda_n \geq 4(A + A')\{\log(np)/n\}^{1/2} + 8\kappa^2 M_f \lambda p \}

\[ A' = \left\{ 16\sqrt{3}(1 + c)^{1/2}M_f + 4\sqrt{3}C_1(1 + c)^{1/2}M_f^{1/2}K_1^{-1/2} + 8C_2(1 + c) \right. \]

\[ + 8C_3(1 + c)^{3/2}M_f^{1/2}K_1^{1/2}K_1^{-1/2} + 8C_4(1 + c)^{2}M_KK_1^{-1} + 8M_f(c + 2)K_1^{-1} \],

where \( \kappa_u \) and \( \kappa_\ell \) are positive constants.
for positive absolute constant \(c\), \(M_f = M + MM_K C_0\), and \(C_1, \ldots, C_4\) as defined in (A3.2). Suppose we have
\[
n > \max \left\{ 64(c + 2)^2(c + 1) \{ \log(np) \}^3 / 3, 3 \right\},
\]
\[
\frac{48 \sqrt{6} M_K \kappa_2 q}{K_1 p \{ \log(np) \}^{1/2}} \left( \frac{2^{10} \cdot 6 \cdot \sqrt{6} M_f \kappa_2^2 q}{\kappa_\ell M_p} \right)^{2/3} \frac{144 \kappa_4^2}{K_0^2 p^2 \log(np)},
\]
\[
\left[ \frac{2^{11} \cdot 6 \cdot \sqrt{3}(2 + c)^{1/2} C_1 M_K^{1/2} M_f^{1/2} \kappa_2^2}{K_1^{1/2} \kappa_\ell M_\ell} \right]^{1/3} \cdot q^{4/3} \{ \log(np) \}^{1/3},
\]
\[
\frac{2^8 \cdot 6 \cdot (20 + 7.5c)(2 + c) C_2 M_f \kappa_4^2}{\kappa_\ell M_\ell} \cdot q^{1/2} \log(np),
\]
\[
\frac{2^8 \cdot 6(c + 2)^{3/2} C_3 \{ 144(2 + c)^2 M_K M_f \kappa_4^4 K_1^{-1} + 192 M_f^2 \kappa_4^4 + 8 M_f \kappa_2^4 \}^{1/2}}{\kappa_\ell M_\ell} \cdot q^{4/3} \{ \log(np) \}^{1/3},
\]
where \(q = 2305s\). Then under Assumptions 6-12, 14-15, we have
\[
\| \hat{\beta}_{hn} - \beta^* \|^2 \leq \frac{288s \lambda_n^2}{M_f^2 \kappa_\ell^2},
\]
with probability at least \(1 - 12.54 \exp(-c \log p) - 2 \exp(-c' n) - \epsilon_n \cdot p\), where \(c' = (\kappa_\ell^2 M_f^2 \wedge 64 \kappa_\ell M_\ell) / [2^{16} \{ 3 M_f^2 \kappa_2^2 + 2 M_f^2 M_K^2 C_0^2 \kappa_2^2 \} \vee 2 M_1 \kappa_4^2].
\]

**Proof.** We adopt the framework as described in Section 2.1 for \(\theta^* = \beta^*\), \(\Gamma_0(\theta) = L_0(\beta)\), \(\tilde{\Gamma}_n(\theta, h) = \tilde{L}_n(\beta, h)\), and take \(\hat{\theta}_{hn} = \beta^*\), which yields \(s_n \leq s\) and \(p_n = 0\).

In addition to (3.2), denote
\[
U_{1k} = \left( \frac{n}{2} \right)^{1} \sum_{i<j} \frac{1}{h_{ij}} K \left( \frac{\tilde{W}_{ij}}{h_{ij}} \right) \tilde{X}_{ij} \tilde{u}_{ij},
\]
\[
U_{2k} = \left( \frac{n}{2} \right)^{1} \sum_{i<j} \frac{1}{h_{ij}} K \left( \frac{\tilde{W}_{ij}}{h_{ij}} \right) \tilde{X}_{ij} \tilde{X}_{ij}^T (\beta_{hn} - \beta^*),
\]
and observe that
\[
| \nabla_k \tilde{L}_n(\beta^*) | \leq 2 \left\{ |U_{1k} - \mathbb{E}[U_{1k}]| + |U_k - \mathbb{E}[U_k]| + |\mathbb{E}[U_{2k}]| \right\}, \tag{A4.4}
\]
where \(U_k\) is defined in (3.2). Apply Lemma A4.21 on \(D_i = (X_{ik}, u_i, W_i)\), with conditions of lemma satisfied by Assumptions 7, 8, 11 and 12, and then we have
\[
\mathbb{P} \{ |U_{1k} - \mathbb{E}[U_{1k}]| \geq A' \log(np)/n \} \leq 6.77 \exp \{ -(c + 1) \log p \}, \tag{A4.5}
\]
for positive absolute constant $c$, and $A'$ as defined in (A4.48), and when $n > \max\{16(c + 2)^2(c + 1)\{\log(np)\}^{3/3}, 3\}$.

Apply Lemma A3.2 on $Z = |\tilde{X}_{ijk} \tilde{X}_{ij}^T (\beta^*_h - \beta^*)|$, with conditions of lemma satisfied by Assumptions 7, 8, 11, and 14, and then we have

$$
|E[U_2]| \leq E[|\tilde{X}_{ijk} \tilde{X}_{ij}^T (\beta^*_h - \beta^*)|][W = 0] + MMKc_0E[|\tilde{X}_{ijk} \tilde{X}_{ij}^T (\beta^*_h - \beta^*)|] 
\leq 2\kappa_x^2(M + MMKc_0)\zeta h_n^2.
$$

(A4.6)

Combining (A4.4)-(A4.6), and Assumption 15, we have

$$
\Pr\{\text{for any } k \in [p], |\nabla k \hat{L}_n(\beta^*)| \leq (2A + 2A')\{\log(np)/n\}^{1/2} + 4\kappa_x^2(M + MMKc_0)\zeta h_n^2\} 
\geq 1 - 6.77 \exp(-c \log p) - p \cdot \epsilon_n,
$$

for positive absolute constant $c$, and when we appropriately take $n$ bounded from below. Assume $\lambda_n \geq 4(A + A')\{\log(np)/n\}^{1/2} + 8\kappa_x^2(M + MMKc_0)\zeta h_n^2$, which verifies Assumption 2.

We verify Assumption 3 by applying Corollary A3.1, and complete the proof by Theorem 2.1.

\textbf{A4.4 Proof of Theorem 3.3}

\textbf{Theorem A4.3} (Theorem 3.3). Assume Assumption 14 holds with a general $\gamma \in [1/4, 1]$. Further assume $h_n \geq K_1\{\log(np)/n\}^{1/2}$ for positive absolute constant $K_1$, and assume $h_n \leq C_0$ for positive constant $C_0$. We also take $\lambda_n \geq 4(A'' + A + M\eta_n)\{\log(np)/n\}^{1/2} + 8MMKc_0^{1/2}\kappa_x^2h_n$, where

$$
A'' = \{16\sqrt{3}Mf(1 + c)^3 + 4\sqrt{3}C_1M_{1/2}K_1^{-1/2}(1 + c)^2 + 8C_2(1 + c) + 8C_3M_{1/2}M_{1/2}K_1^{-1/2}K_1^{-1/2}(1 + c)^2 + 8C_4MK_1^{-1}(1 + c)^2 + 8Mf(c + 2)\} \cdot (\kappa_x \kappa_u + C\kappa_x^2)
$$

$$
\eta_n = \|E[\tilde{X} \tilde{X}^T|W = 0]\|_\infty.
$$
Here, \( C_1, \ldots, C_4 \) are as defined in (A3.2), \( C > \zeta^2 C_0^2 \) and \( c > 0 \) are some absolute constants, and \( M_f = M + MM_RC_0 \). Suppose we have

\[
 n > \max \{ (C - \zeta^2 C_0^2) s \log(np), 64(c + 2)^2 (c + 1) \{ \log(np) \}^3 / 3, 3, \\
 \frac{48 \sqrt{6} M_K \kappa^2 q}{K_1 p \{ \log(np) \}^{1/2}}, \frac{2^{10} \cdot 6 \cdot \sqrt{6} M_f \kappa^2 q}{k_\ell M_p}, \frac{144 \kappa^2}{K_1^2 p^2 \log(np)}, \\
 \frac{2^{11} \cdot 6 \cdot \sqrt{9}(2 + c)^{1/2} C_1 M_K^{1/2} M_f^{1/2} \kappa^2}{K_1^{1/2} k_\ell M_\ell}, \frac{4^{1/3} \cdot q^{4/3} \{ \log(np) \}^{1/3}}, \\
 \frac{2^{8} \cdot 6 \cdot (20 + 7.5c)(2 + c) C_2 M_f \kappa^2}{\kappa_\ell M_\ell}, \frac{q^{1/2} \log(np)}, \\
 \frac{2^{20} \{ (3M^2 \kappa_x^2 + 2M^2 M_K^2 C_0^2 \kappa_x^2 ) \vee 2M \} \kappa^2}{(\kappa_\ell M_\ell)^2}, q \log \left( \frac{6p}{q} \right), \\
 \frac{2^{24} K_1^2 M^2 M_K^2 \kappa_\ell^2 \log(np)}{(\kappa_\ell M_\ell)^2}
\]

where \( q = 2305 \{ s + \zeta^2 nh_n^{2\gamma} / \log(np) \} \). Then under Assumptions 6-8, 10-12, 14-16, we have

\[
\| \hat{\beta}_{\ell_n} - \beta^* \|_2 \leq \frac{288 s^2 \lambda^2}{M_\ell \kappa_\ell^2} + \frac{2s \log(np)}{n} + \left\{ \frac{288 n \lambda^2}{M_\ell^2 \kappa_\ell^2 \log(np)} + 2 \right\} \cdot \zeta^2 h_n^{2\gamma},
\]

with probability at least \( 1 - 19.31 \exp(-c \log p) - 2 \exp(-c'n) - \epsilon_n \cdot p \), where \( c' = (\kappa_\ell^2 M_\ell^2 \wedge 64 \kappa_\ell M_\ell) / [2^{16} \{ (3M^2 \kappa_x^2 + 2M^2 M_K^2 C_0^2 \kappa_x^2 ) \vee 2M \} \kappa_x^2] \).

**Proof.** We adopt the framework as described in Section 2.1 for \( \theta^* = \beta^* \), \( \Gamma_0(\theta) = L_0(\beta) \), \( \bar{\Gamma}_n(\theta, h) = \bar{L}_n(\beta, h) \).

We take \( \hat{\beta}_{\ell_n} = \beta^* \) such that, for each \( j \in [p] \),

\[
\hat{\beta}_{\ell_n,j} = \begin{cases} 
\beta^*_j & \text{if } |\beta^*_{\ell_n,j}| > \{ \log(np) / n \}^{1/2}; \\
0, & \text{if otherwise.}
\end{cases}
\]

Then under Assumption 14, we have

\[
\rho_n^2 \leq s \log(np) / n + \zeta^2 h_n^{2\gamma},
\]

\[
s_n \leq s + \frac{\zeta^2 nh_n^{2\gamma}}{\log(np)}.
\]

We verify Assumption 2 by applying Lemma A4.4 below with \( A'' = A' + A'' \), verify Assumption 3 by applying Corollary A3.1 (2) under Assumption 16, and complete the proof by Theorem 2.1.

\[\square\]
Lemma A4.4. Assume $h_n \geq K_1\{\log(np)/n\}^{1/2}$ for positive absolute constant $K_1$, and assume $h_n \leq C_0$ for positive constant $C_0$. Denote $\eta_n = \|E[\tilde{X}\tilde{X}^T | \tilde{W} = 0]\|_{\infty}$. We also take $\lambda_n \geq 4(A' + A'' + A + M\eta_n)\{\log(np)/n\}^{1/2} + 8M_M C^{1/2} \kappa_2^2 h_n$, where $A'$ and $A''$ are as specified in (A4.48), and $C > \zeta^2 C_0^{2\gamma}$ is some positive absolute constants. Suppose we have

$$n > \max \{(C - \zeta^2 C_0^{2\gamma})s \log(np), 64(c + 2)^2(c + 1)\{\log(np)\}^3/3, 3\},$$

for positive absolute constant $c > 0$. Then under Assumptions Assumptions 6-8, 10-12, 14-15, we have

$$P\{2|\nabla_k \tilde{L}_n(\beta^{*}_{h_n}, h_n)| \leq \lambda_n \text{ for all } k \in [p]\} \geq 1 - 13.54 \exp(-c \log p) - c_n \cdot p.$$ 

Proof of Lemma A4.4. In addition to (3.2), denote

$$U_{1k} = \left(\frac{n}{2}\right)^{-1} \sum_{i<j} \frac{1}{h_n} K\left(\frac{W_{ij}}{h_n}\right) \tilde{X}_{ijk} \tilde{u}_{ij},$$

$$U_{2k} = \left(\frac{n}{2}\right)^{-1} \sum_{i<j} \frac{1}{h_n} K\left(\frac{W_{ij}}{h_n}\right) \tilde{X}_{ijk} \tilde{X}_i^T (\beta^{*} - \beta^{*}_{h_n}),$$

$$U_{3k} = \left(\frac{n}{2}\right)^{-1} \sum_{i<j} \frac{1}{h_n} K\left(\frac{W_{ij}}{h_n}\right) \tilde{X}_{ijk} \tilde{X}_j^T (\beta^{*}_{h_n} - \beta^{*}_{h_n}),$$

and observe that

$$|\nabla_k \tilde{L}_n(\beta^{*}_{h_n}, h_n)| \leq 2||U_{1k} - E[U_{1k}]| + |U_{2k} - E[U_{2k}]| + |U_{k} - E[U_k]| + |E[U_{3k}]|),$$

(A4.9)

where in decomposing the left hand side, we have utilized the fact that $E[\nabla_k \tilde{L}_n(\beta^{*}_{h_n}, h_n)] = 0$. Result of (A4.44) holds, thus bounding $|U_{1k} - E[U_{1k}]|$, i.e.,

$$P\{|U_{1k} - E[U_{1k}]| \geq A'' \{\log(np)/n\}^{1/2} \} \leq 6.77 \exp\{-c \log p\}. \quad (A4.10)$$

We bound the rest of the components on the right hand side of the last display.

We have $\|\beta^{*} - \beta^{*}_{h_n}\|^2 \leq s \log(np)/n + \zeta^2 h_n^{2\gamma} < C$ for some positive absolute constant $C > \zeta^2 C_0^{2\gamma}$, when $n > (C - \zeta^2 C_0^{2\gamma})s \log(np)$. Apply Lemma A4.21 on $D_i = (X_{ik}, X_i^T (\beta^{*} - \beta^{*}_{h_n}), W_i)$, with conditions of lemma satisfied by Assumptions 7, 8, 11, and that $\|\beta^{*} - \beta^{*}_{h_n}\|^2 < C$, and we have

$$P\{|U_{2k} - E[U_{2k}]| \geq A' \{\log(np)/n\}^{1/2} \} \leq 6.77 \exp\{-c \log p\}, \quad (A4.11)$$

for positive constants $A'$ and $c$, and when we assume $n > \max \{64(c + 2)^2(c + 1)\{\log(np)\}^3/3, 3\}$. Here, $A'$ is as specified in (A4.48).

Apply Lemma A3.3 with conditions of lemma satisfied by Assumptions 7 (Lemma A4.15) and 8 (Lemma A4.16), and we have

$$|E[U_{3k}]| \leq ME[|\tilde{X}_{ijk} \tilde{X}_j^T (\beta^{*}_{h_n} - \beta^{*}_{h_n})|] E[|\tilde{W}_{ij}|] + MM_K h_n E[|\tilde{X}_{ijk} \tilde{X}_j^T (\beta^{*}_{h_n} - \beta^{*}_{h_n})|]$$

$$\leq M\eta_n \{\log(np)/n\}^{1/2} + MM_K C^{1/2} \cdot 2\kappa_2^2 h_n$$

(A4.12)

where the second inequality is due to Cauchy-Schwarz and Assumption 11 (Lemmas A4.17 and A4.18).
Combining (A4.9)-(A4.12) and Assumption 15, we have
\[ \mathbb{P}\{ \text{for any } k \in [p], |\nabla_k \tilde{L}_n(\tilde{\beta}_n^*, h_n)| \leq 2(A' + A'' + A + M\eta_n)\{ \frac{\log(np)}{n} \}^{1/2} + 4MMK^{1/2}\kappa_x^2 h_n \} \]
\[ \geq 1 - 13.54p \exp\{- (c + 1) \log p\} - \epsilon_n \cdot p, \]
for positive absolute constant c, and when we appropriately take n bounded from below. Here A' and
A'' are as specified in (A4.48). Assume \( \lambda_n \geq 4(A' + A'' + A + M\eta_n)\{ \log(np)/n \}^{1/2} + 8MMK^{1/2}\kappa_x^2 h_n. \)
This completes the proof. \( \Box \)

### A4.5 Proof of Theorem 3.4

**Theorem A4.5** (Theorem 3.4). Assume \( h \leq C_0 \) for positive constant \( C_0 \), and that \( h^2 \leq \kappa_\ell M_\ell \cdot (4MMK\kappa_x^2)^{-1} \). Under Assumptions 6-8, 9', 10-11, and 13, and when \( g \) is \((L, \alpha)\)-Hölder for \( \alpha \geq 1 \)
\( (g \) has bounded support when \( \alpha > 1 \)), we have
\[ \| \beta_n^* - \beta^* \|_2 \leq \zeta h, \]
where
\[ \zeta = \max \left\{ 4 \cdot \left( \frac{L_\alpha^2 MMK + MMK\mathbb{E}n^2/2}{\kappa_\ell M_\ell} \right)^{1/2}, \frac{16\kappa_x(M + MMK\kappa_x^2)^{1/2} \cdot L_\alpha^2 MMK}{\kappa_\ell M_\ell} \right\}, \]
where \( L_\alpha \) is the Lipschitz constant for \( g \) \( (L_\alpha = L \) when \( \alpha = 1 \)).

**Proof.** Refer to Proof of Theorem 3.5 when \( g \) is \((L, 1)\)-Hölder, taking \( M_g = L \) and \( d = M_a = 0 \), in which case Assumption 5 is not needed. Note that higher-order Hölder with compact support implies \((L, 1)\)-Hölder. Thus we complete the proof. \( \Box \)

### A4.6 Proof of Theorem 3.5

**Theorem A4.6.** Assume \( h \leq C_0 \) for positive constant \( C_0 \), and that \( h^2 \leq \kappa_\ell M_\ell \cdot (4MMK\kappa_x^2)^{-1} \).
Under Assumptions 5, 6-8, 9', 10-11, and 13, we have
\[ \| \beta_n^* - \beta^* \|_2 \leq \zeta h^\gamma, \]
where
\[ \zeta = \max \left\{ 4 \cdot \left( \frac{M_\alpha^2 MMK C_0^{2\alpha-2\gamma} + M_d^2 M_a C_0^{1-2\gamma} + MMK\mathbb{E}n^2 C_0^{2-2\gamma}/2}{\kappa_\ell M_\ell} \right)^{1/2}, \frac{16\kappa_x(M + MMK\kappa_x^2)^{1/2} \cdot (M_g^2 MMK C_0^{2\alpha-2\gamma} + M_d^2 M_a C_0^{1-2\gamma})^{1/2}}{\kappa_\ell M_\ell} \right\}, \]
\( \gamma = \alpha \) if \( M_d M_a = 0 \), and \( \gamma = \min \{ \alpha, 1/2 \} \) if otherwise.

**Proof of Theorem 3.5.** We prove the lemma in three steps.

**Step I.** We show that \( |L_0(\beta_n^*) - L_0(\beta^*)| \) is lower bounded for \( L_0(\beta) = \mathbb{E}[\mathbb{I} - \tilde{X}^T \beta]^2|\tilde{W} = 0] f_\tilde{W}(0) \). By Assumptions 10 and 9', we have
\[ \lambda_{\min} \left( \frac{\partial^2 L_0(\beta)}{\partial \beta^2} \right) = 2\lambda_{\min} \left( \mathbb{E}[\tilde{X}^2|\tilde{W} = 0] \right) f_\tilde{W}(0) \geq 2\kappa_\ell M_\ell. \]
Therefore, for some \( \beta_t = \beta_{h_\alpha} + t(\beta^* - \beta_{h_\alpha}^*), \ t \in [0,1], \) we have
\[
L_0(\beta_h^*) - L_0(\beta^*) = \frac{1}{2} (\beta_h^* - \beta^*)^T \frac{\partial^2 L_0(\beta)}{\partial \beta^2} \bigg|_{\beta = \beta_t} (\beta_h^* - \beta^*) \geq \kappa_t M_t \| \beta_h^* - \beta^* \|^2.
\]

**Step II.** We show that \( |L_{h_\alpha}(\beta) - L_0(\beta)| \) is upper bounded. Observe that
\[
|L_{h_\alpha}(\beta) - L_0(\beta)| \leq \left| \mathbb{E} \left[ \frac{1}{h} K \left( \frac{\nabla}{h} \right) \{ \widetilde{X}(\beta - \beta^*) \} \right]^2 - \mathbb{E} \left[ \{ \widetilde{X}(\beta - \beta^*) \} \widetilde{W} = 0 \right] f_{\widetilde{W}}(0) \right| \\
+ \mathbb{E} \left[ \frac{1}{h} K \left( \frac{\nabla}{h} \right) \{ g(W_i) - g(W_j) \} \right]^2
\]
\[
+ \left| \mathbb{E} \left[ \frac{1}{h} K \left( \frac{\nabla}{h} \right) \widetilde{u} \right]^2 - \mathbb{E} \left[ \widetilde{u}^2 \widetilde{W} = 0 \right] f_{\widetilde{W}}(0) \right| \\
+ 2 \mathbb{E} \left[ \frac{1}{h} K \left( \frac{\nabla}{h} \right) \widetilde{X}(\beta - \beta^*) \{ g(W_i) - g(W_j) \} \right].
\]

And we bound each component on the right hand side of above inequality.

By Taylor’s expansion, we have
\[
\left| \mathbb{E} \left[ \frac{1}{h} K \left( \frac{\nabla}{h} \right) \{ \widetilde{X}(\beta - \beta^*) \} \right]^2 - \mathbb{E} \left[ \{ \widetilde{X}(\beta - \beta^*) \} \widetilde{W} = 0 \right] f_{\widetilde{W}}(0) \right| \\
= \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K \left( \frac{\nabla}{h} \right) v^2 f_{\widetilde{W} | \widetilde{X}(\beta - \beta^*)} (w,v) \, dw \, dF_{\widetilde{X}(\beta - \beta^*)} (v) \\
- \int_{-\infty}^{\infty} v^2 f_{\widetilde{W} | \widetilde{X}(\beta - \beta^*)} (0,v) \, dF_{\widetilde{X}(\beta - \beta^*)} (v) \right| \\
= \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K \left( \frac{\nabla}{h} \right) v^2 \left\{ f_{\widetilde{W} | \widetilde{X}(\beta - \beta^*)} (wh,v) - f_{\widetilde{W} | \widetilde{X}(\beta - \beta^*)} (0,v) \right\} \, dw \, dF_{\widetilde{X}(\beta - \beta^*)} (v) \right| \\
= \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K \left( \frac{\nabla}{h} \right) v^2 \left\{ \frac{\partial f_{\widetilde{W} | \widetilde{X}(\beta - \beta^*)} (w,v)}{\partial w} \right|_{(0,v)} \, dw \, dF_{\widetilde{X}(\beta - \beta^*)} (v) \right| \\
+ \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K \left( \frac{\nabla}{h} \right) v^2 \left\{ \frac{\partial^2 f_{\widetilde{W} | \widetilde{X}(\beta - \beta^*)} (w,v)}{\partial w^2} \right|_{(wh,v)} \, dw \, dF_{\widetilde{X}(\beta - \beta^*)} (v) \right|,'
Using an identical argument, by Assumptions 7, 8 (Lemmas A4.15 and A4.16), and finite second moment assumption \( E[\tilde{u}^2] < \infty \), we have

\[
E \left[ \frac{1}{h} K \left( \frac{\tilde{W}}{h} \right) \tilde{u}^2 \right] - E[\tilde{u}^2]E[\tilde{W} = 0] f_{\tilde{W}}(0) \leq M M_K E[\tilde{u}^2] h^2. \tag{A4.15}
\]

By Assumption 5, we have

\[
\begin{align*}
&\left. E \left[ \frac{1}{h} K \left( \frac{\tilde{W}}{h} \right) \{ g(W_i) - g(W_j) \} \right] \right| \\
&\leq 2 M_s^2 E \left[ \frac{1}{h} K \left( \frac{\tilde{W}}{h} \right) \tilde{W}^{2 \alpha} \right] + 2 M_d^2 E \left[ \frac{1}{h} K \left( \frac{\tilde{W}}{h} \right) \delta \{ (W_i, W_j) \in A \} \right] \\
&\leq 2 M_s^2 E \left[ \frac{1}{h} K \left( \frac{\tilde{W}}{h} \right) \tilde{W}^{2 \alpha} \right] + 2 M_d^2 M_a h,
\end{align*}
\]

where

\[
E \left[ \frac{1}{h} K \left( \frac{\tilde{W}}{h} \right) \tilde{W}^{2 \alpha} \right] = \int_{-\infty}^{\infty} K(w)|w|^{2 \alpha} h^{2 \alpha} f_{\tilde{W}}(wh) \, dw \leq M M_K h^{2 \alpha}.
\]

Therefore, we have

\[
E \left[ \frac{1}{h} K \left( \frac{\tilde{W}}{h} \right) \{ g(W_i) - g(W_j) \} \right] \leq 2 M_s^2 M M_K h^{2 \alpha} + 2 M_d^2 M_a h. \tag{A4.16}
\]

By (A4.14), (A4.16), and applying Hölder’s inequality, we also have

\[
\begin{align*}
&\left. \left| E \left[ \frac{1}{h} K \left( \frac{\tilde{W}}{h} \right) X_{ij}^T(\beta - \beta^*) \{ g(W_i) - g(W_j) \} \right] \right| \right| \\
&\leq E \left[ \frac{1}{h} K \left( \frac{\tilde{W}}{h} \right) \{ X_{ij}^T(\beta - \beta^*) \} \right]^{1/2} \cdot E \left[ \frac{1}{h} K \left( \frac{\tilde{W}}{h} \right) \{ g(W_i) - g(W_j) \} \right]^{1/2} \tag{A4.17}
\end{align*}
\]

\[
\begin{align*}
&\leq (2 M M_K \gamma^2 \| \beta - \beta^* \|^2 2^2 h^2 + 2 \gamma^2 \| \beta - \beta^* \|^2 2^2 M) \| \beta - \beta^* \|^2 2^2 M_a h \| \beta - \beta^* \|^2 2^2 h^2 \\
&\leq a_1 \| \beta - \beta^* \|^2 2^2 h^2,
\end{align*}
\]

where \( \gamma = \alpha \) if \( M_d M_a = 0 \), and \( \gamma = \min \{ \alpha, 1/2 \} \) if otherwise, and \( a_1 = 2 \gamma^2 (M + M M_K C_0^2)^{1/2} \cdot (M^2 h M M_K C_0^2)^{1/2} \cdot (M^2 h M M_K C_0^2)^{1/2} \).

Combining (A4.13)-(A4.17), we have

\[
|L_h(\beta) - L_0(\beta)| \leq 2 a_1 \| \beta - \beta^* \|^2 2^2 h^2 + a_2 h^{2 \gamma} + a_3 \| \beta - \beta^* \|^2 2^2 h^2,
\]

where \( a_1 = 2 M_s^2 M M_K C_0^2 + 2 M_d^2 M_a C_0^2 + M M_K E[u^2 C_0^{2 - 2 \gamma}] \), and \( a_3 = 2 M M_K \gamma^2 \).

**Step III.** We combine Step I and Step II, and verify Assumption 14. Using results from Step I and Step II, we have

\[
\kappa_\ell M \ell \| \beta^*_h - \beta^* \|^2 2 \leq L_0(\beta^*_h) - L_0(\beta^*)
\]

\[
= L_0(\beta^*_h) - L_h(\beta^*_h) + L_h(\beta^*) - L_0(\beta^*) + L_h(\beta^*_h) - L_h(\beta^*)
\]

\[
\leq |L_0(\beta^*_h) - L_h(\beta^*_h)| + |L_h(\beta^*) - L_0(\beta^*)|
\]

\[
\leq 2 a_1 \| \beta^*_h - \beta^* \|^2 2^2 h^2 + 2 a_2 h^{2 \gamma} + a_3 \| \beta^*_h - \beta^* \|^2 2^2 h^2.
\]

When \( h^2 \leq \kappa_\ell M \ell / (2 a_3) \), we have

\[
\kappa_\ell M \ell \| \beta^*_h - \beta^* \|^2 2 \leq 4 a_1 \| \beta^*_h - \beta^* \|^2 2^2 h^2 + 4 a_2 h^{2 \gamma},
\]

\[
16
\]
which further implies that
\[ \|\beta_h^* - \beta^*\|_2 \leq \max \left\{ \left( \frac{8a_2}{\kappa_\ell M_\ell} \right)^{1/2}, \frac{8a_1}{\kappa_\ell M_\ell} \right\} \cdot h^\gamma. \]
This completes the proof. \(\square\)

**A4.7 Proof of Theorem 2.3**

**Theorem A4.7** (Theorem 2.3). Assume \( h_n \geq K_1 \{ \log(np)/n \}^{1/2} \) for positive absolute constant \( K_1 \), and assume \( h_n \leq C_0 \) for positive constant \( C_0 \). We denote \( c \) to be some positive absolute constant, 
\[ c' = \left( \kappa^2_\ell M^2_\ell \wedge 64 \kappa_\ell M_\ell \right) / \left[ 2^{16} \{ (3M^2\kappa^2_\ell + 2M^2M^2_\ell \kappa^2_0 \kappa^2_\ell) \vee 2M \} \kappa^2_\ell \right], \]
\[ M_f = M + MM_K C_0, \]
and \( C_1, \ldots, C_4 \) as defined in (A3.2) Also denote 
\[ \tau_1 = \sqrt{2}(2 + c)^{1/2} \kappa_x K_1^{-1} (BM_K C_0^a + D M_K), \]
\[ \tau_2 = \sqrt{2}(2 + c)^{1/2} \kappa_x \{ BM_K M(1 + C_0) C_0^a + D M_f \}, \]
\[ \tau_3 = 4M^2_\ell M^2 \cdot (BC_0^a + D)^2 \cdot (1 + C_0^2) \cdot \kappa^2_x, \]
\[ \tau_4 = \left\{ 4B^2 M M_K \kappa^2_x (1 + C_0) C_0^{2a-\gamma_1} + 2D^2 \cdot (12 M_f \kappa^4_\ell)^{1/2} \cdot E^{1/2} C_0^{-1/2-\gamma_1} \right\} \cdot M_K K_1^{-1}, \]
\[ \tau_5 = 4(2 + c) \kappa^2_x \{ BM M_K (1 + C_0) C_0^{2a} + D^2 M_f \} M_K K_1^{-1}, \]
and
\[ A' = \left\{ 16\sqrt{3} M_f (1 + c)^{1/2} + 4 \sqrt{3} C_1 M^3_\ell K_1^{-\frac{1}{2}} (1 + c)^{1/2} + 8C_2 (1 + c) + 8C_3 M^2_\ell M^2_f K_1^{-\frac{1}{2}} (1 + c)^{3/2} \right\} \cdot (\kappa_x \kappa_a + C \kappa^2_x) \]
\[ A'' = 4\tau_3^{1/2} (1 + c)^{1/2} + 2C_1 \tau_4^{1/2} (1 + c)^{1/2} + 2C_2 \tau_2 (1 + c) + 2C_3 \tau_5^{1/2} (1 + c)^{3/2} \]
\[ + 2C_4 \tau_1 (1 + c)^2 + 4M_f \cdot (BC_0^a + D) \cdot (c + 2) \kappa_x, \]

17
where $\gamma_1 = \min \{2\alpha - 1, -1/2\}$. Consider lower bound on $n$,

$$
n > \max \left\{ 64(c + 2)^2(c + 1)\{\log(np)\}^3/3, 64(c + 2)^2(c + 1)\tau_2^2\tau_3^{-1}\{\log(np)\}^4, \{\log(np)\}^{5/3}, 3, \frac{48\sqrt{6}MK\kappa_x^2q}{K_1p\{\log(np)\}^{1/2}}, \left(\frac{2^{10} \cdot 6 \cdot \sqrt{6}Mf\kappa_x^2q}{\kappa_\ell M_p}\right)^2/3, \frac{144\kappa_x^4}{K_1^2\tau^2 p^2\log(np)}, \left[\frac{211 \cdot 6 \cdot \sqrt{3}(2 + c)^{1/2}C_1M_1^{1/2}M_2^{1/2}\kappa_x^2}{K_1^{1/2}\kappa_\ell M_\ell}\right]^{4/3} \cdot q^{4/3}\{\log(np)\}^{1/3}, \left[\frac{2^{8} \cdot 6 \cdot (20 + 7.5c)(2 + c)C_2M_\ell\kappa_x^2}{\kappa_\ell M_\ell}\right]^{1/2} \cdot q^{1/2}\log(np), \left[\frac{2^{8} \cdot 6(c + 2)^{1/2}C_3\{144(2 + c)^2MKM_\ell\kappa_x^4K_1^{-1} + 192M_2^2\kappa_x^4 + 8M_\ell^2\kappa_x^2\}^{1/2}}{\kappa_\ell M_\ell}\right]^{4/3} q^{4/3}\{\log(np)\}^{1/3}, \frac{2^{10} \cdot 6 \cdot \sqrt{6}(2 + c)^3C_4\kappa_x^2}{K_1\kappa_\ell M_\ell} \cdot q^{2/3}\{\log(np)\}^{5/3}, \frac{211 \cdot 6 \cdot (20 + 7.5c)(c + 2)M_\ell\kappa_x^2}{\kappa_\ell M_\ell} q\{\log(np)\}^2, \frac{2^{6} \cdot 3q}{(20 + 7.5c)M_\ell^2\kappa_x^2\log(np)}, \frac{2^{20}\{\{3M_\ell^2\kappa_x^2 + 2M_2^2\kappa_x^2\}^2 / 2 M_\ell^2\kappa_x^2\} q\log\left(\frac{6ep}{q}\right)}{(\kappa_\ell M_\ell)^2 \cdot (16\kappa_\ell M_\ell)^2}, \frac{2^{24}K_1^2M_\ell^2\kappa_x^4\log(np)}{(\kappa_\ell M_\ell)^2} \right\}.$$

(A4.18)

Here, $q$, $B$, $D$, $E$ and $a$ are to be specified in different cases. Suppose that Assumptions 6-8, 9', 10-12, and 13 hold.

1. Assume that $g$ is $(L, \alpha)$-Hölder for $\alpha \geq 1$, and $g$ has bounded support when $\alpha > 1$. Also suppose (A4.18) holds with $q = 2305s$. We take $B = L_\alpha$, where $L_\alpha$ is the Lipschitz constant for $g$ ($L_\alpha = L$ when $= 1$), $D = E = 0$, $a = 1$, and assume $\lambda_n \geq 4(A^\theta + A')\{\log(np)/n\}^{1/2} + 8\kappa_x^2M_\ell^2\zeta h_n$, where

$$
\zeta = \max \left\{ 4 \cdot \left( \frac{L_\alpha^2MMK + MMK^2\kappa_x^2/2}{\kappa_\ell M_\ell} \right)^{1/2}, 16\kappa_x(M + MMK^2\kappa_x^2)^{1/2} \cdot L_\alpha^2MMK_\ell \right\}.
$$

Then we have

$$
\|\hat{\beta}_{h_n} - \beta^*\|_2^2 \leq \frac{288s\lambda_n^2}{M_\ell^2\kappa_x^4},
$$

with probability at least $1 - 17.81 \exp(-c\log p) - 2 \exp(-c'n)$. 

2. Assume that Assumption 5 holds with $\alpha \in (0, 1]$. Suppose that (A4.18) holds with $q = 2305s$, and we take $B = M_q$, $D = M_d$, $E = M_a$ and $a = \alpha$. Further assume that
where \(\gamma = \alpha\) if \(M_dM_a = 0\), and \(\gamma = \min\{\alpha, 1/2\}\) if otherwise. Then we have

\[
\|\hat{\beta}_n - \beta^*\|_2^2 \leq \frac{288s\lambda_n^2}{M_f^2\kappa_f^2},
\]

with probability at least \(1 - 17.81 \exp(-c\log p) - 2\exp(-c'n)\).

(3) Assume that Assumption 5 holds with \(\alpha \in [1/4, 1]\). Suppose that (A4.18) holds with \(q = 2305\{s + \zeta^2h_n^2\gamma / \log(np)\}\), and take \(B = M_g, D = M_d, E = M_a\) and \(a = \alpha\). Denote \(C\) to be some positive absolute constant \(C > \zeta^2C_0^2\gamma\), and suppose \(n \geq (C - \zeta^2C_0^2\gamma)s\log(np)\), where

\[
\zeta = \max \left\{4 \cdot \left( \frac{M_g^2M_MK_0^{2\alpha-2\gamma} + M_d^2M_aC_0^{1-2\gamma} + M_MK\tilde{u}^2C_0^{2-2\gamma}/2}{\kappa_fM_f} \right)^{1/2}, \right.
\]

\[
\left. \frac{16\kappa_x(M + M_MK_0^{2\alpha-2\gamma})^{1/2}}{\kappa_fM_f} \cdot \left( \frac{M_g^2M_MK_0^{2\alpha-2\gamma} + M_d^2M_aC_0^{1-2\gamma}}{\kappa_fM_f} \right)^{1/2} \right\},
\]

where \(\gamma = \alpha\) if \(M_dM_a = 0\), and \(\gamma = \min\{\alpha, 1/2\}\) if otherwise. Further assume \(\lambda_n \geq 4(A' + A'' + M_{\eta})\{\log(np)/n\}^{1/2} + 8MMK^{-1/2}\kappa_x^2h_n\). Then we have

\[
\|\hat{\beta}_n - \beta^*\|_2^2 \leq \frac{288s\lambda_n^2}{M_f^2\kappa_f^2} + \frac{2s\log(np)}{n} + \frac{288n\lambda_n^2}{M_f^2\kappa_f^2\log(np)} + 2\cdot \zeta^2h_n^2\gamma,
\]

with probability at least \(1 - 24.58 \exp(-c\log p) - 2\exp(-c'n)\).

**Proof.** We prove the theorem for the case when \(g\) is Lipschitz. We verify Assumptions 14 and 15, and then apply Theorem 3.1. Assumption 14 is verified by applying Theorem 3.4, and Assumption 15 is verified by applying Lemma A4.22. We complete the proof by Theorem 3.1.

The rest of the theorem can be proved based on similar arguments.

### A4.8 Proof of Theorem 2.2

**Theorem A4.8** (Theorem 2.2). Assume \(h_n \geq K_1\{\log(np)/n\}^{1/2}\) for positive absolute constant \(K_1\), and assume that \(h_n \leq C_0\) for positive constant \(C_0\). Further assume \(\lambda_n \geq 4(A + A')\cdot\{\log(np)/n\}^{1/2} + 4\sqrt{2}M_gM_MK\kappa_x(1 + C_0)h_n\), where

\[
A = \left\{16\sqrt{3}M_f(1 + c)^{1/2} + 4\sqrt{3}C_1M_f^{1/2}K^{-1/2}(1 + c)^{1/2} + 8C_2(1 + c) + 8C_3M_K^{1/2}M_f^{1/2}K^{-1/2}(1 + c)^{3/2} + 8C_4M_KK^{-1}(1 + c)^2 + 8M_f(c + 2)\right\}\kappa_x\kappa_u,
\]

\[
A' = 8MM_KM_gC_0(1 + C_0)\kappa_x(1 + c)^{1/2} + 2C_1M_gM_K^{1/2}M_K^{3/2}\kappa_x(1 + C_0)^{1/2}C_0^{5/4}K^{-1/4}(1 + c)^{1/2} + 2\sqrt{2}C_2M_MK_g^2(1 + C_0)\kappa_xK(1 + c)^{3/2} + 4C_3M_K^{3/2}M_g^{1/2}(1 + C_0)^{1/2}C_0^{1/2}\kappa_x(1 + c)^2 + 2\sqrt{2}C_4M_KM_gC_0\kappa_xK(1 + c)^{5/2} + 2\sqrt{2}MM_KM_g(1 + C_0)C_0,
\]
for positive absolute constant $c$, $M_f = M + MMKc_0$, and $C_1, \ldots, C_4$ as defined in (A3.2). Suppose we have

$$n > \max \left\{ 64 (c + 2)^2 (c + 1) \{ \log (np) \}^3 / 3, 64 (c + 2)^3 (c + 1) \{ \log (np) \}^4, \{ \log (np) \}^{5 / 3}, 3, \right.$$  

$$\frac{48 \sqrt{6} M K \kappa^2 q}{K_1^2 \rho \{ \log (np) \}^{1 / 2}}, \left( \frac{2^{10} \cdot 6 \cdot \sqrt{2} M_f \kappa^2 q}{K_1^2 \kappa \mu \rho \{ \log (np) \}^{2 / 3}}, \frac{144 \kappa^4}{M^2 \kappa^2} \right)^{2 / 3}, \frac{211 \cdot 6 \cdot \sqrt{3} (2 + c)^{1 / 2} C_4 M_f^{1 / 2} M_2^{1 / 2} \kappa^2}{K_1^{1 / 2} \kappa \mu \rho \{ \log (np) \}^{4 / 3}}, q^{1 / 3} \{ \log (np) \}^{1 / 3},$$  

$$\left[ \frac{2^8 \cdot 6 \cdot (20 + 7.5 c) (2 + c) C_2 M_f \kappa^2}{K_1^2 \kappa \mu \rho \{ \log (np) \}^{1 / 2}}, q^{1 / 2} \log (np), \frac{2^8 \cdot 6 (c + 2)^2 \kappa \mu \rho \{ \log (np) \}^{1 / 2}}{K_1^{1 / 2} \kappa \mu \rho \{ \log (np) \}^{5 / 3}}, \frac{2^{11} \cdot 6 \cdot (20 + 7.5 c) (c + 2) M_f \kappa^2}{K_1 \kappa \mu \rho \{ \log (np) \}^2}, \frac{2^9 (3 M_f^2 \kappa^2 + 2 M_f^2 M_2^2 C_4^2 C_2^2) \kappa^2}{(K_1 \kappa \mu \rho \{ \log (np) \}^2 \land (16 \kappa \mu \rho \{ \log (np) \})^2} \cdot q \log \left( \frac{6 \exp \left( \frac{6 \kappa^2}{q} \right)}{q} \right), \right.$$  

$$\frac{2^{24} K_2^2 M_f^2 M_2^2 C_2^2 \kappa^2}{(K_1 \kappa \mu \rho \{ \log (np) \}^2 \land (16 \kappa \mu \rho \{ \log (np) \})^2} \right\},$$

(A4.19)

where $q = 2305 s$. Then under Assumptions 6-12, and 4, we have

$$\| \beta_{h_n} - \beta^* \|^2 \leq \frac{288 s \lambda^2}{M_f^2 \kappa^2},$$

with probability at least $1 - 17.81 \exp(-c \log p) - 2 \exp(-c' n)$, where

$$c' = (K_2^2 M_f^2 \kappa \mu \rho \{ \log (np) \}^2 / 2^{16} \{ (3 M_f^2 \kappa^2 + 2 M_f^2 M_2^2 C_4^2 C_2^2) \kappa^2 \} \land (2 M_f^2 M_2^2 C_4^2 C_2^2) \kappa^2 \right\},$$

Proof of Theorem 2.2. We adopt the framework as described in Section 2.1 for $\theta^* = \beta^*$, $\Gamma_0(\theta) = L_0(\beta)$, $\Gamma_n(\theta, h) = \tilde{L}_n(\beta, h)$, $\Gamma_n(\theta) = \mathbb{E} \tilde{L}_n(\beta, h)$, and take $\theta^*_h = \beta^*$, which yields $s_n \leq s$ and $\rho_n = 0$. We verify Assumption 2 by applying Lemma A4.20, and verify Assumption 3 by applying Corollary A3.1. We complete the proof by Theorem 2.1. \qed
A4.9 Proof of Theorem A3.1

Theorem A4.9 (Theorem A3.1). For $q \in [p]$, suppose that

$$
n > \max \left\{ \frac{48 \sqrt{6} M_K \kappa_x^2 q}{K_1 p \{\log (np)\}^{1/2}}, \left( \frac{384 \sqrt{6} M_f \kappa_x^2 q}{t K_1^2 p^2 \log (np)} \right)^{2/3}, \frac{144 \kappa_{x_2}^4}{K_1^2 p^2 \log (np)} \right\},
$$

$$
\left[ \frac{768 (2 + c)^{1/2} C_1 M_K^{1/2} M_f^{1/2} \kappa_x^2}{K_1^{1/2} t} \right]^{4/3} \cdot q^{4/3} \left\{ \log (np) \right\}^{1/3},
$$

$$
\left( \frac{96 (20 + 7.5 c) (2 + c) C_2 M_f \kappa_x^2}{t} \right)^{1/2} \cdot q^{1/2} \log (np),
$$

$$
\left( \frac{384 (2 + c)^3 C_4 \kappa_x^2}{K_1 t} \right)^{2/3} \cdot q^{2/3} \left\{ \log (np) \right\}^{5/3},
$$

$$
\frac{768 (20 + 7.5 c) (c + 2) M_f \kappa_x^2}{t} q \left\{ \log (np) \right\}^2, \frac{12 q}{(20 + 7.5 c) M_f \kappa_x^2 t \log (np)},
$$

$$
\frac{2^{12} \{(3 M^2 \kappa_x^2 + 2 M^2 M_f^2 C_0^2 \kappa_x^2) \lor 2 M^2 \kappa_x^2 \}}{t^2 (16 t)} q \log \left( \frac{6 e p}{q} \right),
$$

$$
\frac{2^{16} K_1^2 M^2 M_f^2 \kappa_x^2 \log (np)}{t^2},
$$

(A4.20)

for positive absolute constant $t$ and $c > 1$. Under Assumptions 7, 8, and 11, we have

$$
\|T_n - E T_n\|_{2,q} \leq t
$$

with probability at least $1 - 5.77 \exp(-c \log p) - 2 \exp(-c' n)$, where $c' = (t^2 \land 4 t) / \left\{ 2^8 \{(3 M^2 \kappa_x^2 + 2 M^2 M_f^2 C_0^2 \kappa_x^2) \lor 2 M^2 \kappa_x^2 \} \right\}$.

Proof. We denote

$$
X_{h_n} = \left( \frac{1}{n} K^{1/2} \left( \frac{W_{ij} h_n}{h_n} \right) \tilde{X}_{ij}^T \right)_{n \times p} \text{ to be a } \binom{n}{2} \times p \text{ matrix},
$$

$$
\Sigma_{h_n} = E \left[ \frac{1}{h_n} K \left( \frac{W_i h_n}{h_n} \right) \tilde{X} \tilde{X}^T \right].
$$

And we aim to show that with high probability

$$
\left| \binom{n}{2}^{-1} v^T X_{h_n} X_{h_n} v - v^T \Sigma_{h_n} v \right| \leq \theta' \|v\|^2_2 \text{ for all } v \in \mathbb{R}^p, \|v\|_0 \leq q \text{ simultaneously}
$$

holds for some $\theta' > 0$ under conditions of Theorem A3.1. We split the proof into three steps.

**Step I.** For set $J \subseteq [p]$, consider $E_J \cap S_j^{p-1}$, where $E_J = \text{span} \{ e_j : j \in J \}$. Construct $\epsilon$-net $\Pi_J$, such that $\Pi_J \subset E_J \cap S_j^{p-1}$ and $|\Pi_J| \leq (1 + 2 \epsilon^{-1})^q$. The existence of $\Pi_J$ can be guaranteed by Lemma 23 of Rudelson and Zhou (2013). Define $\Pi = \bigcup_{|J| = q} \Pi_J$, then for $0 < \epsilon < 1$ to be determined later, we have

$$
|\Pi| \leq \left( \frac{3}{\epsilon} \right)^q \left( \frac{p}{q} \right)^p \leq \left( \frac{3 e p}{q \epsilon^3} \right)^q = \exp \left\{ q \log \left( \frac{6 e p}{q} \right) \right\}.
$$

21
For any \( v \in E_J \cap S_2^{p-1} \), let \( \Pi(v) \) be the closest point in \( \epsilon \)-net \( \Pi_J \). Then we have
\[
\frac{v - \Pi(v)}{\|v - \Pi(v)\|_2} \in E_J \cap S_2^{p-1}, \quad \|v - \Pi(v)\|_2 \leq \epsilon.
\]

**Step II.** Denote \( D_i = (W_i, X_i, V_i) \) for \( i \in [n] \), and \( D = (W, X, V) \) to be an i.i.d copy. We upper bound
\[
\mathbb{P}\left( \max_{v \in \Pi} \left\{ \left( \frac{n}{2} \right)^{-1} \sum_{i<j} g_v(D_i, D_j) - \mu_v \right\} \geq \theta \right),
\]
for some \( \theta > 0 \), where
\[
g_v(D_i, D_j) = \frac{1}{h_n} K\left( \frac{W_{ij}}{h_n} \right) \left( \tilde{X}_{ij}^T v \right)^2,
\]
and \( \mu_v = \mathbb{E}[g_v(D_i, D_j)]. \)

Also, denote \( f_v(D_i) = \mathbb{E}[g_v(D_i, D_j) | D_i] \). Observe that
\[
\left| \left( \frac{n}{2} \right)^{-1} \sum_{i<j} g_v(D_i, D_j) - \mu_v \right| 
\leq \left| \left( \frac{n}{2} \right)^{-1} \sum_{i<j} \{ g_v(D_i, D_j) - f_v(D_i) - f_v(D_j) + \mu_v \} \right| + \left| \frac{2}{n} \sum_{i=1}^n \{ f_v(D_i) - \mu_v \} \right|.
\]

We bound two components on the right hand side of inequality above separately, and then combine the result.

**Step II.1.** We bound
\[
\mathbb{P}\left( \frac{1}{n} \sum_{i=1}^n \{ f_v(D_i) - \mu_v \} \geq t \right), \tag{A4.21}
\]
for \( t > 0 \) to be determined, and for each \( v \in E_J \cap S_2^{p-1} \). Apply Lemma A3.3 with conditions of lemma satisfied by Assumptions 7 (Lemma A4.15) and 8 (Lemma A4.16), and we have
\[
|f_v(D_i) - f_1(D_i)| \leq |MMK h_n f_2(D_i)|, \tag{A4.22}
\]
where \( f_1(D_i) = \mathbb{E}\left[ (\tilde{X}_{ij}^T v)^2 | \tilde{W}_{ij} = 0, D_i \right] f_W(W_i), \) and \( f_2(D_i) = \mathbb{E}\left[ (\tilde{X}_{ij}^T v)^2 | X_i \right]. \) Also, we have
\[
|\mu_v - \mu_1| \leq |MMK h_n \mu_2|, \tag{A4.23}
\]
where \( \mu_1 = \mathbb{E}\left[ (\tilde{X}_{ij}^T v)^2 | \tilde{W}_{ij} = 0 \right] f_W(0), \) and \( \mu_2 = \mathbb{E}[f_2(D_i)] = \mathbb{E}[(\tilde{X}_{ij}^T v)^2]. \) And we bound (A4.21) as
bounded due to Jensen’s inequality, where (Lemma A4.17, and Lemma A4.18). The fourth inequality is by noting that
\[E[e^{a\left(\sum_{i=1}^{n} f_i(D_i) - \mu_v\right)}] \leq e^{-nat} \cdot E[e^{a\left(\sum_{i=1}^{n} f_i(D_i) - \mu_v\right)}] \leq e^{-nat} \cdot E[e^{a\left(\sum_{i=1}^{n} (f_1(D_i) - \mu_1) + MM_K h_n (f_2(D_i) - \mu_2))\right)}] \cdot e^{2MM_K h_n} \cdot e^{\mu_2 a} \leq e^{-nat} \cdot E[e^{2a\sum_{i=1}^{n} (f_1(D_i) - \mu_1)}]^{1/2} \cdot E[e^{2MM_K C_\theta \sum_{i=1}^{n} (f_2(D_i) - \mu_2)}]^{1/2} \cdot e^{4\kappa_2^2 MM_K h_n} \cdot e^{\mu_1} \leq e^{-nat} \cdot E[e^{2aM \sum_{i=1}^{n} \left|\mathbb{E}((\tilde{X}_{ij}^T v)^2|\tilde{W}_{ij}=0, D_i) - \mathbb{E}((\tilde{X}_{ij}^T v)^2|\tilde{W}_{ij}=0)\right|}^2]^{1/2} \cdot e^{2\kappa_4^2 MM_K h_n} \cdot e^{\mu_1} \leq e^{-nat} \cdot e^{2\kappa_4^2 a_2^2 n} \cdot e^{2\kappa_4^2 a_2^2 n} \cdot e^{4\kappa_4^2 a_2^2 n} \cdot e^{4MM_K \kappa_4^2 h_n},
for 0 < a \leq (4MM_K^2)^{-1}, where the first inequality is by Markov’s, the second is an application of (A4.22) and (A4.23), the third is by Cauchy-Schwarz and the result that \(\mu_2 \leq 2\kappa_2^2\) (Assumption 11, Lemma A4.17, and Lemma A4.18). The fourth inequality is by noting that \(f_iW(0) = \mathbb{E}[f_iW(W_i)]\), and applying the following inequality
\[|V_1V_2 - \mathbb{E}[V_1]\mathbb{E}[V_2]| \leq |V_1 - \mathbb{E}[V_1]| \cdot |V_2| + |\mathbb{E}[V_1]| \cdot |V_2 - \mathbb{E}[V_2]|,
where \(V_1 = \mathbb{E}((\tilde{X}_{ij}^T v)^2|\tilde{W}_{ij}=0, D_i)\), \(|\mathbb{E}[V_1]| \leq 2\kappa_2^2\) by Assumption 11, Lemma A4.17, and Lemma A4.18, and \(V_2 = f_iW(W_i) \in [0, M]\). For the fifth inequality, the second component in product is bounded due to Jensen’s inequality, where \((X_i^T, W_i)\), \(i = 1, \ldots, n\) are independent copies of \((X_i, W_i)\); the third is bounded because \(f_iW(W_i) \in [0, M]\) and \(\mathbb{E}((\tilde{X}_{ij}^T v)^2|\tilde{W}_{ij}=0) \leq 2\kappa_2^2\) by Assumption 11, Lemma A4.17, and Lemma A4.18. The sixth inequality is again an application of Assumption 11, Lemma A4.17, and Lemma A4.18.

Take \(a = (1 \land t) \cdot (2a_1)^{-1}\), and \(h_n \leq t \cdot (4a_2)^{-1}\), where \(a_1 = (2M^2\kappa_4^2 + 2M^2M_K^2C_\theta^2\kappa_4^4 + M^2\kappa_4^4) \land 2M\kappa_2^2\) and \(a_2 = 4MM_K\kappa_4^2\). Then we further have
\[P\left(\frac{1}{n} \sum_{i=1}^{n} \{f_i(D_i) - \mu_v\} \geq t\right) \leq \exp\left\{-\frac{n(t^2 \land t)}{8a_1}\right\}.
By the same argument, we have
\[P\left(\frac{1}{n} \sum_{i=1}^{n} \{f_i(D_i) - \mu_v\} \leq -t\right) \leq \exp\left\{-\frac{n(t^2 \land t)}{8a_1}\right\}.
We take \(t = \theta/4\), and have
\[P\left(\frac{1}{n} \sum_{i=1}^{n} \{f_i(D_i) - \mu_v\} \geq \frac{\theta}{4}\right) \leq 2 \exp\left\{-\frac{n(\theta^2 \land 4\theta)}{128a_1}\right\}.
\]
Step II.2. Observe that
\[
\left| \left( \frac{n}{2} \right)^{-1} \sum_{i<j} \{ g_v(D_i, D_j) - f_v(D_i) - f_v(D_j) + \mu_v \} \right| \leq \left( \frac{n}{2} \right)^{-1} \max_{k,l} \left\{ \left| \sum_{i<j} \tilde{\varphi}_{kl}(D_i, D_j) \right| \right\},
\]
where
\[
\tilde{\varphi}_{kl}(D_i, D_j) = \frac{1}{h_n} K(\frac{\tilde{W}_{ij}}{h_n}) \tilde{X}_{ijk} \tilde{X}_{ijl} - \mathbb{E} \left[ \frac{1}{h_n} K(\frac{\tilde{W}_{ij}}{h_n}) \tilde{X}_{ijk} \tilde{X}_{ijl} \middle| D_i \right] - \mathbb{E} \left[ \frac{1}{h_n} K(\frac{\tilde{W}_{ij}}{h_n}) \tilde{X}_{ijl} \tilde{X}_{ijl} \middle| D_j \right] + \mathbb{E} \left[ \frac{1}{h_n} K(\frac{\tilde{W}_{ij}}{h_n}) \tilde{X}_{ijl} \tilde{X}_{ijl} \middle| D_i \right] + \mathbb{E} \left[ \frac{1}{h_n} K(\frac{\tilde{W}_{ij}}{h_n}) \tilde{X}_{ijl} \tilde{X}_{ijl} \middle| D_j \right].
\]

We then bound \( \sum_{i<j} \tilde{\varphi}_{kl}(D_i, D_j) \) for each \( k, l \in [p] \).

Apply truncation \( |X_{ik} - \mathbb{E}[X_{ik}]| \leq \tau_n/2 \) for each \( i \in [n], k \in [p] \), and \( \tau_n = \sqrt{6}(2+c)^{1/2} \kappa_s \{ \log(np) \}^{1/2} \), for positive absolute constant \( c \). Define events
\[
A_i = \{ |X_{ik} - \mathbb{E}[X_{ik}]| \leq \frac{\tau_n}{2}, k \in [p] \}, \quad A_{i[n]} = \{ |X_{ik} - \mathbb{E}[X_{ik}]| \leq \frac{\tau_n}{2}, i \in [n], k \in [p] \}.
\]
Consider truncated U-statistic \( \sum_{i<j} \tilde{\varphi}_{kl}(D_i, D_j) \), where
\[
\varphi_{kl}(D_i, D_j) = \frac{1}{h_n} K \left( \frac{\tilde{W}_{ij}}{h_n} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} \mathbb{I} \left( A_i \cap A_j \right) - \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{\tilde{W}_{ij}}{h_n} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} \middle| D_i \right] \mathbb{I} \left( A_i \right) - \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{\tilde{W}_{ij}}{h_n} \right) \tilde{X}_{ijl} \tilde{X}_{ijl} \middle| D_j \right] \mathbb{I} \left( A_j \right) + \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{\tilde{W}_{ij}}{h_n} \right) \tilde{X}_{ijl} \tilde{X}_{ijl} \middle| D_i \right] \mathbb{I} \left( A_i \right) + \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{\tilde{W}_{ij}}{h_n} \right) \tilde{X}_{ijl} \tilde{X}_{ijl} \middle| D_j \right] \mathbb{I} \left( A_j \right).
\]

First, we bound \( \mathbb{E}[\varphi_{kl}(D_i, D_j)] \). We have
\[
\mathbb{E}[\varphi_{kl}(D_i, D_j)]
\leq \left[ \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{\tilde{W}_{ij}}{h_n} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} \mathbb{I} \left( A_i \cap A_j \right) \right] - 2 \mathbb{E} \left[ \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{\tilde{W}_{ij}}{h_n} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} \middle| D_i \right] \mathbb{I} \left( A_i \right) \right] \right] (A.25)
\leq \left[ \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{\tilde{W}_{ij}}{h_n} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} \mathbb{I} \left( A_i \cap A_j \right) \right] + 2 \mathbb{E} \left[ \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{\tilde{W}_{ij}}{h_n} \right) \tilde{X}_{ijl} \tilde{X}_{ijl} \middle| D_i \right] \mathbb{I} \left( A_i \right) \right] \right] (A.26)
\]
\[
\leq MK \frac{1}{h_n} \mathbb{E}[\tilde{X}_{ijk}^2 \tilde{X}_{ijl}^2]^{1/2} \mathbb{P}(A_i \cap A_j)^{1/2}
\leq MK \frac{1}{h_n} \mathbb{E}[\tilde{X}_{ijk}^4 \tilde{X}_{ijl}^4]^{1/4} \mathbb{P}(A_i \cap A_j)^{1/2}
\leq MK \frac{1}{h_n} (12\kappa_s^4)^{1/2} \cdot (2p^{-1}n^p)^{1/2}
\leq \frac{2\sqrt{6}MK^2}{K_1 np \{ \log(np) \}^{1/2} p^{1/2}} \leq \frac{\theta}{24q},
\]
where the first and second inequalities are by Cauchy-Schwarz, the third is by subgaussianity of \( X_i, X_j \), the fourth is by choice of \( h_n \), and the last holds true when we have
\[
n \geq 48\sqrt{6}MK^2 q \frac{1}{K_1 \theta \{ \log(np) \}^{1/2} p}.\]
We also have
\[
\left| \mathbb{E}\left( \frac{1}{h_n} K\left( \frac{\tilde{W}_{ij}}{h_n} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} | D_i \right) \right| \leq \frac{\mathbb{E}\left( \frac{1}{h_n} K\left( \frac{\tilde{W}_{ij}}{h_n} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} | D_i \right)^2}{\mathbb{E}\left( \frac{1}{h_n} K\left( \frac{\tilde{W}_{ij}}{h_n} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} | D_i \right)} \cdot \mathbb{P}(A_c)^{1/2} \\
\leq \left\{ 24(M + M_K C_0)^2 \kappa_n^4 \right\}^{1/2} \cdot \frac{1}{n^{3/2} p} \\
\leq \frac{\theta}{48 q},
\]
where the first inequality is by Cauchy-Schwarz, the second is by \((A4.51)\) and subgaussianity of \(X_i\) (Assumption 11), and the last holds true when we have
\[
n \geq \left\{ \frac{96\sqrt{6}(M + M_M C_0)\kappa_n^2 + 2 + c}{\theta p} \right\}^{2/3}.
\]
Combining \((A4.25)\), \((A4.26)\), and \((A4.27)\), we have
\[
\left| \mathbb{E}[\varphi_{kl}(D_i, D_j)] \right| \leq \frac{\theta}{12 q},
\]
when we appropriately choose \(n\) bounded from below.

Next, we bound \(| \sum_{i<j} \varphi_{kl}(D_i, D_j) |\) by applying Lemma A3.4. We bound constants in Lemma A3.4 as follows.

For bounding \(B_g\), we have \(B_g \leq 4M_K \tau_n^2 \cdot h_n^{-1} \leq \{4\sqrt{6}(2 + c)M_K \kappa_n^2 \cdot K_1^{-1}\} \cdot \{n \log(np)\}^{1/2}.\) For bounding \(B_f\), we have
\[
\mathbb{E}\left[ |\varphi_{kl}(D_i, D_j)| | D_j \right] \\
\leq \mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{\tilde{W}_{ij}}{h_n} \right) | \tilde{X}_{ijk} \tilde{X}_{ijl} | | \Pi(A_i \cap A_j) | D_j \right] + \mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{\tilde{W}_{ij}}{h_n} \right) | \tilde{X}_{ijk} \tilde{X}_{ijl} | | \Pi(A_i) \right] \\
+ \mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{\tilde{W}_{ij}}{h_n} \right) | \tilde{X}_{ijk} \tilde{X}_{ijl} | | \Pi(A_i \cap A_j) | D_j \right] + \mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{\tilde{W}_{ij}}{h_n} \right) | \tilde{X}_{ijk} \tilde{X}_{ijl} | | \Pi(A_j) \right] \\
+ 2 \cdot \mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{\tilde{W}_{ij}}{h_n} \right) | \tilde{X}_{ijk} \tilde{X}_{ijl} | \right],
\]
Apply Lemma A3.3 on \(\varphi = 1\), with \(M_1 = M\) and \(M_2 = M_K\) as given by Assumptions 8 (Lemma A4.16) and 7 (Lemma A4.15), we have
\[
\mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{\tilde{W}_{ij}}{h_n} \right) | \tilde{X}_{ijk} \tilde{X}_{ijl} | | \Pi(A_i \cap A_j) | D_i \right] \\
\leq \tau_n^2 (M + M_M C_0) = 6(c + 2)(M + M_M C_0) \kappa_n^2 \log(np).
\]
Apply Lemma A3.3 on \(\varphi = | \tilde{X}_{ijk} \tilde{X}_{ijl} |\), with \(M_1 = M\) and \(M_2 = M_K\) as given by Assumptions
Combining (A4.29)-(A4.32), we have

\[ E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) | \tilde{X}_{ijk}, \tilde{X}_{ijl} | D_j \right] \text{I}(A_j) \]

\[ \leq M \cdot E \left[ | \tilde{X}_{ijk} \tilde{X}_{ijl} | D_j, \tilde{W}_{ij} = 0 \right] \text{I}(A_j) + MM_K C_0 E \left[ | \tilde{X}_{ijk} \tilde{X}_{ijl} | D_j \right] \text{I}(A_j) \]

\[ \leq M E \left[ \tilde{X}_{ijk}^2 | D_j, \tilde{W}_{ij} = 0 \right]^{1/2} E \left[ \tilde{X}_{ijl}^2 | D_j, \tilde{W}_{ij} = 0 \right]^{1/2} \text{I}(A_j) \]

\[ + MM_K C_0 E \left[ \tilde{X}_{ijk}^2 | D_j \right]^{1/2} E \left[ \tilde{X}_{ijl}^2 | D_j \right]^{1/2} \text{I}(A_j) \]

\[ \leq (1.5c + 4) \cdot (M + MM_K C_0) \cdot \kappa_x^2 \log(np), \]

where the second inequality is by Cauchy Schwarz, and the last is due to

\[ E[\tilde{X}_{ijk}^2 | D_j, \tilde{W}_{ij} = 0] \text{I}(A_j) = \left\{ E[(X_{ik} - E[X_{ik}])^2] + (X_{ik} - E[X_{jk}])^2 \right\} \text{I}(A_j) \]

\[ \leq \kappa_x^2 / 4 \leq (1.5c + 4) \kappa_x^2 \log(np), \]

and based on an identical argument

\[ E[\tilde{X}_{ijl}^2 | D_j, \tilde{W}_{ij} = 0] \text{I}(A_j) \leq (1.5c + 4) \kappa_x^2 \log(np), \]

for any \( k \in [p] \).

Apply Lemma A3.2 on \( Z = | \tilde{X}_{ijk} \tilde{X}_{ijl} | \), and with \( M_1 = M, M_2 = M_K \) as given by Assumptions 8 (Lemma A4.16) and 7 (Lemma A4.15), we have

\[ E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) | \tilde{X}_{ijk}, \tilde{X}_{ijl} | \right] \leq 2(M + MM_K C_0) \kappa_x^2 \]

(C4.32)

Combining (A4.29)-(A4.32), we have \( B_{\ell} \leq (20 + 7.5c) \cdot (M + MM_K C_0) \cdot \kappa_x^2 \cdot \log(np) \).

For bounding \( E \left[ \varphi_{kl}(D_i, D_j) | D_j \right] \), we observe that

\[ \varphi_{kl}(D_i, D_j) = \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} \text{I}(A_i \cap A_j) - E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} \text{I}(A_i \cap A_j) | D_i \right] \]

\[ - E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} \text{I}(A_i \cap A_j) | D_j \right] + E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} \text{I}(A_i \cap A_j) | D_i \right] \]

\[ + E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} \text{I}(A_i \cap A_j) | D_j \right] - E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} | D_i \right] \text{I}(A_i) \]

\[ + E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} | D_j \right] - E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} | D_j \right] \text{I}(A_j) \]

\[ + E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} \right] - E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} \text{I}(A_i \cap A_j) \right], \]

which further implies that

\[ | E \left[ \varphi_{kl}(D_i, D_j) | D_j \right] | \]

\[ \leq \left| E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} | A_i \right] \right| + \left| E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} | A_i^c \right] \right| | D_j | \]

\[ + \left| E \left[ \frac{1}{h_n} K \left( \tilde{W}_{ij} \right) \tilde{X}_{ijk} \tilde{X}_{ijl} | A_i \cup A_j \right] \right| \]

26
Therefore we have
\[
\mathbb{E}\left[\mathbb{E}\left\{\varphi_{kl}(D_i, D_j)\right| D_j\right\}^2\right]
\leq \frac{3}{h_n^2}\left\{\mathbb{E}\left[\tilde{X}_{ijk}\tilde{X}_{ijl}\mathbb{I}(A_i^c)\right]^2 + \mathbb{E}\left[\tilde{X}_{ijk}\tilde{X}_{ijl}\mathbb{I}(A_j^c)\right]^2 + \mathbb{E}\left[\tilde{X}_{ijk}\tilde{X}_{ijl}\mathbb{I}(A_i^c \cup A_j^c)\right]^2\right\}
\leq \frac{3n}{K_f^2 \log(np)}\left\{2\mathbb{E}\left[\tilde{X}_{ijk}^4\right]^{1/2}\mathbb{E}\left[\tilde{X}_{ijl}^4\right]^{1/2}\mathbb{P}(A_i^c) + \mathbb{E}\left[\tilde{X}_{ijk}^4\right]^{1/2}\mathbb{P}(A_j^c)\right\}
\leq \frac{3n}{K_f^2 \log(np)}\left\{2 \cdot 12\kappa^4 h^4 \frac{1}{n^3 p^2} + 12\kappa_x^4 \frac{2}{n^3 p^2}\right\} \leq \frac{1}{n},
\]
where the first inequality is due to the fact that $K(\cdot) \in [0, 1]$ and by Jensen’s inequality, the second is by Cauchy-Schwarz, the third by subgaussianity of $X_i$, $X_j$ and $\tilde{X}_{ij}$, and last holds true when we have
\[
n \geq \frac{144\kappa_x^4}{K_f^2 p^2 \log(np)}.
\]

For bounding $\sigma^2$, apply Lemma A3.2 on $Z = \tilde{X}_{ijk}^2 \tilde{X}_{ijl}^2$ with $M_1 = M$ and $M_2 = M_K$ as given by Assumptions 8 (Lemma A4.16) and 7 (Lemma A4.15), we have
\[
\sigma^2 \leq \frac{16 M K}{h_n^2} \mathbb{E}\left[\mathbb{I}\left(\frac{\tilde{W}_{ij}}{h_n}\right)\tilde{X}_{ijk}^2 \tilde{X}_{ijl}^2\right]
\leq \frac{16 M K}{h_n^2} \left\{M \cdot \mathbb{E}\left[\tilde{X}_{ijk}^2 \tilde{X}_{ijl}^2 | \tilde{W}_{ij} = 0\right] + M M_K C_0 \mathbb{E}\left[\tilde{X}_{ijk}^2 \tilde{X}_{ijl}^2\right]\right\}
\leq \frac{16 M K}{h_n^2} \left\{M \mathbb{E}\left[\tilde{X}_{ijk}^4 | \tilde{W}_{ij} = 0\right]^{1/2} \mathbb{E}\left[\tilde{X}_{ijl}^4 | \tilde{W}_{ij} = 0\right]^{1/2} + M M_K C_0 \mathbb{E}\left[\tilde{X}_{ijk}^4\right]^{1/2} \mathbb{E}\left[\tilde{X}_{ijl}^4\right]^{1/2}\right\}
\leq \frac{192 M K (M + M M_K C_0) \kappa_x^4}{K_1} \left\{\frac{n}{\log(np)}\right\}^{1/2},
\]
where the third inequality is by Cauchy-Schwarz, and the last is by subgaussianity of $\tilde{X}$ and choice of $h_n$.

For bounding $B^2$, we have
\[
B^2 = n \sup_{D_j} \mathbb{E}\left[\varphi_{kl}(D_i, D_j) | D_j\right]
\leq 4 M_K n h_n^{-1} \sup_{D_j} \mathbb{E}\left[\mathbb{I}\left(\frac{\tilde{W}_{ij}}{h_n}\right) \tilde{X}_{ijk} \tilde{X}_{ijl} | A_i \cap A_j\right] \mathbb{I}(D_j)
+ 4 n \sup_{D_j} \mathbb{E}\left[\mathbb{I}\left(\frac{\tilde{W}_{ij}}{h_n}\right) \tilde{X}_{ijk} \tilde{X}_{ijl} | D_j\right]^2 \mathbb{I}(A_i)
+ 4 n \sup_{D_j} \mathbb{E}\left[\mathbb{I}\left(\frac{\tilde{W}_{ij}}{h_n}\right) \tilde{X}_{ijk} \tilde{X}_{ijl} | D_j\right]^2 \mathbb{I}(A_j)
+ 4 n \mathbb{E}\left[\mathbb{I}\left(\frac{\tilde{W}_{ij}}{h_n}\right) \tilde{X}_{ijk} \tilde{X}_{ijl}\right]^2
\leq \frac{4 M_K n h_n^{-1}}{h_n} + 192 M_f^2 \kappa_x^4 n + 8 M_f \kappa_x^2 n
\leq \{144(2 + c)^2 M_K M_f \kappa_x^4 K_1^{-1} + 192 M_f^2 \kappa_x^4 + 8 M_f \kappa_x^2\} \cdot \{n \log(np)\}^{3/2},
\]
where $M_f = M + MM_KC_0$.

We take
\[
t = \left(\frac{n}{2}\right) \frac{\theta}{12q},
\]
\[
u = (2 + c) \log p,
\]
and require that
\[
n > \max \left\{ \frac{48 \sqrt{6} M_K \kappa_x^2 q}{K_1 p \{\log(np)\}^{1/2}} \left(\frac{96 \sqrt{6} M_f \kappa_x^2 q}{\theta p}\right)^{2/3}, \frac{144 \kappa_x^4}{K_1^2 p^2 \log(np)}, \frac{192 \sqrt{3}(2 + c)^{1/2} C_1 M_K^{1/2} M_f^{1/2} \kappa_x^2}{K_1^{1/2} \theta} \cdot q^{4/3} \{\log(np)\}^{1/2}, \frac{24(20 + 7.5c)(2 + c) M_f \kappa_x^2}{\theta} \cdot q^{1/2} \log(np), \frac{24(c + 2)^3 C_3 \{144(2 + c)^2 M_K M_f \kappa_x^4 K_1^{-1} + 192 M_f^2 \kappa_x^4 + 8 M_f \kappa_x^4\}^{1/2}}{\theta} \cdot q^{3/2} \{\log(np)\}^{3/2}, \frac{96 \sqrt{6}(2 + c)^3 C_4 \kappa_x^2}{K_1 \theta} \cdot q^{2/3} \{\log(np)\}^{5/3}, \frac{192(20 + 7.5c)(c + 2) M_f \kappa_x^2}{\theta} q \{\log(np)\}^2, \frac{12q}{(20 + 7.5c) M_f \kappa_x^2 \theta \log(np)} \right\}^{1/2} \right\}
\]
(A4.33)

for some positive absolute constant $c$, and $C_1, \ldots, C_4$ as defined in (A3.2). Then by Lemma A3.4, we have
\[
\mathbb{P}\left(\left| \frac{1}{\sqrt{n}} \sum_{i<j} \varphi_{kl}(D_i, D_j) - \mathbb{E}[\varphi_{kl}(D_i, D_j)] \right| \geq \frac{5\theta}{12q}\right) \leq 2 \exp\{-2 + (2 + c) \log p\} + 2.77 \exp\{-2 + (2 + c) \log p\}
\]

Combined with (A4.25), the last display further implies that
\[
\mathbb{P}\left(\left| \frac{1}{\sqrt{n}} \sum_{i<j} \varphi_{kl}(D_i, D_j) \right| \geq \frac{\theta}{2q}\right) \leq \mathbb{P}\left(\left| \frac{1}{\sqrt{n}} \sum_{i<j} \varphi_{kl}(D_i, D_j) \right| \geq \frac{\theta}{2q} \cap A_{[n]}\right) + \mathbb{P}(A_{[n]}^c)
\]
\[
\leq \mathbb{P}\left(\left| \frac{1}{\sqrt{n}} \sum_{i<j} \varphi_{kl}(D_i, D_j) - \mathbb{E}[\varphi_{kl}(D_i, D_j)] \right| \geq \frac{5\theta}{12m}\right) + \mathbb{P}(A_{[n]}^c)
\]
\[
\leq 2 \exp\{-2 + (2 + c) \log p\} + 2.77 \exp\{-2 + (2 + c) \log p\} + np \exp\{-2 + (2 + c) \log(np)\}
\]
\[
\leq 5.77 \exp\{-1 + (2 + c) \log p\},
\]
for positive absolute constant $c$.  

28
Step II.3 Combining results of Step II.1, Step II.2 and Step I, when we have (A4.33), and that
\[ n > \max \left\{ \frac{256 \{ (3M^2 \kappa_x^2 + 2M^2 M_k^2 C_0^2 \kappa_x^2) \vee 2M \} \kappa_x^2}{\theta^2 \wedge 4\theta}, \frac{4096 \kappa_1^2 M^2 \kappa_x^2}{\theta^2} \log (np) \right\}, \]
we have
\[ P\left( \max_{v \in \Pi} \left\{ \left( \frac{n}{2} \right)^{-1} \sum_{i<j} g_v(D_i, D_j) - \mu_v \right\} \geq \theta \right) \leq 5.77 \exp\{-c + 1\} \log p + 2 \exp(-c' n), \]
where \( c' = (\theta^2 \wedge 4\theta) / [256 \{ (3M^2 \kappa_x^2 + 2M^2 M_k^2 C_0^2 \kappa_x^2) \vee 2M \} \kappa_x^2] \).

Step III. Denote
\[ \Gamma = \left( \frac{n}{2} \right)^{-1/2} X_{h_n} - \Sigma_{h_n}^{1/2}. \]
From Step II.2, we have that, with probability at least 1 - 5.77 \exp\{-c + 1\} \log p - 2 \exp(-c' n), simultaneously for all \( v_0 \in \Pi \),
\[ \| \Gamma v_0 \|_2^2 \leq \theta, \]
which further implies that
\[ \| \Gamma v_0 \|_2 \leq \theta^{1/2}. \]
Then we obtain bounds on entire \( E_J \cap S_{2}^{p-1} \) by approximation.

For any \( v \in E_J \cap S_{2}^{p-1} \) for some \( |J| = q \), denote \( v_0 = \Pi(v) \). We have
\[ \| \Gamma v \|_2 \leq \| \Gamma \Pi(v) \|_2 + \| \Gamma \{ v - \Pi(v) \} \|_2. \quad (A4.34) \]
Define \( \| \Gamma \|_{2,E_J} = \sup_{y \in E_J \cap S_{2}^{p-1}} \| \Gamma y \|_2 \). Then by (A4.34), we have
\[ \| \Gamma \|_{2,E_J} \leq \theta^{1/2} + \epsilon \| \Gamma \|_{2,E_J}, \]
which further implies that
\[ \| \Gamma \|_{2,E_J}^2 \leq \frac{\theta}{(1 - \epsilon)^2}. \]

Take \( \epsilon = 1/2 \), then we have
\[ \| \Gamma \|_{2,E_J}^2 \leq 4 \theta. \]
We take \( \theta' = 4 \theta \). This completes the proof.

A4.10 Proof of Lemma A3.4

Proof. Denote \( \mu = \mathbb{E}[g(Z_1, Z_2)] \), \( \tilde{f}(z) = f(z) - \mu \), \( \tilde{g}(Z_i, Z_j) = g(Z_i, Z_j) - f(Z_i) - f(Z_j) + \mu \), and \( D_n(\tilde{g}) = \sum_{i<j} \tilde{g}(Z_i, Z_j) \). Also, denote \( \| \tilde{g} \|_\infty = \tilde{B}_g \), \( \| \tilde{f} \|_\infty = \tilde{B}_f \), \( \tilde{\sigma}^2 = \mathbb{E}[\tilde{g}(Z_1, Z_2)^2] \), and
\[ \tilde{B}^2 = n \sup_z \mathbb{E}[\tilde{g}(Z, z)^2], \]
\[ \tilde{D} = \sup \left\{ \mathbb{E} \left[ \sum_{i<j} \tilde{g}(Z_i, Z_j) a_i(Z_i) b_j(Z_j) \right] : \mathbb{E} \left[ \sum_{i=2}^{n} a_i(Z_i)^2 \right] \leq 1, \mathbb{E} \left[ \sum_{j=1}^{n-1} b_j(Z_j)^2 \right] \leq 1 \right\}. \]
Hoeffding decomposition gives us
\[ U_n(g) - \mathbb{E}[U_n(g)] = 2(n-1) \sum_{i=1}^{n} \tilde{f}(Z_i) + D_n(\tilde{g}), \]
where \( D_n(\tilde{g}) \) is a degenerate U-statistic of bounded kernel. By Bernstein inequality, we have
\[
\mathbb{P}\left( \left| \sum_{i=1}^{n} \tilde{f}(Z_i) \right| \geq \frac{t}{2(n-1)} \right) \leq 2 \exp \left( \frac{-t^2/8(n-1)^2}{n\mathbb{E}[\tilde{f}(Z_i)^2] + B_f \cdot t/6(n-1)} \right), \tag{A4.35}
\]
when \( n \geq 3 \). By Theorem 3.4 in Houdré and Reynaud-Bouret (2003), for any \( u > 0 \), we have
\[
\mathbb{P}(|D_n(\tilde{g})| \geq C_1 n \tilde{\sigma} u^{1/2} + C_2 \tilde{D} u/4 + C_3 \tilde{B} u^{3/2} + C_4 \tilde{B}_g u^2/4) \leq C_5 e^{-u}, \tag{A4.36}
\]
where positive absolute constants \( C_1, \ldots, C_5 \) are as defined in (A3.2). Combining (A4.35) and (A4.36), we have
\[
\mathbb{P}(\left| \sum_{i=1}^{n} \tilde{f}(Z_i) \right| \geq \frac{t}{2(n-1)}) + \mathbb{P}(\left| D_n(\tilde{g}) \right| \geq C_1 n \tilde{\sigma} u^{1/2} + C_2 \tilde{D} u/4 + C_3 \tilde{B} u^{3/2} + C_4 \tilde{B}_g u^2/4)
\leq 2 \exp \left( \frac{-t^2/n^2}{8n\mathbb{E}[\tilde{f}(X)^2] + 2B_f \cdot t/n} \right) + C_5 e^{-u}. \tag{A4.37}
\]

It is easy to see that \( \tilde{B}_g \leq B_g + 3B_f \leq 4B_g, \tilde{B}_f \leq 2B_f, \) and \( \mathbb{E}[\tilde{f}(Z)^2] \leq \mathbb{E}[f(Z)^2] \). It remains to bound \( \tilde{\sigma}^2, \tilde{B}, \) and \( \tilde{D} \).

By some algebra, we have
\[
\mathbb{E}[\tilde{g}(X_1, X_2)^2|X_2] \leq \mathbb{E}[g(X_1, X_2)^2|X_2],
\]
which implies that
\[
\tilde{\sigma}^2 = \mathbb{E}[\tilde{g}(X_1, X_2)^2]
= \mathbb{E}\left[ \mathbb{E}\left[ g(X_1, X_2)^2 | X_2 \right] | X_2 \right]
\leq \mathbb{E}\left[ \mathbb{E}\left[ g(X_1, X_2)^2 | X_2 \right] | X_2 \right]
= \mathbb{E}[g(X_1, X_2)^2] = \sigma^2,
\]
and that
\[
\tilde{B}^2 \leq n \sup_{X_2} \mathbb{E}[\tilde{g}(X_1, X_2)^2|X_2]
\leq n \sup_{X_2} \mathbb{E}[g(X_1, X_2)^2|X_2] = B^2.
\]
Meanwhile, we have
\[
\mathbb{E}[|\tilde{g}(X_i, X_j)||X_j] \leq 4B_f.
\]

30
By Hölder’s inequality, and combining with the last display, we have
\[
\mathbb{E}[\bar{g}(X_i, X_j)|a_i(X_i)b_j(X_j)] =\mathbb{E}[b_j(X_j)]\mathbb{E}\{\bar{g}(X_i, X_j)|a_i(X_i)|X_j\} \\
\leq \mathbb{E}[b_j(X_j)]\mathbb{E}\{|\bar{g}(X_i, X_j)|X_j\}^{1/2}\mathbb{E}\{|\bar{g}(X_i, X_j)|a_i(X_i)|X_j\}^{1/2} \\
\leq (4B_f)^{1/2}\mathbb{E}[b_j(X_j)]\mathbb{E}\{|\bar{g}(X_i, X_j)|a_i(X_i)|X_j\}^{1/2} \\
\leq (4B_f)^{1/2}\mathbb{E}[b_j(X_j)]^{1/2}\mathbb{E}\{|\bar{g}(X_i, X_j)|a_i(X_i)|X_j\}^{1/2} \\
= (4B_f)^{1/2}\mathbb{E}[a_i(X_i)^2]^{1/2}\mathbb{E}[b_j(X_j)^2]^{1/2}.
\]
Therefore, we further have
\[
\bar{D} \leq 4B_f \sum_{i=2}^{n} \sum_{j=1}^{i-1} \left\{ \mathbb{E}[a_i(X_i)^2]^{1/2}\mathbb{E}[b_j(X_j)^2]^{1/2} \right\} \\
\leq 4B_f \sum_{i=2}^{n} \sum_{j=1}^{i-1} \frac{1}{2} \left\{ \mathbb{E}[a_i(X_i)^2] + \mathbb{E}[b_j(X_j)^2] \right\} \\
\leq 4B_f.
\]
Combining these upper bounds on constants with (A4.37), we complete the proof. \(\square\)

A4.11 Proof of Corollary A3.1

Corollary A4.1 (Corollary A3.1). Suppose Assumptions 6-8 and 10-11 are satisfied.

(1) Assume Assumption 9 holds, and that (A4.20) is satisfied with \(q = 2305s\) and \(t = \kappa \ell M_\ell/16\). Then we have
\[
\mathbb{P}\left(\delta \hat{L}_n(\Delta, h_n) \geq \frac{\kappa \ell M_\ell}{4}\|\Delta\|_i^2 \text{ for all } \Delta \in \{\Delta' \in \mathbb{R}^p : \|\Delta_{S'}\|_1 \leq 3\|\Delta_S\|_1\}\right) \\
\geq 1 - 5.77 \exp(-c \log p) - 2 \exp(-c'n),
\]
where \(c > 1\) is an absolute constant, and \(c' = (\kappa^2 M_\ell^2 \wedge 64\kappa_\ell M_\ell)/(2^{16}((3M^2\kappa_\ell^2 + 2M^2\kappa_\ell^2 C_0^2\kappa_\ell^2) \vee 2M)\kappa_\ell^2)\).

(2) Assume Assumption 16 holds, and that (A4.20) holds with \(q = 2305\{s + \zeta^2 nh_n^2/\log(np)\}\) and \(t = \kappa \ell M_\ell/16\). Then we have
\[
\mathbb{P}\left(\delta \hat{L}_n(\Delta, h_n) \geq \frac{\kappa \ell M_\ell}{4}\|\Delta\|_i^2 \text{ for all } \Delta \in C_{S_n}\right) \\
\geq 1 - 5.77 \exp(-c \log p) - 2 \exp(-c'n),
\]
where \(C_{S_n} = \{v \in \mathbb{R}^p : \|v_{\mathcal{J}}\|_1 \leq 3\|v_{\mathcal{J}}\|_1 \text{ for some } \mathcal{J} \subset [p] \text{ and } |\mathcal{J}| \leq s + \zeta^2 nh_n^2/\log(np)\}\),
\(c > 1\) is an absolute constant, and \(c' = (\kappa^2 M_\ell^2 \wedge 64\kappa_\ell M_\ell)/(2^{16}((3M^2\kappa_\ell^2 + 2M^2\kappa_\ell^2 C_0^2\kappa_\ell^2) \vee 2M)\kappa_\ell^2)\).

Proof. (1) Denote \(C_{S} = \{v \in \mathbb{R}^p : \|v_S\|_1 \leq 3\|v_S\|_1\}\). By Lemma 13 in Rudelson and Zhou (2013), \(C_{S} \cap S_2^{p-3} \subset 2\text{conv}(\bigcup_{|\mathcal{J}| \leq s} E_{\mathcal{J}} \cap S_2^{p-3})\), where \(\text{conv}(\cdot)\) means convex hull of a set, \(E_{\mathcal{J}} = \text{span}\{e_j : j \in \mathcal{J}\}\).
\[ j \in \mathcal{J} \}, \text{ and } d = 2305s. \text{ Denote} \]
\[ \Sigma_{h_n} = \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{\bar{W}}{h_n} \right) \bar{X} \bar{X}^T \right], \]
\[ \Gamma = \left( \frac{n}{2} \right)^{-1} \sum_{i<j} \left\{ \frac{1}{h_n} K \left( \frac{W_{ij}}{h_n} \right) \bar{X}_{ij} \bar{X}_{ij}^T \right\} - \Sigma_{h_n}, \]
\[ \Sigma_0 = \mathbb{E} \left[ \bar{X} \bar{X}^T | \bar{W} = 0 \right] . \]

For any \( v \in \mathcal{C}_S \cap \mathcal{S}_2^{p-1} \), we have
\[ |v^T \Gamma v| \leq 4 \max_{v' \in \text{conv} (|\mathcal{J}| \leq d E_\mathcal{J} \cap \mathcal{S}_2^{p-1})} v'^T \Gamma v' \]
\[ = 4 \max_{v' \in |\mathcal{J}| \leq d E_\mathcal{J} \cap \mathcal{S}_2^{p-1}} v'^T \Gamma v' \]
\[ = 4 \| \Gamma \|_{2,d}, \]

where the second line is because maximum of \( v'^T \Gamma v' \) occurs at extreme points of set \( \text{conv}( \bigcup_{|\mathcal{J}| \leq d} E_\mathcal{J} \cap \mathcal{S}_2^{p-1}) \). Apply Theorem A3.1 with \( q = d = 2305s \) and \( t = \kappa \ell M \ell / 16 \), when (A4.20) is satisfied, we have
\[ |v^T \Gamma v| \leq \frac{\kappa \ell M \ell}{4} \]
holds simultaneously for all \( v \in \mathcal{C}_S \cap \mathcal{S}_2^{p-1} \) with probability at least \( 1 - 5.77 \exp(-c \log p) - 2 \exp(-c' n) \), where \( c > 1 \) is some absolute constant and \( c' = (\kappa \ell^2 M \ell^2 \wedge 64 \kappa \ell M \ell)/[65536 \{2M^2 \kappa^2_x + 2M^{2.5} C \wedge \kappa^2_x + M^2 \kappa^2_x \} \vee 2M \} \kappa^2_x] \).

(A4.38) further implies that \( \delta \hat{L}_n(v, h_n) \geq v^T \Sigma_{h_n} v - \kappa \ell M \ell / 4 \), where
\[ v^T \Sigma_{h_n} v \geq v^T \Sigma_0 v - MMK \mathbb{E} \left[ (\bar{X}^T v)^2 \right] h_n \]
\[ \geq \kappa \ell M \ell \| v \|_2^2 - MMK \cdot 2\kappa^2_x \| v \|_2^2 \cdot h_n \]
\[ \geq \kappa \ell M \ell \| v \|_2^2 / 2 = \kappa \ell M \ell / 2. \]

Therefore \( \delta \hat{L}_n(v, h_n) \geq \kappa \ell M \ell / 4 \) holds simultaneously for all \( v \in \mathcal{C}_S \cap \mathcal{S}_2^{p-1} \) with probability at least \( 1 - 5.77 \exp(-c \log p) - 2 \exp(-c' n) \). By linearity of \( \delta \hat{L}_n(v, h_n) \), this completes the proof for (1).

(2) Using an identical argument as used in (1), replacing \( \mathcal{C}_S \) by set
\[ \left\{ v \in \mathbb{R}^p : \| v_{\mathcal{J}} \|_1 \leq 3 \| v_{\mathcal{J}} \|_1 \text{ for some } \mathcal{J} \subset [p] \text{ and } |\mathcal{J}| \leq s + \kappa^2 \gamma h^2_n / \log,np \right\}, \]
and using \( d = 2305 \{ s + \kappa^2 \gamma h^2_n / \log(np) \} \) instead, we complete the proof for (2). \( \square \)

### A4.12 Proof of Lemma 3.1

**Lemma A4.10** (Lemma 3.1). Assume \( h_n \geq K_1 \{ \log(np)/n \}^{1/2} \) for positive absolute constant \( K_1 \), and assume \( h_n \leq C_0 \) for positive constant \( C_0 \). We further assume that \( u \) satisfies Assumption 17, and take \( c, c' < 3c/4 + 1/2 \) to be positive absolute constants. We take \( \xi = (1 + c')/(2 + c) \), and suppose we have
\[ n > \max \left\{ \left\lfloor \{ 16(c + 2)^2(c + 1)C_0^2 Mu^2(2+\epsilon) \}^{1/(3-2\epsilon)} \vee 1 \right\} \cdot (\log p)^{2/(3-2\epsilon)}, \{ \log(np) \}^{5/(3-4\epsilon)} \right\}, \]

32
Then under Assumptions 7, 8, 11, and 17, we have
\[ \mathbb{P}(\max_{k \in [p]} \{U_k - \mathbb{E}[U_k]\}) \geq C\{\log(np)/n\}^{1/2} \leq 4.77\exp(-c\log p) + \exp(-c'\log n), \]
where \( C = \sqrt{2C_1M_1^{1/2}M_f^{1/2}M_u^{1/(2+\varepsilon)}\kappa_x\epsilon^{1/2}K_1^{-1/2} + 2C_2M_f(c + 1/2)^{1/2}c + 8M_fM_u^{1/(2+\varepsilon)}\kappa_x\epsilon^{1/2} + 2C_3M_1^{1/2}M_f^{1/2}(c + 2)^{1/2}\epsilon^{3/2}K_1^{-3/2} + 2C_4M_1(c + 2)^{1/2}\epsilon^{2}K_1^{-1}. \) Here \( M_f = M + MM_KC_0, \) and \( C_1, \ldots, C_4 \) are as defined in (A3.2).

**Proof.** We apply truncation on \( \tilde{X}_{ijk} \) and \( \tilde{u}_i \) at levels \( \tau_n \) and \( \theta_n/2 \) respectively, and first focus on U-statistic
\[
\tilde{U}_k = \left(\frac{n}{2}\right)^{-1} h_n K\left(\frac{W_{ij}}{h_n}\right) \tilde{X}_{ijk}\tilde{u}_{ij} \mathbb{I}(\mathcal{A}_{k,ij} \cap \mathcal{B}_i \cap \mathcal{B}_j),
\]
where we denote events
\[
\mathcal{A}_{k,ij} = \{ |\tilde{X}_{ijk}| \leq \tau_n \}, \quad \mathcal{B}_i = \{|u_i - \mathbb{E}[u]| \leq \theta_n/2\}.
\]
We also denote events
\[
\mathcal{A}_{k,[n]} = \{ |\tilde{X}_{ijk}| \leq \tau_n, \ i < j \in [n]\}, \quad \mathcal{B}_{[n]} = \{|u_i - \mathbb{E}[u]| \leq \theta_n/2, \ i \in [n]\}.
\]

Denote
\[
g(D_i, D_j) = \frac{1}{h_n} K\left(\frac{W_{ij}}{h_n}\right) \tilde{X}_{ijk}\tilde{u}_{ij} \mathbb{I}(\mathcal{A}_{k,ij} \cap \mathcal{B}_i \cap \mathcal{B}_j), \text{ and } f(D_i) = \mathbb{E}[g(D_i, D_j) | D_i].
\]
We complete the proof in two steps.

**Step I.** We bound \( B_g, B_f, \mathbb{E}[f(D_2)^2], \sigma^2, \) and \( B^2 \) as in Lemma A3.4, and apply Lemma A3.4.

For bounding \( B_g \), we have \( B_g \leq M_K\tau_n\theta_n/h_n \). For bounding \( B_f \), apply Lemma A3.3 on \( \varphi = 1 \) with lemma conditions satisfied by 7 and 8, and we have
\[
B_f \leq \tau_n\theta_n \mathbb{E}\left[\left|\frac{1}{h_n} K\left(\frac{W_1 - W_2}{h_n}\right) W_1\right|\right] \leq M_f\tau_n\theta_n,
\]
where \( M_f = M + MM_KC_0 \).

For bounding \( \sigma^2 \), we have
\[
\sigma^2 = \mathbb{E}\left[g(D_1, D_2)^2\right] \leq \frac{M_K}{h_n} \mathbb{E}\left[\frac{1}{h_n} K\left(\frac{W_{ij}}{h_n}\right) \tilde{X}_{ijk}\tilde{u}_{ij}\right] \leq \frac{M_K}{h_n} \mathbb{E}\left[\tilde{X}_{ijk}\tilde{u}_{ij}^2\right] | W_{ij} = 0 | M + MM_KC_0 \mathbb{E}\left[\tilde{X}_{ijk}\tilde{u}_{ij}^2\right] \leq 2M_KM_fM_u^{2/(2+\varepsilon)}\kappa_x^2/h_n,
\]
where the first inequality is due to \( K(\cdot) \in [0, 1] \), the second inequality is by applying Lemma A3.2 on \( Z = \tilde{X}_{ijk}\tilde{u}_{ij} \) with lemma assumptions satisfied by Assumptions 7 and 8, and the last inequality is by Assumptions 11, 12, and independence of \( \tilde{X}_{ijk} \) and \( \tilde{u}_{ij} \).

For bounding \( \mathbb{E}[f(D_2)^2], \) apply Lemma A3.3 on \( \varphi = \tilde{X}_{ijk}\tilde{u}_{ij} \mathbb{I}(\mathcal{A}_{k,ij} \cap \mathcal{B}_i \cap \mathcal{B}_j), \) with lemma assumptions satisfied by Assumptions 7 and 8, and we have \( |f(D_2) - f_1(D_2)| \leq MM_KC_0f_2(D_2), \)

\[
33
\]
where
\[ f_1(D_2) = \mathbb{E}[\tilde{X}_{12k} \tilde{u}_{12} \mathbb{I}(A_{k,12} \cap B_1 \cap B_2) | W_1 = W_2, D_2] \cdot f_{W_1}(W_2), \]
\[ f_2(D_2) = \mathbb{E}[\tilde{X}_{12k} \tilde{u}_{12} | \mathbb{I}(A_{k,12} \cap B_1 \cap B_2) | D_2]. \]

We have, by Assumptions 11, 12, and independence of \( \tilde{X}_{ijk} \) and \( \tilde{u}_{ij} \),
\[ \mathbb{E}[f_1(D_2)^2] \leq \mathbb{E}[\tilde{X}_{12k}^2 \tilde{u}_{12}^2 | W_1 = W_2] M^2 \leq 2MM_u^2/(2+\epsilon)\kappa_x^2, \]
\[ \mathbb{E}[f_2(D_2)^2] \leq \mathbb{E}[\tilde{X}_{12k}^2 \tilde{u}_{12}^2] \leq 2M_u^2/(2+\epsilon)\kappa_x^2. \]
This further implies that
\[ \mathbb{E}[f(D_2)^2] \leq 2\mathbb{E}[f_1(D_2)^2] + 2M^2M^2_0C_0^2 \mathbb{E}[f_2(D_2)^2] \leq 4M^2M_u^2/(2+\epsilon)\kappa_x^2. \]
For bounding \( B^2 \), we have
\[ B^2 = n \sup_{D_2} \mathbb{E}[g(D_1, D_2)^2 | D_2] \]
\[ \leq n\frac{MK}{h_n^2} \sup_{D_2} \mathbb{E}\left[ \frac{1}{h_n}K\left(\frac{W_1 - W_2}{h_n}\right)(X_{1k} - X_{2k})^2(u_1 - u_2)^2 \mathbb{I}(A_{k,12} \cap B_1 \cap B_2) | D_2 \right] \]
\[ \leq MK\frac{n\tau_n^2\theta_n^2}{h_n}. \]

We take for some positive absolute constant \( c > 1 \),
\[ t = 8M_fM^1_u/(2+\epsilon)\kappa_xc^{1/2} \cdot \left(\frac{n}{2}\right) \{\log(np)/n\}^{1/2}, \]
\[ \tau_n = \max \{c, 2\}^{1/2} \cdot \{\log(np)\}^{1/2}, \quad \theta_n = n^\alpha, \quad 0 < \alpha < 3/4, \]
\[ c_u = c \log p, \]
and we have that
\[ n > \max \left\{ \left[\{16c^3(c + 1)C^2_0M_u^2/(2+\epsilon)\kappa_x^2 \right]^{1/(3-2\alpha)} \cdot (\log p)^{2/(3-2\alpha)}, \{\log(np)\}^{5/(3-4\alpha)} \right\}. \]
Then by Lemma A3.4, we have
\[ \mathbb{P}\left\{ \left(\frac{n}{2}\right)^{-1} |\tilde{U}_k - \mathbb{E}[\tilde{U}_k]| \geq A\{\log(np)/n\}^{1/2} \right\} \leq 2 \exp(-c \log(np)) + 2.77 \exp(-c \log p) \]
\[ \leq 4.77 \exp(-c \log p), \]
where with \( C_1, \ldots, C_4 \) defined in (A3.2),
\[ A = 2\sqrt{2}C_1M^1_KM^1_fM^1_u/(2+\epsilon)\kappa_xc^{1/2}K_1^{-1/2} + 2C_2M_f(c + 1/2)^{1/2}c \]
\[ + 2C_3M^1_KM^1_f(c + 2)^{1/2}c^{3/2}K_1^{-1/2} + 2C_4M_K(c + 2)^{1/2}c^2K_1^{-1} + 8M_fM^1_u/(2+\epsilon)\kappa_xc^{1/2}. \]
Step II. We have $\mathbb{E}[\hat{U}_k] = 0$, and thus we have

\[
\mathbb{P}\left( \max_{k \in [p]} \left\{ |U_k - \mathbb{E}[U_k]| \right\} \geq A\{\log(np)/n\}^{1/2} \right) \\
\leq \mathbb{P}\left( \max_{k \in [p]} \left\{ |U_k - \mathbb{E}[U_k]| \right\} \geq A\{\log(np)/n\}^{1/2} \cap B_{[n]} \right) + \mathbb{P}(B_{[n]}^c)
\]

\[
\leq \sum_{k=1}^{p} \left\{ \mathbb{P}(|U_k| > A\{\log(np)/n\}^{1/2} \cap \mathcal{A}_{k,[n]} \cap B_{[n]}) + \mathbb{P}(\mathcal{A}_{k,[n]}^c) \right\} + \mathbb{P}(B_{[n]}^c)
\]

\[
\leq 4.77 \exp(-c \log p + \log p) + \frac{\mathbb{E}[u]^{2+\epsilon}}{n^{\alpha(2+\epsilon)}}
\]

\[
\leq 4.77 \exp(-c \log p + \log p) + \exp(-c' \log n).
\]

The last inequality holds if we take $(c' + 1)/(2 + \epsilon) < 3/4$ and we take $\alpha = (c' + 1)/(2 + \epsilon)$. This completes the proof.

\[\square\]

A4.13 Proof of Corollary A2.1

Assume $h_n \geq K_1\{\log(np)/n\}^{1/2}$ for positive absolute constant $K_1$, and assume $h_n \leq C_0$ for positive constant $C_0$. We further assume that $u$ satisfies Assumption 17, and take $c$ and $c' < 3\epsilon/4 + 1/2$ to be positive absolute constants. We take $\xi = (1 + c')/(2 + \epsilon)$, and suppose we have

\[
n > \max \left\{ \left\{ 16(c + 2)^3(c + 1)C_0^2M_u^{2/(2+c)}\kappa_x^2 \right\}^{1/(3-2\epsilon)} \right\} \cdot (\log p)^{2/(3-2\epsilon)} \cdot \{\log(np)\}^{5/(3-4\epsilon)},
\]

\[
64(c + 2)^2(c + 1)\{\log(np)\}^{3/3}, 3,
\]

\[
\frac{48\sqrt{6}M_K\kappa_x^2 q}{K_1 p \{\log(np)\}^{1/2}}, \frac{2^{10} \cdot 6 \cdot \sqrt{6} M_f \kappa_x^2 q}{K_f \kappa_x^{2} 2^{3/2}}, \frac{144\kappa_x^{4}}{K_f^{2} p^{2} \log(np)},
\]

\[
\frac{2^{11} \cdot 6 \cdot \sqrt{3}(2 + c)^{1/2} C_1 M_f^{1/2} M_f^{1/2} \kappa_x^2}{K_1^{1/2} \kappa_f M_f}, \frac{4^{3}}{q^{4/3} \{\log(np)\}^{1/3}},
\]

\[
\frac{2^{8} \cdot 6 \cdot (20 + 7.5c)(c + 2) C_2 M_f \kappa_x^2}{\kappa_f M_f}, \frac{q^{1/2} \log(np)},
\]

\[
\frac{2^{8} \cdot 6(c + 2)^{3/2} C_3 \{144(2 + c)^2 M_K M_f \kappa_x^2 K_1^{-1} + 192 M_f \kappa_x^2 + 8 M_f \kappa_x^2 \}^{1/2}}{\kappa_f M_f}, \frac{q^{2/3} \{\log(np)\}^{5/3}}{q},
\]

\[
\frac{2^{10} \cdot 6 \cdot \sqrt{6}(2 + c)^{3} C_4 \kappa_x^2}{K_1 \kappa_f M_f}, \frac{q^{2/3} \{\log(np)\}^{5/3}}{q},
\]

\[
\frac{2^{11} \cdot 6 \cdot (20 + 7.5c)(c + 2) M_f \kappa_x^2}{\kappa_f M_f}, \frac{q \{\log(np)\}^{2}}{q}, \frac{2^{6} \cdot 3q}{(20 + 7.5c) M_f \kappa_x^2 \kappa_f M_f \log(np)},
\]

\[
\frac{2^{20} \{(3M_f^2 \kappa_x^2 + 2M_f^2 M_f^2 C_0^2 \kappa_x^2) \vee 2M_f \} \kappa_x^2}{\{\kappa_f M_f \}^{2}}, \frac{6ep \log \left( \frac{q}{q} \right)}{q},
\]

\[
\frac{2^{24} K_f^2 M_f^2 C_0^2 \kappa_x^2}{\{\kappa_f M_f \}^{2}}.
\]
where \( q \) is to be determined in specific cases. Denote \( M_f = M + MM_KC_0 \), and \( C_1, \ldots, C_4 \) are as defined in (A3.2). Also denote \( c \) to be some positive absolute constant, and
\[
A' = \sqrt{2}C_1 M^{1/2}_K M^{1/2}_f M^{1/2}_u \kappa_c c^{1/2} K_1^{-1/2} + 2C_2 M_f (c + 1/2)^{1/2} \kappa_c c^{1/2} + 2C_3 M^{1/2}_K M^{1/2}_f (c + 2)^{1/2} K_1^{-1/2} + 2C_4 M_K (c + 2)^{1/2} \kappa_c c^{1/2},
\]
\[
\epsilon'' = (\kappa_c^2 M^2_f \wedge 64 \kappa_f M_f) \big/ [216(3M^2 \kappa_c^2 + 2M^2 M_K C_0^2 \kappa_e^2) \vee 2M^2 \kappa_c^2].
\]

**Theorem A4.11** (Corollary A2.1(1)). Assume \( \lambda_n \geq 4(A + A') \{ \log(np)/n \}^{1/2} + 8\kappa_c^2 M_f \zeta h_n \). Further assume (A4.40) holds with \( q = 2305s \). Then under Assumptions 6-11, 14, 15, and 17, we have
\[
\| \widehat{\beta}_{h_n} - \beta^* \|_2^2 \leq \frac{288 s \lambda_n^2}{M^2_f \kappa_c^2},
\]
with probability at least \( 1 - 10.54 \exp(-c \log p) - \exp(-c' \log n) - 2 \exp(-c''n) - \epsilon_n \cdot p \).

*Proof.* See Proof of Theorem A4.12. \( \square \)

**Theorem A4.12.** [Corollary A2.1(2)] Assume \( \lambda_n \geq 4(A + A') \{ \log(np)/n \}^{1/2} + 8\kappa_c^2 M_f \zeta h_n \). Further assume (A4.40) holds with \( q = 2305s \). Then under Assumptions 6-11, 14, 15, and 17, we have
\[
\| \widehat{\beta}_{h_n} - \beta^* \|_2^2 \leq \frac{288 s \lambda_n^2}{M^2_f \kappa_c^2},
\]
with probability at least \( 1 - 10.54 \exp(-c \log p) - \exp(-c' \log n) - 2 \exp(-c''n) - \epsilon_n \cdot p \).

*Proof.* We adopt the framework as described in Section 2.1 for \( \theta^* = \beta^* \), \( \Gamma_0(\theta) = L_0(\beta), \widehat{\Gamma}_n(\theta, h) = \widehat{L}_n(\beta, h), \) and take \( \theta^*_{h_n} = \beta^* \), which yields \( s_n \leq s \) and \( \rho_n = 0 \).

We verify Assumption 2, by using results (A4.4), (A4.6), and applying Lemma 3.1. We verify Assumption 3 by applying Corollary A3.1. We complete the proof by Theorem 2.1. \( \square \)

**Theorem A4.13** (Corollary A2.1(3)). Denote \( C \) to be some positive absolute constant \( C > \zeta^2 C_0^2 \gamma \), and suppose \( n \geq (C - \zeta^2 C_0^2 \gamma) s \log(np) \). Assume \( \lambda_n \geq 4(A + A + Mm_n) \{ \log(np)/n \}^{1/2} + 8MM_K C_1^{1/2} \kappa_c^2 h_n \). Further assume that (A4.40) holds with \( q = 2305 \{ s + \zeta^2 n h_n^2 \gamma / \log(np) \} \). Then under Assumptions 6-8, 10-11, 14-16 and 17, we have
\[
\| \widehat{\beta}_{h_n} - \beta^* \|_2^2 \leq \frac{288 s \lambda_n^2}{M^2_f \kappa_c^2} + \frac{2 s \log(np)}{n} + \left\{ \frac{288 n \lambda_n^2}{M^2_f \kappa_c^2 \log(np)} + 2 \right\} \cdot \zeta^2 h_n^2 \gamma,
\]
with probability at least \( 1 - 17.31 \exp(-c \log p) - \exp(-c' \log n) - 2 \exp(-c''n) - \epsilon_n \cdot p \).

*Proof.* We adopt the framework as described in Section 2.1 for \( \theta^* = \beta^* \), \( \Gamma_0(\theta) = L_0(\beta), \widehat{\Gamma}_n(\theta, h) = \widehat{L}_n(\beta, h), \) and take \( \theta^*_{h_n} = \beta^* \), which yields \( s_n \leq s \) and \( \rho_n = 0 \).

We verify Assumption 2, by using results (A4.4), (A4.6), and applying Lemma 3.1. We verify Assumption 3 by applying Corollary A3.1. We complete the proof by Theorem 2.1. \( \square \)
Corollary A4.2 (Corollary A2.1(4)). Denote
\[ \tau_1 = \sqrt{2}(2 + c)^{1/2} \kappa_x K_{15}^{-1} (BMK C_0^a + D M_K), \]
\[ \tau_2 = \sqrt{2}(2 + c)^{1/2} \kappa_x \{ BMK M (1 + C_0) C_0^a + DM_f \}, \]
\[ \tau_3 = 4M_K^2 M^2 \cdot \left( BC_0^a + D \right)^2 \cdot (1 + C_0^a) \cdot \kappa_x^2, \]
\[ \tau_4 = \{ 4B^2 MMK \kappa_x^2 (1 + C_0) C_0^{2a-\gamma_i} + 2D^2 \cdot (12M_f \kappa_x^4)^{1/2} \cdot E^{1/2} C_0^{-1/2-\gamma_i} \} \cdot M_K K_i^{\gamma_i}, \]
\[ \tau_5 = 4(2 + c) \kappa_x^2 \{ BMK M (1 + C_0) C_0^{2a} + D^2 M_f \} M_K K_i^{-1}, \]
and
\[ A'' = 4\tau_3^{1/2} (1 + c)^{1/2} + 2C_1 \tau_4^{1/2} (1 + c)^{1/2} + 2C_2 \tau_2 (1 + c) + 2C_3 \tau_5^{1/2} (1 + c)^{3/2} + 2C_4 \tau_1 (1 + c)^2 + 4M_f \cdot (BC_0^a + D) \cdot (c + 2) \kappa_x, \]
where \( \gamma_1 = \min \{ 2a - 1, -1/2 \} \). Consider lower bound on \( n \),
\[ n > \max \left\{ 64(c + 2)^2 (c + 1) \tau_2^{1/2} \tau_3^{-1} \{ \log(np) \}^{1/4}, \{ \log(np) \}^{5/3} \right\}. \tag{A4.41} \]
Here, \( B, D, E \) and \( a \) are to be specified in different cases.

(1) Assume that \( g \) is \((L, \alpha)\)-Hölder for \( \alpha \geq 1 \), and \( g \) has bounded support when \( \alpha > 1 \). Suppose that (A4.40) holds with \( q = 2305s \), and that (A4.41) holds with \( B = L_a \), where \( L_a \) is the Lipschitz constant for \( g \) \((L_a = L \text{ when } 1) \), \( D = E = 0 \), \( a = 1 \). Further assume that \( \lambda_n \geq 4(A'' + A') \{ \log(np)/n \}^{1/2} + 8 \kappa_x^2 M_f \zeta h_n \), where
\[ \zeta = \max \left\{ 4 \cdot \left( \frac{L_a^2 MMMK + MMMK E\tilde{u}^2/2}{\kappa_x M_f} \right)^{1/2}, \frac{16 \kappa_x (M + MMMK C_0^{2a})^{1/2} \cdot L_a^2 MMMK}{\kappa_x M_f} \right\}. \]
Then under Assumptions 6-8, 9', 10-11, 13, and 17, we have
\[ \| \tilde{\beta}_h - \beta^* \|_2 \leq \frac{288 \lambda_n^2}{M_f \kappa_x^2}, \]
with probability at least \( 1 - 15.81 \exp(-c \log p) - \exp(-c' \log n) - 2 \exp(-c''n) \).

(2) Assume that Assumption 5 holds with \( \alpha \in (0, 1] \). Suppose that (A4.40) holds with \( q = 2305s \), and that (A4.41) holds with \( B = M_g \), \( D = M_d \), \( E = M_a \) and \( a = \alpha \). Assume \( \lambda_n \geq 4(A'' + A') \{ \log(np)/n \}^{1/2} + 8 \kappa_x^2 M_f \zeta h_n^\gamma \), where
\[ \zeta = \max \left\{ 4 \cdot \left( \frac{M_g^2 MMMK C_0^{2a-2\gamma} + M_d^2 M_a C_0^{1-2\gamma} + MMMK E\tilde{u}^2 C_0^{2-2\gamma}/2}{\kappa_x M_f} \right)^{1/2}, \frac{16 \kappa_x (M + MMMK C_0^{2a})^{1/2} \cdot (M_g^2 MMMK C_0^{2a-2\gamma} + M_d^2 M_a C_0^{1-2\gamma})^{1/2}}{\kappa_x M_f} \right\}, \]
\[ \gamma = \alpha \text{ if } M_d M_a = 0, \text{ and } \gamma = \min \{ \alpha, 1/2 \} \text{ if otherwise.} \]
Then under Assumptions 6-8, 9', 10-11, 13, and 17, we have
\[ \| \tilde{\beta}_h - \beta^* \|_2 \leq \frac{288 \lambda_n^2}{M_f \kappa_x^2}, \]
with probability at least \( 1 - 15.81 \exp(-c \log p) - \exp(-c' \log n) - 2 \exp(-c''n) \).

(3) Assume that Assumption 5 holds with \( \alpha \in [1/4, 1] \). Suppose that (A4.40) holds with
q = 2305\{s + \zeta^2 n h^2 / \log(n p)\}, and that \((A4.41)\) holds with \(B = M_4, D = M_4, E = M_4\) and \(a = \alpha\). Further assume \(\lambda_n \geq 4(A' + A'' + M_2)\{n \log(n p)\}^{1/2} + 8M M K C_2 h_n\), where
\[
\zeta = \max \left\{ 4 \cdot \left( \frac{M_4^2 M M K C_0(2\alpha - 2\gamma + M_4 M_4 C_0^{-1} - 2\gamma + M M K 2\alpha^{-2\gamma} + M_4 M_4 C_0^{-2\gamma} / 2)}{16\kappa x (M + M M K C_0^2)} \right)^{1/2}, \frac{16\kappa x (M + M M K C_0^2)^{1/2} \cdot (M_4^2 M M K C_0(2\alpha - 2\gamma + M_4 M_4 C_0^{-1} - 2\gamma)^{1/2})}{\kappa x M_4} \right\},
\]
\[\gamma = \alpha \text{ if } M_4 M_4 = 0, \text{ and } \gamma = \min \{ \alpha, 1/2 \} \text{ if otherwise.} \]

Then under Assumptions 6-8, 9', 10-11, and 17, we have
\[
\| \hat{\beta}_{h_n} - \beta^* \|^2 \leq \frac{288 s \lambda_n^2}{M_4^2 \kappa x^2} + 2 s \log(n p) + \left\{ \frac{288 n \lambda_n^2}{M_4^2 \kappa x^2 \log(n p)} + 2 \right\} \cdot \zeta^2 h_n^2,
\]
with probability at least \(1 - 22.58 \exp(-c \log p) - \exp(-c' \log n) - 2 \exp(-c'' n)\).

**Proof.** The result follows directly from Corollary A2.1(1)-(3). \(\square\)

**Theorem A4.14 (Corollary A2.1(5)).** Assume that \((A4.40)\) holds with \(q = 2305s\). Assume further that \(n > 64(c + 2)\{c + 1\} \{n \log(n p)\}^4\) and \(\lambda_n \geq 4(A' + A'')\{n \log(n p)\}^{1/2} + 4\sqrt{2} M_4 M M K M_4 (1 + C_0) h_n\), where
\[
A'' = 8M M K M_4 C_0(1 + C_0) \kappa x (1 + c)^{1/2} + 2C_1 M_4 M_4^{1/2} M_4^{3/2} \kappa x^{1/2} (1 + C_0)^{1/2} C_0^{5/4} K_1^{-1/4} (1 + c)^{1/2}
\]
\[+ 2\sqrt{2} C_2 M M K M_4 (1 + C_0) \kappa x K_1 (1 + c)^{3/2} + 4C_3 M M K M_4^{1/2} (1 + C_0)^{1/2} C_0^{1/2} \kappa x (1 + c)^2
\]
\[+ 2\sqrt{2} C_4 M M K M_4 C_0 \kappa x K_1^{-1} (1 + c)^{5/2} + 2\sqrt{2} M M K M_4 (1 + C_0) C_0 \kappa x,
\]
Then under Assumptions 6-11, 4, and 17 we have
\[
\| \hat{\beta}_{h_n} - \beta^* \|^2 \leq \frac{288 s \lambda_n^2}{M_4^2 \kappa x^2},
\]
with probability at least \(1 - 15.81 \exp(-c \log p) - \exp(-c' \log n) - 2 \exp(-c'' n)\).

**Proof.** We adopt the framework as described in Section 2.1 for \(\theta^* = \beta^*, \Gamma_0(\theta) = L_0(\beta), \hat{\Gamma}_n(\theta, h) = \hat{L}_n(\beta, h), \Gamma_h(\theta) = \mathbb{E} \hat{L}_n(\beta, h), \) and take \(\hat{\theta}^*_n = \beta^*\), which yields \(s_n \leq s\) and \(\rho_n = 0\).

We verify Assumption 2 by using results \((A4.43), (A4.45), (A4.46), (A4.47)\), and applying Lemma 3.1. We verify Assumption 3 by applying Corollary A3.1. We complete the proof by Theorem 2.1. \(\square\)

### A4.14 Supporting lemmas

**Lemma A4.15.** Assumption 7 implies that, for any \(0 < a < 3\) and \(0 < b < 1\), we have
\[
\int_{-\infty}^{+\infty} |w|^a K(w) \, dw \leq M_K \text{ and } \sup_{w \in \mathbb{R}} |w|^b K(w) \leq M_K.
\]

**Proof of Lemma A4.15.** For any \(0 < a < 3\), we have
\[
\int_{-\infty}^{+\infty} |w|^a K(w) \, dw \leq \left\{ \int_{-\infty}^{+\infty} |w|^3 K(w) \, dw \right\}^{a/3} \leq M_K^{a/3} \leq M_K.
\]
where the first inequality is by Hölder’s inequality, the second is by Assumption 7 and that $a > 0$, and the last is by the fact that $0 < a < 3$ and the choice of $M_K \geq 1$.

For any $0 < b < 1$ and any $w \in \mathbb{R}$, we have
\[
|w|^b K(w) = \{|w|K(w)|^b \cdot K(w)^{1-b} \leq M_K^b M_K^{1-b} = M_K,
\]
where the first inequality is by Assumption 7 and that $0 < b < 1$. Therefore, we have obtained that $\sup_{w \in \mathbb{R}} |w|^b K(w) \leq M_K$. This completes the proof. 

\[\square\]

**Lemma A4.16.** Assumption 8 implies that, for any $\tilde{X}$-measurable function $\psi(\cdot) : \mathbb{R}^p \to \mathbb{R}^m$ mapping to a $m$-dimensional real space, we have
\[
\sup_{w,z} \left\{ \left| \frac{\partial f_{\tilde{W}|\psi(X)}(w,z)}{\partial w} \right|, f_{\tilde{W}|\psi(X)}(w,z), \left| \frac{\partial f_{\tilde{W}}(w)}{\partial w} \right|, f_{\tilde{W}}(w) \right\} \leq M. \tag{A4.42}
\]

**Proof of Lemma A4.16.** For a function $F(\cdot)$, we write $dF(x)/dx = F(x+) - F(x-)$, where $F(x+)$ and $F(x-)$ are right and left limits respectively, when $F(x)$ is discontinuous at $x$. We first show that
\[
\sup_{w,x} \left\{ \left| \frac{\partial f_{\tilde{W}|\tilde{X}}(w,x)}{\partial w} \right|, f_{\tilde{W}|\tilde{X}}(w,x) \right\} \leq M.
\]

We have
\[
F_{\tilde{W}|\tilde{X}=x}(w) = \int \int F_{W_1|X_1=x_2+x}(w_2 + w) \frac{dF_{X_1}(x')}{dx'} \big|_{x' = x_2 + x} dF_{W_2|X_2=x_2}(w_2) dF_{X_2}(x_2).
\]

By dominated convergence theorem, we have
\[
f_{\tilde{W}|\tilde{X}}(w,x) = \int \int f_{W_1|X_1}(w_2 + w, x_2 + x) \frac{dF_{X_1}(x')}{dx'} \big|_{x' = x_2 + x} dF_{W_2|X_2=x_2}(w_2) dF_{X_2}(x_2) \leq M,
\]
and
\[
\left| \frac{\partial f_{\tilde{W}|\tilde{X}}(w,x)}{\partial w} \right| = \left| \int \int \frac{\partial f_{W_1|X_1}(w_2 + w, x_2 + x)}{\partial w} \frac{dF_{X_1}(x')}{dx'} \big|_{x' = x_2 + x} dF_{W_2|X_2=x_2}(w_2) dF_{X_2}(x_2) \right| \leq M.
\]

Based on the same argument, we have
\[
F_{\tilde{W}}(w) = \int F_{\tilde{W}|\tilde{X}=x}(w) dF_{\tilde{X}}(x),
\]
which, by dominated convergence theorem, implies that
\[
f_{\tilde{W}}(w) = \int f_{\tilde{W}|\tilde{X}}(w, x) dF_{\tilde{X}}(x) \leq M,
\]
and
\[
\left| \frac{\partial f_{\tilde{W}}(w)}{\partial w} \right| = \left| \int \frac{\partial f_{\tilde{W}|\tilde{X}}(w, x)}{\partial w} dF_{\tilde{X}}(x) \right| \leq M.
\]

Also, for any $\tilde{X}$-measurable function $\psi(\cdot)$, we have
\[
F_{\tilde{W}|\psi(\tilde{X})=z}(w) = \frac{\partial}{\partial v} \int \{\psi(x) \leq v\} F_{\tilde{W}|\tilde{X}=x}(w) dF_{\tilde{X}}(x)|_{v=z}.
\]
By dominated convergence theorem, we have

\[
\int \frac{\partial}{\partial v} \left( \mathbb{I}\{\psi(x) \leq v\} f_{\tilde{W}|\tilde{X}}(w, x) \right) dF_{\tilde{X}}(x) \big|_{v=z} \leq M,
\]

and

\[
\left| \frac{\partial}{\partial w} \left( \mathbb{I}\{\psi(x) \leq v\} f_{\tilde{W}|\tilde{X}}(w, x) \right) \right| \leq M.
\]

Therefore, Assumption 8 implies (A4.42)

\[\square\]

**Lemma A4.17.** Assumption 11 implies, conditional on \(\tilde{W} = 0\) and unconditionally, \(\langle \tilde{X}, v \rangle\) is mean-zero subgaussian with parameter at most \(2\kappa^2_x\), for any \(v \in \mathbb{R}^p\). Assumption 12 implies that \(\tilde{u}\) is mean-zero subgaussian with parameter at most \(2\kappa^2_u\).

**Proof of Lemma A4.17.** Observe that \(\tilde{X}^T v\) and \(-\tilde{X}^T v\) are identically distributed, and thus we have \(\mathbb{E}[\tilde{X}^T v] = 0\). We have that the moment generating function of \(\tilde{X}^T v\) is

\[
\mathbb{E}[e^{t \tilde{X}^T v}] = \mathbb{E}[e^{t(X^T v - \mathbb{E}[X^T v])}] \cdot \mathbb{E}[e^{t(-X^T v + \mathbb{E}[X^T v])}] \leq e^{t^2 \kappa^2_x \|v\|^2_2},
\]

where the first inequality is because \(X_1\) and \(X_2\) are i.i.d., and the second is an application of Assumption 11. Therefore, \(\tilde{X}^T v\) is mean-zero subgaussian with parameter at most \(2\kappa^2_x\),

Observe that conditional on \(\tilde{W} = 0\), \(\tilde{X}^T v\) and \(-\tilde{X}^T v\) are identically distributed, and thus we have \(\mathbb{E}[\tilde{X}^T v|\tilde{W} = 0] = 0\). We have that the moment generating function of \(\tilde{X}^T v\), conditional on \(\tilde{W} = 0\), is

\[
\mathbb{E}[e^{t \tilde{X}^T v|\tilde{W} = 0}] = \mathbb{E}\left[\mathbb{E}\left[e^{t \tilde{X}^T v}|W_1 = W_2, W_2\right]\right] = \mathbb{E}\left[\mathbb{E}\left[e^{t(X^T v - \mathbb{E}[X^T v]|W_1 = W_2)}|W_1 = W_2\right] \cdot \mathbb{E}\left[e^{t(-X^T v + \mathbb{E}[X^T v]|W_2)}|W_2\right]\right] \leq e^{t^2 \kappa^2_x \|v\|^2_2},
\]

where the second inequality is because \((X_1, W_1)\) and \((X_2, W_2)\) are i.i.d., and the third is an application of Assumption 11. Therefore, conditional on \(\tilde{W} = 0\), \(\tilde{X}^T v\) is mean-zero subgaussian with parameter at most \(2\kappa^2_x\). Apply the same argument on \(u\), we complete the proof.

The following results in Lemma A4.18 can be found in Vershynin (2012).

**Lemma A4.18.** For mean-zero subgaussian random variable \(V\) with parameter at most \(\kappa^2_v\), we have \(\mathbb{E}[V^2] \leq \kappa^2_v\), \(\mathbb{E}[V^4] \leq 3\kappa^4_v\), \(\mathbb{P}(V^2 - \mathbb{E}[V^2] \leq v) \geq 1 - \exp\{-v/(2\kappa^2_v)\}\) for any \(v \geq 2\kappa^2_v\), and that \(\mathbb{E}[e^{sV^2 - s\mathbb{E}[V^2]}/s] \leq e^{2s^2\kappa^2_v}\) for \(|s| \leq (2\kappa^2_v)^{-1}\).

**Lemma A4.19.** Let \(Z\) be some subgaussian random variable, with parameter at most \(\kappa^2_z\). Suppose \(\kappa^2_z \leq a/4\) for some \(a > 0\). Then we have

\[
\int_a^\infty z \, dF_Z(z) \leq (a + 4\kappa^2_z) \exp\{-a/(4\kappa^2_z)\}.
\]

**Proof of Lemma A4.19.** We have \(F_Z(z) = \mathbb{P}(Z^2 - \mathbb{E}[Z^2] \leq z/2) \geq 1 - \exp\{-z/(4\kappa^2_z)\}\) for any
\[ z \geq a \geq 4\kappa_z^2 \quad \text{(Lemma A4.18)} \]

By integration by parts, we have
\[
\int_a^\infty z \, dF_{Z^2}(z) = \int_a^\infty (-z) \, d\{1 - F_{Z^2}(z)\} \\
= (-z)\{1 - F_{Z^2}(z)\} \bigg|_a^\infty + \int_a^\infty 1 - F_{Z^2}(z) \, dz \\
\leq a \exp\{-a/(4\kappa_z^2)\} + \int_a^\infty \exp\{-z/(4\kappa_z^2)\} \, dz \\
= (a + 4\kappa_z^2) \exp\{-a/(4\kappa_z^2)\}.
\]

This completes the proof. \(\square\)

The following Lemma A4.20 is used in the proof of Theorem 2.2 to directly verify Assumption 2.

**Lemma A4.20.** Assume \( h_n \geq K_1 \{\log(np)/n\}^{1/2} \) for positive absolute constant \( K_1 \), and assume \( h_n \leq C_0 \) for positive constant \( C_0 \). Further assume \( \lambda_n \geq 4(A + A') \cdot \{\log(np)/n\}^{1/2} + 4\sqrt{2}M_{M \kappa_0}(1 + C_0)h_n \). Here, \( A' \) is as specified in (A4.48), and \( A'' \) as in (A4.53). Suppose we have

\[ n > \max \{64(c + 2)^2(c + 1)\{\log(np)\}^3/3, 64(c + 2)^3(c + 1)\{\log(np)\}^4, \{\log(np)\}^{5/3}, 3\}, \]

for positive absolute constant \( c > 0 \). Then under Assumptions 7, 8, and 11, 4, 12, we have

\[ \mathbb{P}(2|\nabla_k \hat{L}_n(\beta^*, h_n)| \leq \lambda_n \text{ for all } k \in [p]) \geq 1 - 12.04 \exp(-c \log p). \]

**Proof of Lemma A4.20.** Denote

\[ U_{1k} = \left( \begin{array}{c} n \end{array} \right)^{-1} \sum_{i<j} 1 h_n K \left( \frac{\hat{W}_{ij}}{h_n} \right) \tilde{X}_{ijk} \tilde{u}_{ij} \]

\[ U_{2k} = \left( \begin{array}{c} n \end{array} \right)^{-1} \sum_{i<j} 1 h_n K \left( \frac{\hat{W}_{ij}}{h_n} \right) \tilde{X}_{ijk} \{g(W_i) - g(W_j)\}, \]

and observe that

\[ |\nabla_k \hat{L}_n(\beta^*, h_n)| \leq 2 \{ |U_{1k} - \mathbb{E}[U_{1k}]| + |\mathbb{E}[U_{1k}]| + |U_{2k} - \mathbb{E}[U_{2k}]| + |\mathbb{E}[U_{2k}]| \}. \quad \text{(A4.43)} \]

Apply Lemma A4.21 on \( D_i = (X_i, u_i, W_i) \), with conditions of lemma satisfied by Assumptions 7, 8, 11, 12, we have

\[ \mathbb{P}(|U_{1k} - \mathbb{E}[U_{1k}]| \geq A\{\log(np)/n\}^{1/2}) \leq 6.77 \exp\{-(c + 1) \log p\}, \quad \text{(A4.44)} \]

for positive absolute constant \( A \) and \( c \), and when assuming \( n > \max \{64(c+2)^2(c+1)\{\log(np)\}^3/3, 3\} \). Here \( A \) is as specified in (A4.48).

Apply Lemma A4.22 on \( D_i = (X_i, g(W_i), W_i) \), with conditions of lemma satisfied by Assumptions 7, 8, 11, 4, we have

\[ \mathbb{P}(|U_{2k} - \mathbb{E}[U_{2k}]| \geq A'\{\log(np)/n\}^{1/2}) \leq 5.27 \exp\{-(c + 1) \log p\}, \quad \text{(A4.45)} \]

for positive constants \( A' \) and \( c \), and when assuming \( n > \max \{64(c+2)^3(c+1)\{\log(np)\}^4, \{\log(np)\}^{5/3}\} \). Here \( A' \) is as specified in (A4.53).
By independence of $u$ and $(X, W)$, we have
\[ \mathbb{E}[U_{1k}] = 0. \] (A4.46)

We also have
\[ |\mathbb{E}[U_{2k}]| \leq M_2 \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{\tilde{W}_{ij}}{h_n} \right) |\tilde{X}_{ijk} \tilde{W}_{ij}| \right] \]
\[ = M_2 \int \int K(w) |xwh_n| f_{\tilde{W}_{ij} | \tilde{X}_{ijk}}(w, x) \, dw \, dF_{\tilde{X}_{ijk}}(x) \]
\[ = M_2 \int \int K(w) |xwh_n| \left\{ f_{\tilde{W}_{ij} | \tilde{X}_{ijk}}(0, x) + \frac{\partial f_{\tilde{W}_{ij} | \tilde{X}_{ijk}}(w, x)}{\partial w} \right\} \left( twh_n, x \right) \cdot wh_n \, dw \, dF_{\tilde{X}_{ijk}}(x) \]
\[ \leq M_2 M_K \mathbb{E} \left[ |\tilde{X}_{ijk} \tilde{W}_{ij}| = 0 \right] h_n + M_2 M_K \mathbb{E} [|\tilde{X}_{ijk}] h^2_n \]
\[ \leq \sqrt{2} M_2 M_K M \kappa_x (1 + C_0) h_n, \] (A4.47)

where the first inequality is by Assumption 4, the second equality is by definition, the third equality by Taylor’s expansion at $w = 0$ ($t \in [0, 1]$), the third inequality is by Assumptions 7 (Lemma A4.15) and 8 (Lemma A4.16), and the last inequality is by Assumption 11 (Lemma A4.17).

Combining (A4.43)-(A4.47), we have
\[ \mathbb{P}\{ \text{for any } k \in [p], \ |\nabla_k \tilde{L}_n(\tilde{\beta}, h_n)| \leq 2(A + A') \cdot \left( \log(np)/n \right)^{1/2} + 2\sqrt{2} M_2 M_K M \kappa_x (1 + C_0) \cdot h_n \} \]
\[ \geq 1 - 12.04 \exp(-c \log p), \]
for positive absolute constant $c$, and when we appropriately take $n$ bounded from below. Thus we have completed the proof by noting that $\lambda_n \geq 4(A + A') \cdot \left( \log(np)/n \right)^{1/2} + 4\sqrt{2} M_2 M_K M \kappa_x (1 + C_0) h_n$. \hfill \square

In the following, we collect the proofs of Lemmas A3.2-A3.3 in Section A3.

Proof of Lemma A3.2. By Taylor’s expansion, for some $t_{w,h} \in [0, 1]$, we have
\[ \mathbb{E} \left[ \frac{1}{h} K \left( \frac{W}{h} \right) Z \right] = \int \int K(w) f_{W | Z}(wh, z) \, dw \, dF_Z(z) \]
\[ = \int \int K(w) z \left\{ f_{W | Z}(0, z) + \frac{\partial f_{W | Z}(w, z)}{\partial w} \right\} \left| t_{w,h} wh \right| \, dw \, dF_Z(z), \]
which implies that
\[ \left| \mathbb{E} \left[ \frac{1}{h} K \left( \frac{W}{h} \right) Z \right] - \mathbb{E}[Z | W = 0] f_W(0) \right| \leq M_1 M_2 \mathbb{E} [|Z]| h. \]
This completes the proof. \hfill \square
Proof of Lemma A3.3. By Taylor’s expansion, for some \( t_{w,h} \in [0,1] \), we have

\[
\mathbb{E}\left[ \frac{1}{h} K\left( \frac{W_1 - W_2}{h} \right) \varphi(Z_1, Z_2) | W_2, Z_2 \right] \\
= \int \int \frac{1}{h} K\left( \frac{w - W_2}{h} \right) \varphi(z, Z_2) f_{W_1|Z_1}(w, z) \, dw \, dF_{Z_1}(z) \\
= \int K(w) \varphi(z, Z_2) f_{W_1|Z_1}(W_2 + wh, z) \, dw \, dF_{Z_1}(z) \\
= \int \int K(w) \varphi(z, Z_2) \left\{ f_{W_1|Z_1}(W_2, z) + \frac{\partial f_{W_1|Z_1}(w, z)}{\partial w} \right\}_{w + t_{w,h}wh} \, dw \, dF_{Z_1}(z),
\]

which implies that

\[
\left| \mathbb{E}\left[ \frac{1}{h} K\left( \frac{W_1 - W_2}{h} \right) \varphi(Z_1, Z_2) | W_2, Z_2 \right] - \mathbb{E}\left[ \varphi(Z_1, Z_2) | W_2, Z_2, W_1 = W_2 \right] f_{W_1}(W_2) \right| \\
\leq M_1 M_2 \mathbb{E}[|\varphi(Z_1, Z_2)||Z_2] h.
\]

This completes the proof.

\[ \Box \]

**Lemma A4.21.** Let \( D_i = (X_i, V_i, W_i) \) be i.i.d. for \( i = 1, \ldots, n \), and \( K(\cdot) \) be a positive kernel function, such that \( \int_{-\infty}^{\infty} K(w) \, dw = 1 \) and that max \( \{ \int_{-\infty}^{\infty} |w| K(w) \, dw, \sup_{w \in \mathbb{R}} K(w) \} \leq M_K \), for positive absolute constant \( M_K \). Assume that conditional on \( W_i = w \) for any \( w \) in the range of \( W_i \), and unconditionally, \( X_i \) and \( V_i \) are subgaussian with parameters at most \( \kappa_x^2 \) and \( \kappa_v^2 \) respectively, for positive absolute constants \( \kappa_x \) and \( \kappa_v \). Assume that there exists positive absolute constant \( M \), such that

\[
\max \left\{ \left| \frac{\partial f_{W_i|X,V}(w, x, v)}{\partial w} \right|, | f_{W_i|X,V}(w, x, v) \right\} \leq M,
\]

for any \( w, x, v \in \mathbb{R} \) such that the densities are defined. Take \( h_n \geq K_1 \{ \log(np)/n \}^{1/2} \) for positive absolute constant \( K_1 \), and assume that \( h_n \leq C_0 \) for positive constant \( C_0 \). Suppose \( n > \max \{ 64(c+2)^2(c+1)\{ \log(np) \}^3/3, 3 \} \) for positive absolute constant \( c \). Consider U-statistic

\[
U = \sum_{i<j} \left\{ \frac{1}{h_n} K\left( \frac{W_i - W_j}{h_n} \right) (X_i - X_j)(V_i - V_j) \right\}.
\]

Then we have

\[
\mathbb{P}\left\{ \left( \frac{n}{2} \right)^{-1} |U - \mathbb{E}[U]| \geq C\left( \frac{\log(np)}{n} \right)^{1/2} \right\} \leq 6.77 \exp\{ -(c + 1) \log p \},
\]

where

\[
C = \{ 16\sqrt{3}(1 + c)^{1/2} M_f + 4\sqrt{3}C_1(1 + c)^{1/2} M_f^{1/2} K_1^{-1/2} + 8C_2(1 + c) + 8C_3(1 + c)^{3/2} M_f^{1/2} M_f^{1/2} K_1^{-1/2} + 8C_4(1 + c)^2 M_f K_1^{-1} + 8M_f(c + 2) \} \kappa_x \kappa_v,
\]

(A4.48)

with \( C_1, \ldots, C_4 \) as defined in (A3.2) and \( M_f = M + M M_K C_0 \).

**Proof of Lemma A4.21.** Denote \( Z_{ij} = (X_i - X_j)(V_i - V_j) \). We apply truncation to \((X_i - X_j)^2 \) at level \( C_x^2 \log(np) \), and to \((V_i - V_j)^2 \) at level \( C_v^2 \log(np) \), for some positive absolute constants \( C_x \) and \( C_v \). Denote \( \mathcal{A}_{[n]} = \{ (X_i - X_j)^2 \leq C_x^2 \log(np), (V_i - V_j)^2 \leq C_v \log(np), \ i, j \in [n], \ i < j \} \), and first

43
focus on U-statistic
\[ \bar{U} = \sum_{i<j} \left[ \frac{1}{h_n} K \left( \frac{W_i - W_j}{h_n} \right) Z_{ij} \mathbb{I}\{(X_i - X_j)^2 \leq C_v \log(np), (V_i - V_j)^2 \leq C_v \log(np)\} \right]. \]

Denote
\[ g(D_i, D_j) = \frac{1}{h_n} K \left( \frac{W_i - W_j}{h_n} \right) Z_{ij} \mathbb{I}\{(X_i - X_j)^2 \leq C_v \log(np), (V_i - V_j)^2 \leq C_v \log(np)\}, \]

and
\[ f(D_i) = \mathbb{E}[g(D_i, D_j)|D_i]. \]

Assume \( h_n \geq K_1 \{\log(np)/n\}^{1/2} \) for some positive absolute constant \( K_1 \). Denote \( \bar{X} = X_1 - X_2, \bar{V} = V_1 - V_2, \) and \( \bar{W} = W_1 - W_2 \). Note that by argument of Lemma A4.16, we have all the necessary smooth conditions of densities. Denote \( C = C_x \cdot C_v \) and note that \((X_i - X_j)^2 \leq C_v \log(n)\), \((V_i - V_j)^2 \leq C_v \log(np)\) implies that \(|Z_{ij}| \leq C \log(n)\).

**Step I.** We bound \( B_g, B_f, \mathbb{E}[f(D)^2], \sigma^2, \) and \( B^2 \) as in Lemma A3.4, and apply Lemma A3.4.

We have \( B_g \leq CM_K \log(np)/h_n \leq (CM_K/K_1) \cdot \{\log(np)\}^{1/2} \). For \( B_f \), apply Lemma A3.3 on \( \varphi = 1 \) and with \( M_1 = M, M_2 = M_K \), and we have
\[ B_f \leq C \log(np) \cdot \mathbb{E}\left[ \frac{1}{h_n} K \left( \frac{W_i - W_j}{h_n} \right) \right] \leq C \log(np) \cdot \{f_W(W_j) + MM_K C_0\} \leq CM_f \log(np), \]

where \( M_f = M + M_K MM_K C_0 \), and the last inequality used the fact that \( f_W(W_j) \in [0, M] \).

For bounding \( \mathbb{E}[f(D)^2] \), apply Lemma A3.3 on \( \varphi = Z_{ij} \mathbb{I}\{ (X_i - X_j)^2 \leq C_v \log(np), (V_i - V_j)^2 \leq C_v \log(np)\} \) and with \( M_1 = M, M_2 = M_K \), and then we have
\[ |f(D_2) - f_1(D_2)| \leq M_K M f_2(D_2) h_n, \]

where
\[ f_1(D_2) \leq \mathbb{E}[Z_{12} \mathbb{I}\{|Z_{12}| \leq C \log(np)\}]W_1 = W_2, D_2 f_W(W_2) \]
\[ f_2(D_2) \leq \mathbb{E}\left[Z_{12} \mathbb{I}\{|Z_{12}| \leq C \log(np)\}\right]D_2]. \]

Therefore, we have
\[ \mathbb{E}[f(D_2)^2] = \mathbb{E}\left\{ (f(D_2) - f_1(D_2) + f_1(D_2))^2 \right\} \leq 2M_K^2 M^2 C_0^2 \mathbb{E}[f_2(D_2)^2] + 2 \mathbb{E}\left[f_1(D_2)^2\right], \quad (A4.49) \]

and meanwhile,
\[ \mathbb{E}[f_1(D_2)^2] \leq M^2 \mathbb{E}[Z_{12}^2] \leq M^2 \mathbb{E}[\bar{X}^4]^{1/2} \mathbb{E}[\bar{V}^4]^{1/2} \leq 12M^2 \kappa_x^2 \kappa_v^2, \] and
\[ \mathbb{E}[f_2(D_2)^2] \leq \mathbb{E}[Z_{12}^2] \leq \mathbb{E}[\bar{X}^4]^{1/2} \mathbb{E}[\bar{V}^4]^{1/2} \leq 12 \kappa_x^2 \kappa_v^2, \quad (A4.50) \]

where the first inequalities are by Jensen’s inequality, the second are by Cauchy-Schwarz inequality, and the third are due to the fact that \( \mathbb{E}[\bar{X}^4] \leq 12 \kappa_x^2, \mathbb{E}[\bar{V}^4] \leq 12 \kappa_v^2 \) (Lemma A4.18). Combining (A4.49) and (A4.50), we have
\[ \mathbb{E}[f(D_2)^2] \leq (M^2 M^2 C_0^2 + M^2) \cdot 24 \kappa_x^2 \kappa_v^2 < 24M^2 \kappa_x^2 \kappa_v^2. \quad (A4.51) \]

For bounding \( \sigma^2 \), apply Lemma A3.2 on \( Z = Z_{ij} \) and with \( M_1 = M, M_2 = M_K \), and then we
where the last inequality is by applying Lemma A3.3 with \( M \) and \( \tilde{C} \) of \( \tilde{X} \) and \( \tilde{V} \), both conditional on \( \tilde{W} = 0 \) and unconditionally.

For bounding \( B^2 \), we have

\[
B^2 = n \sup_{D_2} \mathbb{E} [g(D_1, D_2)^2 | D_2] 
\leq \frac{nM_K}{h_n} \sup_{D_2} \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_1 - W_2}{h_n} \right) Z_{12}^2 \mathbb{1} \{ |Z_{12}| \leq C \log(np) \} | D_2 \right] 
\leq \frac{nM_K}{h_n} (C \log(np))^2 \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_1 - W_2}{h_n} \right) | D_2 \right] 
\leq \frac{C^2 M_f M_K}{K_1} n \log(np)^{3/2},
\]

where the last inequality is by applying Lemma A3.3 with \( M_1 = M \) and \( M_2 = M_K \), and noticing that \( f_W(W_2) \in [0, M] \).

We take

\[
C_x = C_Z \cdot 2 \kappa_x^2, \quad C_v = C_Z \cdot 2 \kappa_v^2, \quad \text{for} \quad C_Z \geq 4,
\]

\[
t = C_t \cdot 16 \sqrt{3} M_f \kappa_x \kappa_v \left( \frac{n}{2} \right) \{ \log(np) / n \}^{1/2},
\]

\[
u = C_u \log p, \quad \text{for} \quad C_u > 1,
\]

and require \( n > \max \{ 16 C_Z^2 C_t^2 \{ \log(np) \}^3 / 3, \ 3 \} \). Then by Lemma A3.4, we have

\[
P \left\{ \left( \frac{n}{2} \right)^{-1} | \tilde{U} - \mathbb{E} [\tilde{U}] | \geq A_1 \left\{ \log(np) / n \right\}^{1/2} \right\} \leq 2 \exp(-C_t^2 \log(np)) + C_5 \exp(-C_u \log p),
\]

where

\[
A_1 = (16 \sqrt{3} C_t M_f + 4 \sqrt{3} C_1 C_u^2 M_f^2 K_1^{-1} ) K_1^{-1/2} + 8 C_2 C_u + 8 C_3 C_u^2 / 2 M_f^2 K_1^{-1} + 8 C_4 C_u^2 M_K K_1^{-1}) \kappa_x \kappa_v.
\]

Here, \( C_1, \ldots, C_5 \) are as defined in (A3.2).
Step II. We bound $|\mathbb{E}[\tilde{U}] - \mathbb{E}[\bar{U}]|$, and complete the proof. We have
\[
\left(\begin{array}{c}n \\ 2\end{array}\right)^{-1} |\mathbb{E}[\tilde{U}] - \mathbb{E}[U]| = |\mathbb{E}\left[\frac{1}{h_n}K\left(\frac{\bar{W}_{ij}}{h_n}\right)Z_{ij} \mathbb{I}\{|Z_{ij}| > C \log(np)\}\right]| \\
\leq \mathbb{E}\left[\frac{1}{h_n}K\left(\frac{\bar{W}}{h_n}\right)\tilde{X}^2 \mathbb{I}\{\tilde{X}^2 > 2C\tilde{Z}\kappa_2 \log(np)\}\right]^{1/2} \times \\
\mathbb{E}\left[\frac{1}{h_n}K\left(\frac{\bar{W}}{h_n}\right)\bar{V}^2 \mathbb{I}\{\bar{V}^2 > 2C\bar{Z}\kappa_2 \log(np)\}\right]^{1/2},
\]
where
\[
\mathbb{E}\left[\frac{1}{h_n}K\left(\frac{\bar{W}}{h_n}\right)\tilde{X}^2 \mathbb{I}\{\tilde{X}^2 > 2C\tilde{Z}\kappa_2 \log(np)\}\right] \\
\leq \mathbb{E}[\tilde{X}^2 \mathbb{P}\{\tilde{X}^2 > 2C\tilde{Z}\kappa_2 \log(np)\}] + M_1M_KC_0 \mathbb{E}[\tilde{X}^2 \mathbb{P}\{\tilde{X}^2 > 2C\tilde{Z}\kappa_2 \log(np)\}] \\
\leq M_f \{2C\tilde{Z}\kappa_2 \log(np) + 8\kappa_2^2\} \exp\{-2C\log(np)/(8\kappa_2^2)\} \leq 4M_f C\tilde{Z}\kappa_2 \{\log(np)/np\}^{1/2},
\]
where the first inequality is by applying Lemma A3.2 on $Z = \tilde{X}^2 \mathbb{P}\{\tilde{X}^2 > 2C\tilde{Z}\kappa_2 \log(np)\}$ and with $M_1 = M$, $M_2 = M_K$, and the second is by the fact that $X_{ij}$ is subgaussian with parameter at most $\kappa_2^2$ (Lemma A4.18), both conditional on $W_i = W_j$ and unconditionally, and by applying Lemma A4.19 with $a = 2C\tilde{Z}\kappa_2 \log(np) \geq 4\kappa_2$. By an identical argument, we have
\[
\mathbb{E}\left[\frac{1}{h_n}K\left(\frac{\bar{W}}{h_n}\right)\bar{V}^2 \mathbb{I}\{\bar{V}^2 > 2C\bar{Z}\kappa_2 \log(np)\}\right] \leq 4M_f C\bar{Z}\kappa_2 \{\log(np)/np\}^{1/2}.
\]
Combining the last three displays, we have
\[
\left(\begin{array}{c}n \\ 2\end{array}\right)^{-1} |\mathbb{E}[\tilde{U}] - \mathbb{E}[U]| = |\mathbb{E}\left[\frac{1}{h_n}K\left(\frac{\bar{W}_{ij}}{h_n}\right)Z_{ij} \mathbb{I}\{|Z_{ij}| > C \log(np)\}\right]| \leq A_2 \{\log(np)/np\}^{1/2}, \quad (A4.52)
\]
where $A_2 = 4M_f C\tilde{Z}\kappa_2 \kappa_v$.

We have
\[
P\left\{\left(\begin{array}{c}n \\ 2\end{array}\right)^{-1} |U - \mathbb{E}[U]| \geq (A_1 + A_2) \cdot \left\{\frac{\log(np)}{n}\right\}^{1/2}\right\} \\
\leq P\left\{\left(\begin{array}{c}n \\ 2\end{array}\right)^{-1} |U - \mathbb{E}[U]| \geq (A_1 + A_2) \cdot \left\{\frac{\log(np)}{n}\right\}^{1/2} \cap \mathcal{A}_i\right\} + P(\mathcal{A}_i) \\
\leq P\left\{\left(\begin{array}{c}n \\ 2\end{array}\right)^{-1} |\tilde{U} - \mathbb{E}[U]| \geq (A_1 + A_2) \cdot \left\{\frac{\log(np)}{n}\right\}^{1/2} \cap \mathcal{A}_i\right\} + P(\mathcal{A}_i) \\
\leq P\left\{\left(\begin{array}{c}n \\ 2\end{array}\right)^{-1} |\tilde{U} - \mathbb{E}[\tilde{U}]| \geq (A_1 + A_2) \cdot \left\{\frac{\log(np)}{n}\right\}^{1/2} \cap \mathcal{A}_i\right\} + P(\mathcal{A}_i) \\
\leq 2 \exp(-C^2 \log(np)) + C_5 \exp(-C_a \log p) + 2n^2 \exp(-C \log(np)/2) \\
\leq 2 \exp(-C^2 \log(np)) + C_5 \exp(-C_a \log p) + 2 \exp(-C \log p/2),
\]
where (i) is by (A4.52). We take $C^2 = C_a = c > 1$, and $C \leq \max\{2c, 4\} \leq 2c + 2$, for positive absolute constant $c$. This completes the proof. \nobreak
Lemma A4.22. Let $D_i = (X_i, V_i, W_i)$ be i.i.d. for $i = 1, \ldots, n$, and $K(\cdot)$ be a positive kernel function, such that $\int_{-\infty}^{+\infty} K(w) \, dw = 1$, and that
\[
\max \left\{ \int_{-\infty}^{+\infty} |w|^{2\alpha+1} K(w) \, dw, \sup_{w} |w|^\alpha K(w) \right\} \leq M_K,
\]
for positive absolute constant $M_K \geq 1$. Let $V_i = v(W_i)$ for function $v(\cdot)$, such that
\[
|v(w_1) - v(w_2)| \leq M_v |w_1 - w_2|^\alpha + M_d \I \{ (w_1, w_2) \in A \},
\]
for positive absolute constant $M_v$, $M_d$, $0 < \alpha \leq 1$, and set $A$ such that $(w_1, w_2) \in A$ implies $(w_2, w_1) \in A$, and that
\[
\mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_i}{h_n} \right) \I \{ (W_i, W_j) \in A \} \right] \leq M_a h_n,
\]
for positive absolute constant $M_a$. Assume that conditional on $W_i = w$ for any $w$ in the range of $W_i$, and unconditionally, $X_i$ is subgaussian with parameter at most $\kappa_x^2$, for positive absolute constant $\kappa_x$. Assume that there exists positive absolute constant $M$ such that
\[
\max \left\{ |\partial f_{w|X}(w, x)|, f_{w|X}(w, x) \right\} \leq M,
\]
for any $w, x \in \mathbb{R}$ such that the densities are defined. Take $h_n \geq K_1 \{ \log(np)/n \}^{1/2}$ for positive absolute constant $K_1$, and assume that $h_n \leq C_0$ for positive constant $C_0$. Suppose $n > \max \left\{ (64(c+2)^2(c+1)\tau_2^2\tau_3^{-1}(\log(np))^4, (\log(np))^{5/3} \right\}$, for positive absolute constant $c$. Consider U-statistics
\[
U = \sum_{i<j} \left\{ \frac{1}{h_n} K \left( \frac{W_i - W_j}{h_n} \right) (X_i - X_j)(V_i - V_j) \right\}.
\]
Then there exists positive absolute constants $C$, such that
\[
\mathbb{P} \left\{ \left( \binom{n}{2} \right)^{-1} |U - \mathbb{E}[U]| \geq C \{ \log(np)/n \}^{1/2} \right\} \leq 5.27 \exp\{ -(c + 1) \log p \},
\]
where
\[
C = 4\tau_1 1/2(1+c)^{1/2} + 2C_1 \tau_2^{1/2}(1+c)^{1/2} + 2C_2 \tau_2(1+c) + 2C_3 \tau_5^{1/2}(1+c)^{3/2} + 2C_4 \tau_1(1+c)^2 + 4(M + MM_K C_0) \cdot (M_0 C_0 + M_d) \cdot (c+2)\kappa_x \quad (A4.53)
\]
Here $C_1, \ldots, C_5$ are as defined in (A3.2), and $\tau_1 = \sqrt{2}(2+c)^{1/2}\kappa_x K_1^{-1}(M_0 M_K C_0 + M_d M_K)$,
$\tau_2 = \sqrt{2}(2+c)^{1/2}\kappa_x \{ M_0 M_K M(1+C_0) C_0 + M_d(M + MM_K C_0) \}$,
$\tau_3 = 4M_0^2 C_0^2 + M_d^2 \cdot (1 + C_0^2) \cdot (c+2)\kappa_x^2$,
$\tau_4 = \{ 4M_0^2 M_K \kappa_x^2 (1+C_0) C_0^{2a-\gamma} + 2M_d^2 \cdot (12M_K^4 + 12M_0 M_K C_0 \kappa_x^4)^{1/2} \cdot M_{a,1}^{1/2} C_0^{-1/2 - \gamma_1} \} \cdot M_K K_1^{-\gamma_1}$,
$\tau_5 = 4(2+c)\kappa_x^2 \{ M_0 M_K (1+C_0) C_0^{2a} + M_d^2 (M + MM_K C_0) \} M_K K_1^{-1}$,
where $\gamma_1 = \min \{ 2a - 1, -1/2 \}$.

Proof of Lemma A4.22. We apply truncation to $(X_i - X_j)^2$ at level $C \log(np)$ for some positive

47
absolute constant $C$, and first focus on U-statistic
\[
\bar{U} = \sum_{i<j} \left[ \frac{1}{h_n} K \left( \frac{W_i - W_j}{h_n} \right) (X_i - X_j)(V_i - V_j) \mathbb{I} \{ (X_i - X_j)^2 \leq C \log (np) \} \right].
\]
Denote
\[
g(D_i, D_j) = \frac{1}{h_n} K \left( \frac{W_i - W_j}{h_n} \right) (X_i - X_j)(V_i - V_j) \mathbb{I} \{ (X_i - X_j)^2 \leq C \log (np) \},
\]
and
\[
f(D_i) = \mathbb{E}[g(D_i, D_j) | D_i].
\]
Assume $h_n \geq K_1 \{ \log (np) / n \}^{1/2}$ for some positive absolute constant $K_1$. Note that by the argument of Lemma A4.16, we have all the necessary smooth conditions of densities.

**Step I.** We bound $B_g$, $B_f$, $\mathbb{E}[f(D_2)^2]$, $\sigma^2$, and $B^2$ as in Lemma A3.4, and apply Lemma A3.4. For bounding $B_g$, we have
\[
B_g \leq h_n^{-1}(C \log (np))^{1/2} \left\| K \left( \frac{W_i - W_j}{h_n} \right) [M_v(W_i - W_j)^\alpha + M_d \mathbb{I} \{ (W_i, W_j) \in A \}] \right\|_{\infty}
\leq C^{1/2} K_1^{-1} (M_v M_K C_0^\alpha + M_d M_K) \cdot n^{1/2} = \tau_1 n^{1/2},
\]
where the second inequality is by $|w|^\alpha K(w) \leq M_K$ and $K(w) \leq M_K$.

For bounding $B_f$, we have
\[
B_f \leq (C \log (np))^{1/2} \left\| \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_i - W_j}{h_n} \right) (V_i - V_j) \mathbb{I} \{ (W_i, W_j) \in A \} | D_j \right] \right\|_{\infty}
\leq (C \log (np))^{1/2} \left\{ M_v \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_i - W_j}{h_n} \right) |W_i - W_j|^\alpha | D_j \right] + M_d \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_i - W_j}{h_n} \right) \mathbb{I} \{ (W_i, W_j) \in A \} | D_j \right] \right\} \right\|_{\infty},
\]
where
\[
\mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_i - W_j}{h_n} \right) |W_i - W_j|^\alpha | D_j \right]
= \int \int K(w)|w h_n|^\alpha f_{W_1}(w_2 + w h_n) dw
= \int \int K(w)|w h_n|^\alpha \left\{ f_{W_1}(w_2) + \frac{\partial f_{W_1}(w)}{\partial w} \bigg|_{w_2 + w h_n} \cdot w h_n \right\} dw
\leq M M_K (1 + C_0) h_n^\alpha.
\]
Therefore $B_f \leq C^{1/2} \left\{ M_v M_K M (1 + C_0) C_0^\alpha + M_d (M + M M_K C_0) \right\} \cdot \{ \log (np) \}^{1/2} = \tau_2 \{ \log (np) \}^{1/2}$.

For bounding $\mathbb{E}[f(D_2)^2]$, we have
\[
|f(D_2)| \leq M_v \mathbb{E} \left[ \frac{|W_i - W_j|^\alpha K \left( \frac{W_i - W_j}{h_n} \right) |X_1 - X_2| | D_2 \right]
+ M_d \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_i - W_j}{h_n} \right) |X_1 - X_2| | D_2 \right] \right)
\]
(A4.54)
Apply Lemma A3.3 on $\varphi = |X_1 - X_2|$ and with $M_1 = M$, $M_2 = M_K$, we have

$$
\mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{W_1 - W_2}{h_n} \right) |X_1 - X_2| |D_2\right] \\
\leq \mathbb{E}[|X_1 - X_2||W_1 = W_2, D_2] f_{W_1}(W_2) + M M_K C_0 \mathbb{E}[|X_1 - X_2| |D_2],
$$

(4.55)

while using a similar argument as used in proof of Lemma A3.3, for some $t \in [0, 1]$, we have

$$
\mathbb{E}\left[ |W_1 - W_2|^{\alpha} \frac{h_n}{h_n} K\left( \frac{W_1 - W_2}{h_n} \right) |X_1 - X_2| |D_2\right] \\
= \int \int K(w)|x - x_2| \cdot |w h_n|^{\alpha} f_{W|X}(W_2 + w h_n, x) \, dw \, dF_X(x) \\
= \int \int K(w)|x - x_2| \cdot |w h_n|^{\alpha} \left\{ f_{W|X}(W_2, x) + \frac{\partial f_{W|X}(w, x)}{\partial w} \right\} \mid_{(W_2 + t w h_n, x)} \cdot wh_n \, dw \, dF_X(x) \\
\leq M_K M C_0^\alpha \cdot \mathbb{E}[|X_1 - X_2| |W_1 = W_2, D_2] + M_K M C_0^{\alpha + 1} \cdot \mathbb{E}[|X_1 - X_2| |D_2].
$$

(4.56)

Combining (4.54)-(4.56), and by Jensen’s inequality, we have

$$
\mathbb{E}[f(D_2)^2] \leq \mathbb{E}\left\{ \left( (M_e M_K M C_0^{\alpha} + M_d M) \mathbb{E}[|X_1 - X_2| |W_1 = W_2, D_2] \\
+ (M_e M_K M C_0^{\alpha + 1} + M_d M M M_K C_0) \mathbb{E}[|X_1 - X_2| |D_2] \right)^2 \right\} \\
\leq 2(M_e M_K M C_0^{\alpha} + M_d M)^2 \mathbb{E}[\mathbb{E}[|X_1 - X_2| |W_1 = W_2, D_2]^2] \\
+ 2(M_e M_K M C_0^{\alpha + 1} + M_d M M M_K C_0)^2 \mathbb{E}[\mathbb{E}[|X_1 - X_2| |D_2]^2] \\
\leq 2M_K^2 M^2 (M_e C_0^{\alpha} + M_d)^2 (1 + C_0^2) \mathbb{E}[|X_1 - X_2|^2] \\
\leq 4M_K^2 M^2 \cdot (M_e C_0^{\alpha} + M_d)^2 \cdot (1 + C_0^2) \cdot \kappa _x^2 = \tau_3
$$

For bounding $\sigma^2$, we have

$$
\sigma^2 = \mathbb{E}[g(D_1, D_2)^2] \\
\leq \mathbb{E}\left[ \frac{2M_e^2 |W_1 - W_2|^{\alpha} + 2M_d^2 \mathbb{I}\{(W_1, W_2) \in A\}}{h_n^2} K^2\left( \frac{W_1 - W_2}{h_n} \right) |X_1 - X_2|^2 \right] \\
\leq \frac{2M_e^2 M_K}{h_n} \mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{W_1 - W_2}{h_n} \right) |W_1 - W_2|^{\alpha} (X_1 - X_2)^2 \right] \\
+ \frac{2M_d^2 M_K}{h_n} \mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{W_1 - W_2}{h_n} \right) \mathbb{I}\{(W_1, W_2) \in A\} (X_1 - X_2)^2 \right].
$$

(4.57)

Using a similar argument as used in proof of Lemma A3.2, for some $t \in [0, 1]$, we have

$$
\mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{W_1 - W_2}{h_n} \right) |W_1 - W_2|^{\alpha} (X_1 - X_2)^2 \right] \\
= \int \int K(w)|wh_n|^{\alpha} \cdot x^2 \cdot f_{W|X}(w h_n, x) \, dw \, dF_X(x) \\
= \int \int K(w)|wh_n|^{\alpha} \cdot x^2 \cdot \left\{ f_{W|X}(0, x) + \frac{\partial f_{W|X}(w, x)}{\partial w} \right\} \mid_{(t w h_n, x)} \cdot wh_n \, dw \, dF_X(x) \\
\leq M M_K h_n^2 \mathbb{E}[\overline{X}^2 | \overline{W} = 0] + M M_K h_n^2 \mathbb{E}[\overline{X}^2] \\
\leq 2M M_K \kappa _x^2 (1 + C_0) h_n^2.
$$

(4.58)
We also have
\[
\mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{W_1 - W_2}{h_n} \right) \mathbb{I}\left\{ (W_1, W_2) \in A \right\} (X_1 - X_2)^2 \right] \\
\leq \mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{W_1 - W_2}{h_n} \right) \mathbb{I}\left\{ (W_1, W_2) \in A \right\} \right]^{1/2} \cdot \mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{W_1 - W_2}{h_n} \right) (X_1 - X_2)^2 \right]^{1/2} \\
\leq (M_a h_n)^{1/2} \cdot (\mathbb{E}[\tilde{X}^4|\tilde{W} = 0] M + MM_K C_0 \mathbb{E}[\tilde{X}^4])^{1/2} \\
\leq (12 M_K^4 + 12 M M_K C_0 \kappa^4) h_n^{1/2} \cdot M_a^{1/2} h_n^{1/2},
\]
where the first inequality is by Cauchy-Schwarz inequality, second is by applying Lemma A3.2 on $Z = (X_1 - X_2)^4$ with $M_1 = M$, $M_2 = M_K$, and third is by subgaussianity of $\tilde{X}$ conditional on $\tilde{W} = 0$ and unconditionally. Combining (A4.57)-(A4.59), we have
\[
\sigma^2 \leq \left\{ 4 M_a^2 M M_K \kappa^2 (1 + C_0) C_0^{2\alpha - \gamma_1} + 2 M_a^2 \cdot (12 M_K^4 + 12 M M_K C_0 \kappa^4)^{1/2} \cdot M_a^{1/2} C_0^{1/2 - \gamma_1} \right\} M_K h_n^{\gamma_1} + \tau_4 n^{-\gamma_1/2} \{ \log(np) \}^{\gamma_1/2},
\]
where $\gamma_1 = \min\{2\alpha - 1, -1/2\}$.

For bounding $B^2$, we have
\[
B^2 = n \sup_{D_2} \mathbb{E}\left[ g(D_1, D_2)^2 | D_2 \right] \\
\leq n \sup_{D_2} \mathbb{E}\left[ \frac{2 M_a^2 |W_1 - W_2|^{2\alpha} + 2 M_a^2 \mathbb{I}\left\{ (W_1, W_2) \in A \right\} K^2 \left( \frac{W_1 - W_2}{h_n} \right) \cdot (X_1 - X_2)^2 \mathbb{I}\left\{ (X_1 - X_2)^2 \leq C \log(np) \right\} | D_2 \right]\]
\[
\leq \frac{2 CM_n n \log(np)}{h_n} \left\{ M_a^2 \mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{W_1 - W_2}{h_n} \right) |W_1 - W_2|^{2\alpha} | D_2 \right] \\
+ M_a^2 M_K \mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{W_1 - W_2}{h_n} \right) \mathbb{I}\left\{ (W_1, W_2) \in A \right\} | D_2 \right]\right\}.
\]
By a similar argument as used in (A4.56), for some $t \in [0, 1]$, we have
\[
\mathbb{E}\left[ \frac{1}{h_n} K\left( \frac{W_1 - W_2}{h_n} \right) |W_1 - W_2|^{2\alpha} | D_2 \right] \\
= \int \int K(w)|wh_n|^{2\alpha} f_{W|X}(W_2 + wh_n, x) dw dF_X(x) \\
= \int \int K(w)|wh_n|^{2\alpha} \left\{ f_{W|X}(W_2, x) + \frac{\partial f_{W|X}(w, x)}{\partial w} \bigg|_{(W_2 + twh_n, x)} \cdot wh_n \right\} dw dF_X(x) \\
\leq MM_K (1 + C_0) h_n^{2\alpha}
\]
Combining (A4.60) and (A4.61), we have
\[
B^2 \leq \frac{2 CM_n n \log(np)}{h_n} \left\{ M_a^2 MM_K (1 + C_0) h_n^{2\alpha} + M_a^2 (M + MM_K C_0) \right\} \\
\leq 2 CM_K \left\{ M_a^2 MM_K (1 + C_0) h_n^{2\alpha} + M_a^2 (M + MM_K C_0) \right\} K_1^{-1} n^{3/2} \{ \log(np) \}^{1/2} \\
= \tau_5 n^{3/2} \{ \log(np) \}^{1/2}
\]
We take
\[ C = C_Z \cdot 2\kappa_x^2, \]
\[ t = C_t 4 \tau_3^{1/2} \left( \frac{n}{2} \right) \{ \log(np)/n \}^{1/2}, \]
\[ u = C_u \log p, \] for \( C_u > 1, \)
and require \( n > \max \{ 64c(c + 1)^2 \tau_3^2 \tau_3^{-1} \{ \log(np) \}^4, \{ \log(np) \}^{5/3} \}. \) For simplicity, we further take \( C_t^2 = C_u = c > 1, \) and \( C_Z = \max \{ 2c, 4 \} \leq 2c + 2. \) Then by Lemma A3.4, we have
\[ \mathbb{P}\left\{ \left( \frac{n}{2} \right)^{-1} | \tilde{U} - \mathbb{E}[\tilde{U}] | \geq A_1 \{ \log(np)/n \}^{1/2} \right\} \leq 2 \exp(-C_t^2 \log(np)) + C_5 \exp(-C_u \log p), \]
where \( A_1 = 2 \cdot (2 \tau_3^{1/2} c^{1/2} + C_1 \tau_4^{1/2} c^{1/2} + C_2 \tau_2 c + C_3 \tau_5^{1/2} c^{3/2} + C_4 \tau_1 c^2). \) Here, \( \tau_1, \ldots, \tau_5 \) are given in equations above, and \( C_1, \ldots, C_5 \) are as defined in (A3.2).

**Step II.** We bound \( | \mathbb{E}[\tilde{U}] - \mathbb{E}[U] | \), and complete the proof.

We have
\[ \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_1 - W_2}{h_n} \right) (X_1 - X_2)^2 I \left\{ (X_1 - X_2)^2 > C \log(np) \right\} \right] \]
\[ \leq \mathbb{E} \left[ (X_1 - X_2)^2 I \left\{ (X_1 - X_2)^2 > C \log(np) \right\} | W_1 = W_2 \right] M \]
\[ + MM_K C_0 \mathbb{E} \left[ (X_1 - X_2)^2 I \left\{ (X_1 - X_2)^2 > C \log(np) \right\} \right] \]
\[ \leq 4(M + MM_K C_0) C_Z \kappa_x^2 \cdot \{ \log(np)/n \}, \]
where the first inequality is by applying Lemma A3.2 on \( (X_1 - X_2)^2 I \left\{ (X_1 - X_2)^2 > C \log(np) \right\} \) and with \( M_1 = M, \) \( M_2 = M_K, \) the second is by subgaussianity of \( (X_1 - X_2) \) conditional on \( W_1 = W_2 \) and unconditionally, and by applying Lemma A4.19 with \( a = C \log(np) \geq 4\kappa_x^2. \)

Based on earlier arguments, we have
\[ \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_1 - W_2}{h_n} \right) (V_1 - V_2)^2 \right]^{1/2} \]
\[ \leq M_\nu \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_1 - W_2}{h_n} \right) | W_1 - W_2 |^{2\alpha} \right]^{1/2} + M_\nu \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_1 - W_2}{h_n} \right) \right]^{1/2} \]
\[ \leq (M + MM_K C_0)^{1/2} (M_\nu C_0^\alpha + M_\nu), \]
where the last inequality is by the fact that \( |w|^{\alpha} K(w) < M \) and by applying Lemma A3.2 on \( Z = 1 \) with \( M_1 = M, \) \( M_2 = M_K. \)

Combining the last two displays, and apply Cauchy-Schwarz inequality, we have
\[ n^{-1} \left| \mathbb{E}[\tilde{U}] - \mathbb{E}[U] \right| \]
\[ = \left| \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_1 - W_2}{h_n} \right) (X_1 - X_2)(V_1 - V_2) I \left\{ (X_1 - X_2)^2 > C \log(np) \right\} \right] \right| \]
\[ \leq \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_1 - W_2}{h_n} \right) (X_1 - X_2)^2 I \left\{ (X_1 - X_2)^2 > C \log(np) \right\} \right] \frac{1}{2} \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{W_1 - W_2}{h_n} \right) (V_1 - V_2)^2 \right]^{1/2} \]
\[ \leq A_2 \cdot \{ \log(np)/n \}^{1/2}, \]
where \( A_2 = 2(M + MM_K C_0) \cdot (M_\nu C_0^\alpha + M_\nu) C_Z \kappa_x. \) (A4.62)
Denote $\mathcal{A}_{[n]} = \{(X_i - X_j)^2 \leq C \log(np), \ i, j \in [n], \ i < j\}$, and we have
\[
\P\left\{ \left( \frac{n}{2} \right)^{-1} |U - \E[U]| \geq (A_1 + A_2) \cdot \left( \frac{\log(np)}{n} \right)^{1/2} \right\} \\
\leq \P\left\{ \left( \frac{n}{2} \right)^{-1} |\bar{U} - \E[U]| \geq (A_1 + A_2) \cdot \left( \frac{\log(np)}{n} \right)^{1/2} \cap \mathcal{A}_{[n]} \right\} + \P(\mathcal{A}_{[n]}^c) \\
\leq \P\left\{ \left( \frac{n}{2} \right)^{-1} |\bar{U} - \E[U]| \geq (A_1 + A_2) \cdot \left( \frac{\log(np)}{n} \right)^{1/2} \right\} + \P(\mathcal{A}_{[n]}^c) \\
\leq \P\left\{ \left( \frac{n}{2} \right)^{-1} |\bar{U} - \E[U]| \geq (A_1 + A_2) \cdot \left( \frac{\log(np)}{n} \right)^{1/2} \right\} + \P(\mathcal{A}_{[n]}^c) \\
\leq 2 \exp(-C_i^2 \log(np)) + C_5 \exp(-C_u \log p) + \frac{n^2}{2} \exp\{-2C_Z \log(np)/2\} \\
\leq 2 \exp(-C_i^2 \log(np)) + C_5 \exp(-C_u \log p) + \frac{1}{2} \exp\{-C_Z \log p/2\},
\]
where $(i)$ is by (A4.62). This completes the proof. \hfill \Box

References


