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 $\mathbb{1}(\cdot)$ stand for the indicator function. For arbitrary vectors v, $v' \in \mathbb{R}^p$ and $0 < q < \infty$, we define $\|v\|_0 = \sum_{j=1}^p \mathbb{1}(v_j \neq 0)$, $\|v\|_q^q = \sum_{j=1}^p |v_j|^q$, and $\langle v, v' \rangle = \sum_{j=1}^p v_j v_j'$. For an arbitrary matrix $\Omega = (\Omega_{ij}) \in \mathbb{R}^{p \times q}$, write $\|\Omega\|_\infty = \max_{i \in [p]} \sum_{j=1}^q |\Omega_{ij}|$. For a symmetric real matrix Ω, let $\lambda_{\min}(\Omega)$ denote its smallest eigenvalue. For a set $\mathcal{S}$, we denote $|\mathcal{S}|$ to be its cardinality and $\mathcal{S}^{\mathsf{c}}$ 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 $\|f\|_\infty = \sup_{f \in \mathcal{X}} f(x)$. For an arbitrary function $f : \mathbb{R}^k \to \mathbb{R}$, we use $\nabla f = (\nabla_1 f, \ldots, \nabla_k f)^{\mathsf{T}}$ 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 Σ_X denote its covariance matrix. For some joint continuous random vector $(X^{\mathsf{T}}, W)^{\mathsf{T}} \in \mathbb{R}^{p+1}$ 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: `fanghan@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, α)-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 $|f_W(w)| \le M$ for some positive absolute constant M. Consider set

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

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

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

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

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{h} K\Big(\frac{\widetilde{W}_{ij}}{h}\Big) \, \mathbb{1}\big\{(W_i, W_j) \in A\big\}\Big] \\
=&\frac{2}{h} \sum_{j=1}^{J} \int_{-\infty}^{a_j} \int_{a_j}^{+\infty} \mathbb{1}\big\{|w_1 - w_2| \le h/2\big\} f_W(w_1) f_W(w_2) \, dw_1 \, dw_2 \le \frac{JM^2}{4} h.
\end{aligned}$$

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 $K(w) = \mathbb{1}(w \in [-1/2, 1/2])$, and

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

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

$$\mathbb{E}\Big[\frac{1}{h} K\Big(\frac{\widetilde{W}_{ij}}{h}\Big) \, \mathbb{1}\big\{(W_i, W_j) \in A\big\}\Big] = h/4.$$

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

$$\begin{aligned}
&h_n \in [K_1 (\log p/n)^{1/2}, C_0) \quad \text{and} \\
&n \ge C\big\{(\log p)^{5/(3-4\xi)} \vee (\log p)^3 \vee q^{4/3} (\log p)^{1/3} \vee q (\log p)^2\big\},
\end{aligned}$$

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 $\eta_n = \big|\|\mathbb{E}\big[\widetilde{X}\widetilde{X}^{\intercal} \big| \widetilde{W} = 0\big]\|_{\infty}$. 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

$$\mathbb{P}(\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \leq C' s\lambda_n^2) \geq 1 - c\exp(-c'\log p) - c\exp(-c'\log n) - \epsilon_n.$$

(2) Analogue of Theorem 3.2: Assume Assumption 14 holds with a general $\gamma \in (0, 1]$. Set $q = s$. Assume further that $\lambda_n \geq C\{(\log p/n)^{1/2} + h_n^\gamma\}$, 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

$$\mathbb{P}(\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \leq C' s\lambda_n^2) \geq 1 - c\exp(-c'\log p) - c\exp(-c'\log n) - \epsilon_n.$$

(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 $\lambda_n \geq C\{h_n + \eta_n(\log p/n)^{1/2}\}$, 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 6-8, 10-11, 14-16, and 17, we have

$$\mathbb{P}\Big\{\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \leq C'\Big(s\lambda_n^2 + \frac{s\log p}{n} + \frac{n\lambda_n^2 h_n^{2\gamma}}{\log p}\Big)\Big\} \geq 1 - c\exp(-c'\log p) - c\exp(-c'\log n) - \epsilon_n.$$

(4) Analogue of Theorem 2.3:

a. Assume that $g(\cdot)$ is α-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 $M, M_K, C_0, \kappa_x, M_u, \kappa_\ell, M_\ell, \xi, \epsilon, K_1$, and Hölder parameters of $g(\cdot)$. Then under Assumptions 6-8, 9′, 10-11, 13, and 17, we have

$$\mathbb{P}(\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \leq C' s\lambda_n^2) \geq 1 - c\exp(-c'\log p) - c\exp(-c'\log n).$$

b. Assumption 5 holds with $\alpha \in (0, 1]$. Set $q = s$. Assume further that $\lambda_n \geq C\{(\log p/n)^{1/2} + h_n^\gamma\}$ 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 $M_d M_a = 0$, $\gamma = \alpha \wedge 1/2$ if otherwise. Then under Assumptions 6-8, 9′, 10-11, 13 and 17, we have

$$\mathbb{P}(\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \leq C' s\lambda_n^2) \geq 1 - c\exp(-c'\log p) - c\exp(-c'\log n).$$

c. Assume Assumption 5 holds with $\alpha \in [1/4, 1]$. Set $q = s + nh_n^{2\gamma}/\log p$. Assume further that $\lambda_n \geq C\{h_n + \eta_n(\log p/n)^{1/2}\}$ 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 $M_d M_a = 0$, $\gamma = \alpha \wedge 1/2$ if otherwise. Then under Assumptions 6-8, 9′, 10-11, 13 and 17,

$$\mathbb{P}\Big\{\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \leq C'\Big(s\lambda_n^2 + \frac{s\log p}{n} + \frac{n\lambda_n^2 h_n^{2\gamma}}{\log p}\Big)\Big\} \geq 1 - c\exp(-c'\log p) - c\exp(-c'\log n).$$

(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

$$\mathbb{P}(\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \le C' s \lambda_n^2) \ge 1 - c\exp(-c'\log p) - c\exp(-c'\log n).$$

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 by p real matrix M and an integer $q \in [p]$, the q-restricted spectral norm $\|M\|_{2,q}$ of M is defined to be

$$\|M\|_{2,q} := \max_{v \in \mathbb{R}^p, \|v\|_0 \le q} \Big| \frac{v^\mathsf{T} M v}{v^\mathsf{T} v} \Big|.$$

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 $\widehat{\Gamma}_n(\theta, h_n) = \widehat{L}_n(\beta, h_n)$, simple algebra yields

$$\delta \widehat{L}_n(\Delta, h_n) = \Delta^\mathsf{T} \Big\{ \binom{n}{2}^{-1} \sum_{i<j} \frac{1}{h_n} K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big) \widetilde{X}_{ij} \widetilde{X}_{ij}^\mathsf{T} \Big\} \Delta = \Delta^\mathsf{T} \widehat{T}_n \Delta.$$

Note that $\widehat{T}_n$ 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 $\widetilde{W}_{ij}$ is close to zero. Consequently, when h_n is set in the regime of interest, the variance of the kernel $g_\Delta(D_i, D_j) = h_n^{-1} K(\widetilde{W}_{ij}/h_n)(\widetilde{X}_{ij}^\mathsf{T}\Delta)^2$ 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 $\widehat{T}_n$ is need.

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

Theorem A3.1. For some $q \in [p]$, suppose there exists some absolute constant $C > 0$ such that

$$n \ge C \cdot \big[\big\{q^{4/3}(\log p)^{1/3} \vee q(\log p)^2\big\} + \log(1/\alpha)\big].$$

Then under Assumptions 7, 8, and 11, with probability at least $1 - \alpha$,

$$\|\widehat{T}_n - \mathbb{E}\widehat{T}_n\|_{2,q} \le C' \cdot \Big[\frac{q(\log p)^{1/4}}{n^{3/4}} + \frac{q(\log p)^2}{n} + \frac{\log(1/\alpha)}{n}\Big]^{1/2},$$

where C' is a positive constant only depending on M, M_K, C_0, κ_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 $\delta\widehat{L}_n(\Delta, h_n)$, after decomposition, does not have a natural product structure, namely, it cannot be written as $n^{-1}\sum_{i=1}^n U_i U_i^\top$ 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

$$\Big|\frac{\partial f_{W|Z}(w, z)}{\partial w}\Big| \le M_1,$$

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 $\int_{-\infty}^{+\infty} |w|K(w)\,dw \le M_2$ for some constant $M_2 > 0$. Then we have for any $h > 0$,

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

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

$$\Big|\frac{\partial f_{W_1|Z_1}(w, z)}{\partial w}\Big| \le M_1$$

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 $\int_{-\infty}^{+\infty} |w|K(w)\,dw \le M_2$ for some constant $M_2 > 0$. Let $\varphi : \mathcal{Z}^2 \to \mathbb{R}$ be a measurable function. Then we have for any $h > 0$,

$$\begin{aligned}&\Big|\mathbb{E}\Big[\frac{\varphi(Z_1,Z_2)}{h}K\Big(\frac{W_1-W_2}{h}\Big)\Big|W_2, Z_2\Big] - \mathbb{E}\big[\varphi(Z_1,Z_2)\big|W_1 = W_2, W_2, Z_2\big]f_{W_1}(W_2)\Big| \\ &\le M_1 M_2 \mathbb{E}\big[|\varphi(Z_1, Z_2)|\big|Z_2\big]h.\end{aligned}$$

On the other hand, the second order term of $\delta\widehat{L}_n(\Delta, h_n)$, 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 Gine (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 $\mathbb{E}\big[g(Z_1, Z_2)\big] < \infty$. Write $U_n(g) = \sum_{i<j} g(Z_i, Z_j)$ and $f(z) = \mathbb{E}\big[g(Z, z)\big]$. Let

$$B_g = \|g\|_\infty, B_f = \sup_{Z_2} \mathbb{E}\big[|g(Z_1, Z_2)|\big|Z_2\big], \text{ and } \sigma^2 = \mathbb{E}\big[g(Z_1, Z_2)^2\big].$$

In addition, denote $B^2 = n\sup_{Z_2}\mathbb{E}\big[g(Z_1,Z_2)^2\big|Z_2\big]$. We then have

$$\begin{aligned}&\mathbb{P}\big(|U_n(g)-\mathbb{E}[U_n(g)]|\geq t+C_1 n\sigma u^{1/2}+C_2B_f u+C_3Bu^{3/2}+C_4B_g u^2\big)\\ \leq&2\exp\Big(\frac{-t^2/n^2}{8n\mathbb{E}\big[f(Z_2)^2\big]+4B_f\cdot t/n}\Big)+C_5e^{-u},\end{aligned}\tag{A3.1}$$

where we take positive absolute constants

$$\begin{aligned}&C_1=2(1+\epsilon)^{3/2},\\ &C_2=8\sqrt{2}(2+\epsilon+\epsilon^{-1}),\\ &C_3=e(1+\epsilon^{-1})^2(5/2+32\epsilon^{-1})+\big[\{2\sqrt{2}(2+\epsilon+\epsilon^{-1})\}\vee(1+\epsilon)^2/\sqrt{2}\big],\\ &C_4=\big\{4e(1+\epsilon^{-1})^2(5/2+32\epsilon^{-1})\big\}\vee 4(1+\epsilon)^2/3,\\ &C_5=2.77,\end{aligned}\tag{A3.2}$$

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

$$n\geq C\big\{s^{4/3}(\log p)^{1/3}\vee s(\log p)^2\big\},$$

for some constant $C>0$ only depending on $M, M_K, C_0, \kappa_x, \kappa_\ell, M_\ell$. Then we have

$$\begin{aligned}&\mathbb{P}\Big[\delta\widehat{L}_n(\Delta,h_n)\geq\frac{\kappa_\ell M_\ell}{4}\|\Delta\|_2^2\text{ for all }\Delta\in\big\{\Delta'\in\mathbb{R}^p:\|\Delta_{\mathcal{S}^{\mathsf{c}}}\|_1\leq 3\|\Delta_{\mathcal{S}}\|_1\big\}\Big]\\ \geq&1-c\exp(-c'\log p)-c\exp(-c'n),\end{aligned}$$

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

$$n\geq C\big[\{s+nh_n^{2\gamma}/\log p\}^{4/3}(\log p)^{1/3}\vee\{s+nh_n^{2\gamma}/\log p\}(\log p)^2\big],$$

for some constant $C>0$ only depending on $M, M_K, C_0, \kappa_x, \kappa_\ell, M_\ell, \zeta, \gamma$. Then we have

$$\begin{aligned}&\mathbb{P}\Big\{\delta\widehat{L}_n(\Delta,h_n)\geq\frac{\kappa_\ell M_\ell}{4}\|\Delta\|_2^2\text{ for all }\Delta\in\mathcal{C}_{\widetilde{\mathcal{S}}'_n}\Big\}\\ \geq&1-c\exp(-c'\log p)-c\exp(-c'n),\end{aligned}$$

where $\mathcal{C}_{\widetilde{\mathcal{S}}'_n}:=\big\{v\in\mathbb{R}^p:\|v_{\mathcal{J}^{\mathsf{c}}}\|_1\leq 3\|v_{\mathcal{J}}\|_1\text{ for some }\mathcal{J}\subset[p]\text{ and }|\mathcal{J}|\leq s+\zeta^2nh_n^{2\gamma}/\log p)\big\}$, 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

$$\|\widetilde{\theta}^*_{h_n} - \theta^*\|_2^2 \le \rho_n^2.$$

So it suffices to show that

$$\|\widehat{\theta}_{h_n} - \widetilde{\theta}^*_{h_n}\|_2^2 \le 9\widetilde{s}_n\lambda_n^2/\kappa_1^2$$

holds with probability at least $1 - \epsilon_{1,n} - \epsilon_{2,n}$ whenever $\lambda_n \le \kappa_1 r/3\widetilde{s}_n^{1/2}$. We split the rest of the proof into two main steps.

Step I. Denote $\widehat{\Delta} = \widehat{\theta}_{h_n} - \widetilde{\theta}^*_{h_n}$. Recall definition of sets $\widetilde{\mathcal{S}}_n$ and $\mathcal{C}_{\widetilde{\mathcal{S}}_n}$, and further define function

$$\mathcal{F}(\Delta) = \widehat{\Gamma}_n(\widetilde{\theta}^*_{h_n} + \Delta, h_n) - \widehat{\Gamma}_n(\widetilde{\theta}^*_{h_n}, h_n) + \lambda_n\big(\|\widetilde{\theta}^*_{h_n} + \Delta\|_1 - \|\widetilde{\theta}^*_{h_n}\|_1\big).$$

For the first step, we show that if $\mathcal{F}(\Delta) > 0$ for all $\Delta \in \mathcal{C}_{\widetilde{\mathcal{S}}_n} \cap \big\{\Delta' \in \mathbb{R}^p : \|\Delta'\|_2 = \eta\big\}$, then $\|\widehat{\Delta}\|_2 \le \eta$. To this end, we first show that

$$\widehat{\Delta} \in \mathcal{C}_{\widetilde{\mathcal{S}}_n}. \tag{A4.1}$$

Applying triangle inequality and some algebra, we obtain

$$\|\widetilde{\theta}^*_{h_n} + \Delta\|_1 - \|\widetilde{\theta}^*_{h_n}\|_1 \ge \|\Delta_{\widetilde{\mathcal{S}}_n^c}\|_1 - \|\Delta_{\widetilde{\mathcal{S}}_n}\|_1. \tag{A4.2}$$

We also have, with probability at least $1 - \epsilon_{1,n}$,

$$\begin{aligned} \widehat{\Gamma}_n(\widetilde{\theta}^*_{h_n} + \Delta, h_n) - \widehat{\Gamma}_n(\widetilde{\theta}^*_{h_n}, h_n) &\ge \langle \nabla\widehat{\Gamma}_n(\widetilde{\theta}^*_{h_n}, h_n), \Delta\rangle \\ &\ge -\|\nabla\widehat{\Gamma}_n(\widetilde{\theta}^*_{h_n}, h_n)\|_\infty \cdot \|\Delta\|_1 \\ &\ge -\frac{\lambda_n}{2}\big(\|\Delta_{\widetilde{\mathcal{S}}_n}\|_1 + \|\Delta_{\widetilde{\mathcal{S}}_n^c}\|_1\big), \end{aligned} \tag{A4.3}$$

where the first inequality is by convexity of $\widehat{\Gamma}_n(\theta, h)$ in θ 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 $\mathcal{F}(\widehat{\Delta}) \le 0$, we have

$$0 \ge \frac{\lambda_n}{2}\big(\|\widehat{\Delta}_{\widetilde{\mathcal{S}}_n^c}\|_1 - 3\|\widehat{\Delta}_{\widetilde{\mathcal{S}}_n}\|_1\big),$$

thus proving (A4.1).

Next, we assume that $\|\widehat{\Delta}\|_2 > \eta$. Then, because $\widehat{\Delta} \in \mathcal{C}_{\widetilde{\mathcal{S}}_n}$ and $\mathcal{C}_{\widetilde{\mathcal{S}}_n}$ is star-shaped, there exists some $t \in (0,1)$, such that $t\widehat{\Delta} \in \mathcal{C}_{\widetilde{\mathcal{S}}_n} \cap \big\{\Delta' \in \mathbb{R}^p : \|\Delta'\|_2 = \eta\big\}$. However, by convexity of $\mathcal{F}(\cdot)$,

$$\mathcal{F}(t\widehat{\Delta}) \le t\mathcal{F}(\widehat{\Delta}) + (1-t)\mathcal{F}(0) = t\mathcal{F}(\widehat{\Delta}) \le 0.$$

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 $\Delta \in \mathcal{C}_{\widetilde{\mathcal{S}}_n} \cap \big\{\Delta' \in \mathbb{R}^p : \|\Delta'\|_2 = \eta\big\}$, for some appropriately chosen η, and then complete the proof.

Combining Assumptions 2, 3, and (A4.2), for any $\Delta \in \mathcal{C}_{\widetilde{\mathcal{S}}_n} \cap \big\{\Delta' \in \mathbb{R}^p : \|\Delta'\|_2 = \eta\big\}$, where we take $\eta = 3\widetilde{s}_n^{1/2}\lambda_n/\kappa_1$, and $\lambda_n \le \kappa_1 r/(3\widetilde{s}_n^{1/2})$ so that $\eta \le r$, we have that with probability at least

$1 - \epsilon_{1,n} - \epsilon_{2,n}$,

$$\begin{aligned}\mathcal{F}(\Delta) &\geq \langle \nabla \widehat{\Gamma}_n(\widetilde{\theta}^*_{h_n}, h_n), \Delta\rangle + \kappa_1 \|\Delta\|_2^2 + \lambda_n(\|\widetilde{\theta}^*_{h_n} + \Delta\|_1 - \|\widetilde{\theta}^*_{h_n}\|_1) \\ &\geq -\|\nabla \widehat{\Gamma}_n(\widetilde{\theta}^*_{h_n}, h_n)\|_\infty \cdot \|\Delta\|_1 + \kappa_1\|\Delta\|_2^2 + \lambda_n(\|\Delta_{\widetilde{\mathcal{S}}_n^c}\|_1 - \|\Delta_{\widetilde{\mathcal{S}}_n}\|_1) \\ &\geq -\lambda_n\|\Delta\|_1/2 + \kappa_1\|\Delta\|_2^2 + \lambda_n(\|\Delta_{\widetilde{\mathcal{S}}_n^c}\|_1 - \|\Delta_{\widetilde{\mathcal{S}}_n}\|_1) \\ &\geq \kappa_1\|\Delta\|_2^2 - 3\lambda_n \widetilde{s}_n^{1/2}\|\Delta\|_2/2,\end{aligned}$$

where the first inequality is by Assumption 3, the second is by Hölder's inequality and (A4.2), the third is by Assumption 2, and the last is due to the fact that $\|\Delta_{\widetilde{\mathcal{S}}_n}\|_1 \leq \widetilde{s}_n^{1/2}\|\Delta_{\widetilde{\mathcal{S}}_n}\|_2 \leq \widetilde{s}_n^{1/2}\|\Delta\|_2$.

Then we have

$$\mathcal{F}(\Delta) \geq \kappa_1\eta^2 - 3\widetilde{s}_n^{1/2}\lambda_n\eta/2 = 9\widetilde{s}_n\lambda_n^2/(2\kappa_1) > 0,$$

which, using result from Step I, implies that $\|\widehat{\Delta}\|_2^2 \leq \eta^2 = 9\widetilde{s}_n\lambda_n^2/\kappa_1^2$.

Combining with Assumption 2, we have

$$\|\widehat{\theta}_{h_n} - \theta^*\|_2^2 \leq \frac{18\widetilde{s}_n\lambda_n^2}{\kappa_1^2} + 2\rho_n^2,$$

with probability at least $1 - \epsilon_{1,n} - \epsilon_{2,n}$. This completes the proof of Theorem 2.1. □

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

$$\mathbb{P}\big\{\big|U_k - \mathbb{E}[U_k]\big| \leq A\{\log(np)/n\}^{1/2}, \ \text{ for all } k \in [p]\big\} \geq 1 - \epsilon_n$$

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 $\lambda_n \geq 4(A + A')\{\log(np)/n\}^{1/2} + 8\kappa_x^2 M_f \zeta h_n$, where

$$\begin{aligned}A' =&\{16\sqrt{3}(1+c)^{1/2}M_f + 4\sqrt{3}C_1(1+c)^{1/2}M_f^{1/2}K_1^{-1/2} + 8C_2(1+c) \\ &+ 8C_3(1+c)^{3/2}M_K^{1/2}M_f^{1/2}K_1^{-1/2} + 8C_4(1+c)^2M_KK_1^{-1} + 8M_f(c+2)\}\kappa_x\kappa_u.\end{aligned}$$

for positive absolute constant c, $M_f = M + MM_KC_0$, and $C_1, \ldots, C_4$ as defined in (A3.2). Suppose we have

$$
\begin{aligned}
n > \max\Big\{ & 64(c+2)^2(c+1)\{\log(np)\}^3/3,\ 3, \\
& \frac{48\sqrt{6}M_K\kappa_x^2 q}{K_1 p\{\log(np)\}^{1/2}}, \Big(\frac{2^{10}\cdot 6\cdot\sqrt{6}M_f\kappa_x^2 q}{\kappa_\ell M_\ell p}\Big)^{2/3}, \frac{144\kappa_x^4}{K_1^2p^2\log(np)}, \\
& \Big[\frac{2^{11}\cdot 6\cdot\sqrt{3}(2+c)^{1/2}C_1M_K^{1/2}M_f^{1/2}\kappa_x^2}{K_1^{1/2}\kappa_\ell M_\ell}\Big]^{4/3}\cdot q^{4/3}\{\log(np)\}^{1/3}, \\
& \Big[\frac{2^8\cdot 6\cdot(20+7.5c)(2+c)C_2M_f\kappa_x^2}{\kappa_\ell M_\ell}\Big]^{1/2}\cdot q^{1/2}\log(np), \\
& \Big[\frac{2^8\cdot 6(c+2)^{3/2}C_3\{144(2+c)^2M_KM_f\kappa_x^4K_1^{-1}+192M_f^2\kappa_x^4+8M_f\kappa_x^4\}^{1/2}}{\kappa_\ell M_\ell}\Big]^{\frac{4}{5}}q^{\frac{4}{5}}\{\log(np)\}^{\frac{9}{5}},, \\
& \Big[\frac{2^{10}\cdot 6\cdot\sqrt{6}(2+c)^3C_4\kappa_x^2}{K_1\kappa_\ell M_\ell}\Big]^{2/3}q^{2/3}\{\log(np)\}^{5/3}, \\
& \frac{2^{11}\cdot 6\cdot(20+7.5c)(c+2)M_f\kappa_x^2}{\kappa_\ell M_\ell}q\{\log(np)\}^2,\ \frac{2^6\cdot 3q}{(20+7.5c)M_f\kappa_x^2\kappa_\ell M_\ell\log(np)}, \\
& \frac{2^{20}\{(3M^2\kappa_x^2+2M^2M_K^2C_0^2\kappa_x^2)\vee 2M\}\kappa_x^2}{(\kappa_\ell M_\ell)^2\wedge(16\kappa_\ell M_\ell)^2}q\log\Big(\frac{6ep}{q}\Big), \\
& \frac{2^{24}K_1^2M^2M_K^2\kappa_x^2\log(np)}{(\kappa_\ell M_\ell)^2}\Big\},
\end{aligned}
$$

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

$$\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \le \frac{288s\lambda_n^2}{M_\ell^2\kappa_\ell^2},$$

with probability at least $1 - 12.54\exp(-c\log p) - 2\exp(-c'n) - \epsilon_n\cdot p$, where $c' = (\kappa_\ell^2M_\ell^2 \wedge 64\kappa_\ell M_\ell)/[2^{16}\{(3M^2\kappa_x^2+2M^2M_K^2C_0^2\kappa_x^2)\vee 2M\}\kappa_x^2]$.

Proof. See Proof of Theorem 3.2. □

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 $h_n \ge K_1\{\log(np)/n\}^{1/2}$ for positive absolute constant K_1, and assume $h_n \le C_0$ for positive constant C_0. We also take $\lambda_n \ge 4(A+A')\{\log(np)/n\}^{1/2} + 8\kappa_x^2M_f\zeta h_n^\gamma$, where

$$
\begin{aligned}
A' =& \{16\sqrt{3}(1+c)^{1/2}M_f + 4\sqrt{3}C_1(1+c)^{1/2}M_f^{1/2}K_1^{-1/2} + 8C_2(1+c) \\
& + 8C_3(1+c)^{3/2}M_K^{1/2}M_f^{1/2}K_1^{-1/2} + 8C_4(1+c)^2M_KK_1^{-1} + 8M_f(c+2)\}\kappa_x\kappa_u.
\end{aligned}
$$

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

$$
\begin{aligned}
n > \max\Big\{ & 64(c+2)^2(c+1)\{\log(np)\}^3/3,\ 3, \\
& \frac{48\sqrt{6}M_K\kappa_x^2 q}{K_1 p\{\log(np)\}^{1/2}}, \Big(\frac{2^{10}\cdot 6\cdot\sqrt{6}M_f\kappa_x^2 q}{\kappa_\ell M_\ell p}\Big)^{2/3}, \frac{144\kappa_x^4}{K_1^2 p^2\log(np)}, \\
& \Big[\frac{2^{11}\cdot 6\cdot\sqrt{3}(2+c)^{1/2}C_1 M_K^{1/2}M_f^{1/2}\kappa_x^2}{K_1^{1/2}\kappa_\ell M_\ell}\Big]^{4/3}\cdot q^{4/3}\{\log(np)\}^{1/3}, \\
& \Big[\frac{2^8\cdot 6\cdot(20+7.5c)(2+c)C_2 M_f\kappa_x^2}{\kappa_\ell M_\ell}\Big]^{1/2}\cdot q^{1/2}\log(np), \\
& \Big[\frac{2^8\cdot 6(c+2)^{3/2}C_3\{144(2+c)^2M_K M_f\kappa_x^4 K_1^{-1} + 192M_f^2\kappa_x^4 + 8M_f\kappa_x^4\}^{1/2}}{\kappa_\ell M_\ell}\Big]^{\frac{4}{5}} q^{\frac{4}{5}}\{\log(np)\}^{\frac{9}{5}}, \\
& \Big[\frac{2^{10}\cdot 6\cdot\sqrt{6}(2+c)^3 C_4\kappa_x^2}{K_1\kappa_\ell M_\ell}\Big]^{2/3} q^{2/3}\{\log(np)\}^{5/3}, \\
& \frac{2^{11}\cdot 6\cdot(20+7.5c)(c+2)M_f\kappa_x^2}{\kappa_\ell M_\ell} q\{\log(np)\}^2,\ \frac{2^6\cdot 3q}{(20+7.5c)M_f\kappa_x^2\kappa_\ell M_\ell\log(np)}, \\
& \frac{2^{20}\{(3M^2\kappa_x^2 + 2M^2M_K^2C_0^2\kappa_x^2)\vee 2M\}\kappa_x^2}{(\kappa_\ell M_\ell)^2\wedge(16\kappa_\ell M_\ell)^2} q\log\Big(\frac{6ep}{q}\Big), \\
& \frac{2^{24}K_1^2M^2M_K^2\kappa_x^2\log(np)}{(\kappa_\ell M_\ell)^2}\Big\},
\end{aligned}
$$

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

$$
\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \le \frac{288 s\lambda_n^2}{M_\ell^2\kappa_\ell^2},
$$

with probability at least $1 - 12.54\exp(-c\log p) - 2\exp(-c'n) - \epsilon_n\cdot p$, where $c' = (\kappa_\ell^2 M_\ell^2 \wedge 64\kappa_\ell M_\ell)/[2^{16}\{(3M^2\kappa_x^2 + 2M^2M_K^2C_0^2\kappa_x^2)\vee 2M\}\kappa_x^2]$.

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

In addition to (3.2), denote

$$
U_{1k} = \binom{n}{2}^{-1}\sum_{i<j}\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{u}_{ij},
$$

$$
U_{2k} = \binom{n}{2}^{-1}\sum_{i<j}\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ij}^{\mathsf{T}}(\beta_{h_n}^* - \beta^*),
$$

and observe that

$$
\big|\nabla_k\widehat{L}_n(\beta^*)\big| \le 2\big\{|U_{1k} - \mathbb{E}[U_{1k}]| + |U_k - \mathbb{E}[U_k]| + |\mathbb{E}[U_{2k}]|\big\}, \tag{A4.4}
$$

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

$$
\mathbb{P}\{|U_{1k} - \mathbb{E}[U_{1k}]| \ge A'\{\log(np)/n\}^{1/2}\} \le 6.77\exp\{-(c+1)\log p\}, \tag{A4.5}
$$

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 $Z = |\widetilde{X}_{ijk}\widetilde{X}_{ij}^{\mathsf{T}}(\beta^*_{h_n} - \beta^*)|$, with conditions of lemma satisfied by Assumptions 7, 8, 11, and 14, and then we have

$$\begin{aligned}|\mathbb{E}[U_{2k}]| &\le \mathbb{E}\big[|\widetilde{X}_{ijk}\widetilde{X}_{ij}^{\mathsf{T}}(\beta^*_{h_n} - \beta^*)|\big|\widetilde{W} = 0\big]M + MM_K C_0 \mathbb{E}\big[|\widetilde{X}_{ijk}\widetilde{X}_{ij}^{\mathsf{T}}(\beta^*_{h_n} - \beta^*)|\big] \\ &\le 2\kappa_x^2(M + MM_K C_0)\zeta h_n^{\gamma}.\end{aligned} \tag{A4.6}$$

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

$$\begin{aligned}&\mathbb{P}\big\{\text{for any } k \in [p],\, \big|\nabla_k \widehat{L}_n(\beta^*)\big| \le (2A + 2A')\{\log(np)/n\}^{1/2} + 4\kappa_x^2(M + MM_K C_0)\zeta h_n^{\gamma}\big\} \\ \ge & 1 - 6.77\exp(-c\log p) - p\cdot\epsilon_n,\end{aligned}$$

for positive absolute constant c, and when we appropriately take n bounded from below. Assume $\lambda_n \ge 4(A + A')\{\log(np)/n\}^{1/2} + 8\kappa_x^2(M + MM_K C_0)\zeta h_n^{\gamma}$, which verifies Assumption 2.

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

A4.4 Proof of Theorem 3.3

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

$$\begin{aligned}A''' =&\{16\sqrt{3}M_f(1+c)^{\frac{1}{2}} + 4\sqrt{3}C_1 M_f^{1/2}K_1^{-\frac{1}{2}}(1+c)^{\frac{1}{2}} + 8C_2(1+c) + 8C_3 M_K^{\frac{1}{2}}M_f^{\frac{1}{2}}K_1^{-\frac{1}{2}}(1+c)^{\frac{3}{2}} \\ &+ 8C_4 M_K K_1^{-1}(1+c)^2 + 8M_f(c+2)\}\cdot(\kappa_x\kappa_u + C\kappa_x^2) \\ \eta_n =&\big\|\mathbb{E}\big[\widetilde{X}\widetilde{X}^{\mathsf{T}}\big|\widetilde{W} = 0\big]\big\|_{\infty}.\end{aligned}$$

Here, $C_1, \ldots, C_4$ are as defined in (A3.2), $C > \zeta^2 C_0^{2\gamma}$ and $c > 0$ are some absolute constants, and $M_f = M + MM_K C_0$. Suppose we have

$$
\begin{aligned}
n > \max\Big\{&(C - \zeta^2 C_0^{2\gamma}) s \log(np),\ 64(c+2)^2(c+1)\{\log(np)\}^3/3,\ 3,\\
&\frac{48\sqrt{6} M_K \kappa_x^2 q}{K_1 p \{\log(np)\}^{1/2}}, \Big(\frac{2^{10}\cdot 6 \cdot \sqrt{6} M_f \kappa_x^2 q}{\kappa_\ell M_\ell p}\Big)^{2/3}, \frac{144\kappa_x^4}{K_1^2 p^2 \log(np)},\\
&\Big[\frac{2^{11}\cdot 6\cdot \sqrt{3}(2+c)^{1/2} C_1 M_K^{1/2} M_f^{1/2}\kappa_x^2}{K_1^{1/2}\kappa_\ell M_\ell}\Big]^{4/3}\cdot q^{4/3}\{\log(np)\}^{1/3},\\
&\Big[\frac{2^8\cdot 6\cdot (20+7.5c)(2+c) C_2 M_f \kappa_x^2}{\kappa_\ell M_\ell}\Big]^{1/2}\cdot q^{1/2}\log(np),\\
&\Big[\frac{2^8\cdot 6 (c+2)^{3/2} C_3 \{144(2+c)^2 M_K M_f \kappa_x^4 K_1^{-1} + 192 M_f^2\kappa_x^4 + 8 M_f \kappa_x^4\}^{1/2}}{\kappa_\ell M_\ell}\Big]^{\frac{4}{5}} q^{\frac{4}{5}}\{\log(np)\}^{\frac{9}{5}},\\
&\Big[\frac{2^{10}\cdot 6\cdot\sqrt{6}(2+c)^3 C_4\kappa_x^2}{K_1\kappa_\ell M_\ell}\Big]^{2/3} q^{2/3}\{\log(np)\}^{5/3},\\
&\frac{2^{11}\cdot 6\cdot (20+7.5c)(c+2) M_f\kappa_x^2}{\kappa_\ell M_\ell} q\{\log(np)\}^2,\ \frac{2^6\cdot 3q}{(20+7.5c) M_f \kappa_x^2 \kappa_\ell M_\ell \log(np)},\\
&\frac{2^{20}\{(3M^2\kappa_x^2 + 2M^2 M_K^2 C_0^2 \kappa_x^2)\vee 2M\}\kappa_x^2}{(\kappa_\ell M_\ell)^2 \wedge (16\kappa_\ell M_\ell)^2} q \log\Big(\frac{6ep}{q}\Big),\\
&\frac{2^{24} K_1^2 M^2 M_K^2 \kappa_x^2 \log(np)}{(\kappa_\ell M_\ell)^2}\Big\},
\end{aligned}
$$

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

$$
\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \le \frac{288 s\lambda_n^2}{M_\ell^2\kappa_\ell^2} + \frac{2s\log(np)}{n} + \Big\{\frac{288 n\lambda_n^2}{M_\ell^2\kappa_\ell^2\log(np)} + 2\Big\}\cdot\zeta^2 h_n^{2\gamma},
$$

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 $\theta^* = \beta^*$, $\Gamma_0(\theta) = L_0(\beta)$, $\widehat{\Gamma}_n(\theta, h) = \widehat{L}_n(\beta, h)$, $\Gamma_h(\theta) = \mathbb{E}\widehat{L}_n(\beta, h)$.

We take $\widetilde{\theta}_{h_n} = \widetilde{\beta}^*_{h_n}$ such that, for each $j \in [p]$,

$$
\widetilde{\beta}^*_{h_n,j} = \begin{cases} \beta^*_{h_n,j}, & \text{if } |\beta^*_{h_n,j}| > \{\log(np)/n\}^{1/2};\\ 0, & \text{if otherwise.}\end{cases} \tag{A4.7}
$$

Then under Assumption 14, we have

$$
\begin{aligned}
\rho_n^2 &\le s\log(np)/n + \zeta^2 h_n^{2\gamma},\\
s_n &\le s + \frac{\zeta^2 n h_n^{2\gamma}}{\log(np)}.
\end{aligned} \tag{A4.8}
$$

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.

□

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 $\eta_n = \big\|\mathbb{E}\big[\widetilde{X}\widetilde{X}^{\mathsf{T}}\big|\widetilde{W} = 0\big]\big\|_{\infty}$. We also take $\lambda_n \geq 4(A' + A'' + A + M\eta_n)\{\log(np)/n\}^{1/2} + 8MM_K C^{1/2}\kappa_x^2 h_n$, 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

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

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

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

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

$$U_{1k} = \binom{n}{2}^{-1}\sum_{i<j}\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{u}_{ij},$$

$$U_{2k} = \binom{n}{2}^{-1}\sum_{i<j}\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ij}^{\mathsf{T}}(\beta^* - \widetilde{\beta}^*_{h_n}),$$

$$U_{3k} = \binom{n}{2}^{-1}\sum_{i<j}\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ij}^{\mathsf{T}}(\beta^*_{h_n} - \widetilde{\beta}^*_{h_n}),$$

and observe that

$$\big|\nabla_k \widehat{L}_n(\widetilde{\beta}^*_{h_n}, h_n)\big| \leq 2(|U_{1k} - \mathbb{E}[U_{1k}]| + |U_{2k} - \mathbb{E}[U_{2k}]| + |U_k - \mathbb{E}[U_k]| + |\mathbb{E}[U_{3k}]|), \tag{A4.9}$$

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

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

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

We have $\|\beta^* - \widetilde{\beta}^*_{h_n}\|_2^2 \leq s\log(np)/n + \zeta^2 h_n^{2\gamma} < C$ 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 $D_i = (X_{ik}, X_i^{\mathsf{T}}(\beta^* - \widetilde{\beta}^*_{h_n}), W_i)$, with conditions of lemma satisfied by Assumptions 7, 8, 11, and that $\|\beta^* - \widetilde{\beta}^*_{h_n}\|_2^2 < C$, and we have

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

for positive constants A' and c, and when we assume $n > \max\big\{64(c+2)^2(c+1)\{\log(np)\}^3/3,\, 3\big\}$. 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

$$\begin{aligned}\big|\mathbb{E}[U_{3k}]\big| &\leq M\mathbb{E}\big[|\widetilde{X}_{ijk}\widetilde{X}_{ij}^{\mathsf{T}}(\beta^*_{h_n} - \widetilde{\beta}^*_{h_n})|\big|\widetilde{W}_{ij} = 0\big] + MM_K h_n\mathbb{E}\big[|\widetilde{X}_{ijk}\widetilde{X}_{ij}^{\mathsf{T}}(\beta^*_{h_n} - \widetilde{\beta}^*_{h_n})|\big] \\ &\leq M\eta_n\{\log(np)/n\}^{1/2} + MM_K C^{1/2}\cdot 2\kappa_x^2 h_n\end{aligned} \tag{A4.12}$$

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

$$\mathbb{P}\Big\{\text{for any } k\in[p],\ \big|\nabla_k \widehat{L}_n(\widetilde{\beta}^*_{h_n},h_n)\big| \le \big\{2(A'+A''+A+M\eta_n)\{\frac{\log(np)}{n}\}^{1/2}+4MM_KC^{1/2}\kappa_x^2h_n\big\}$$
$$\ge 1-13.54p\exp\{-(c+1)\log p\}-\epsilon_n\cdot p,$$

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 $\lambda_n \ge 4(A'+A''+A+M\eta_n)\{\log(np)/n\}^{1/2}+8MM_KC^{1/2}\kappa_x^2h_n$. This completes the proof. $\square$

A4.5 Proof of Theorem 3.4

Theorem A4.5 (Theorem 3.4)**.** Assume $h\le C_0$ for positive constant C_0, and that $h^2\le \kappa_\ell M_\ell\cdot(4MM_K\kappa_x^2)^{-1}$. Under Assumptions 6-8, 9′, 10-11, and 13, and when g is (L,α)-Hölder for $\alpha\ge 1$ (g has bounded support when $\alpha>1$), we have

$$\|\beta_h^*-\beta^*\|_2\le \zeta h,$$

where

$$\zeta=\max\Big\{4\cdot\Big(\frac{L_\alpha^2MM_K+MM_K\mathbb{E}\widetilde{u}^2/2}{\kappa_\ell M_\ell}\Big)^{1/2},\ \frac{16\kappa_x(M+MM_KC_0^2)^{1/2}\cdot L_\alpha^2MM_K}{\kappa_\ell M_\ell}\Big\},$$

where L_α 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 $M_d=M_a=0$, 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. $\square$

A4.6 Proof of Theorem 3.5

Theorem A4.6. Assume $h\le C_0$ for positive constant C_0, and that $h^2\le\kappa_\ell M_\ell\cdot(4MM_K\kappa_x^2)^{-1}$. Under Assumptions 5, 6-8, 9′, 10-11, and 13, we have

$$\|\beta_h^*-\beta^*\|_2\le\zeta h^\gamma,$$

where

$$\zeta=\max\Big\{4\cdot\Big(\frac{M_g^2MM_KC_0^{2\alpha-2\gamma}+M_d^2M_aC_0^{1-2\gamma}+MM_K\mathbb{E}\widetilde{u}^2C_0^{2-2\gamma}/2}{\kappa_\ell M_\ell}\Big)^{1/2},$$
$$\frac{16\kappa_x(M+MM_KC_0^2)^{1/2}\cdot(M_g^2MM_KC_0^{2\alpha-2\gamma}+M_d^2M_aC_0^{1-2\gamma})^{1/2}}{\kappa_\ell M_\ell}\Big\},$$

$\gamma=\alpha$ if $M_dM_a=0$, and $\gamma=\min\big\{\alpha,\ 1/2\big\}$ if otherwise.

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

Step I. We show that $|L_0(\beta_h^*)-L_0(\beta^*)|$ is lower bounded for $L_0(\beta)=\mathbb{E}\big[(\widetilde{Y}-\widetilde{X}^\top\beta)^2\big|\widetilde{W}=0\big]f_{\widetilde{W}}(0)$. By Assumptions 10 and 9′, we have

$$\lambda_{\min}\Big(\frac{\partial^2L_0(\beta)}{\partial\beta^2}\Big)=2\lambda_{\min}\big(\mathbb{E}\big[\widetilde{X}\widetilde{X}^\top\big|\widetilde{W}=0\big]\big)f_{\widetilde{W}}(0)\ge 2\kappa_\ell M_\ell.$$

Therefore, for some $\beta_t = \beta^*_{h_n} + t(\beta^* - \beta^*_{h_n})$, $t \in [0,1]$, we have

$$L_0(\beta^*_h) - L_0(\beta^*) = \frac{1}{2}(\beta^*_h - \beta^*)^\top \frac{\partial^2 L_0(\beta)}{\partial \beta^2}\Big|_{\beta=\beta_t}(\beta^*_h - \beta^*) \geq \kappa_\ell M_\ell \|\beta^*_h - \beta^*\|_2^2.$$

Step II. We show that $|L_{h_n}(\beta) - L_0(\beta)|$ is upper bounded. Observe that

$$\begin{aligned}
|L_h(\beta) - L_0(\beta)| \leq & \Big|\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}}{h}\Big)\{\widetilde{X}^\top(\beta-\beta^*)\}^2\Big] - \mathbb{E}\big[\{\widetilde{X}^\top(\beta-\beta^*)\}^2\big|\widetilde{W}=0\big]f_{\widetilde{W}}(0)\Big| \\
& + \mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}_{ij}}{h}\Big)\big\{g(W_i) - g(W_j)\big\}^2\Big] \\
& + \Big|\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}}{h}\Big)\widetilde{u}^2\Big] - \mathbb{E}\big[\widetilde{u}^2\big|\widetilde{W}=0\big]f_{\widetilde{W}}(0)\Big| \\
& + 2\Big|\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}_{ij}}{h}\Big)\widetilde{X}^\top(\beta-\beta^*)\big\{g(W_i)-g(W_j)\big\}\Big]\Big|.
\end{aligned} \tag{A4.13}$$

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

By Taylor's expansion, we have

$$\begin{aligned}
& \Big|\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}}{h}\Big)\{\widetilde{X}^\top(\beta-\beta^*)\}^2\Big] - \mathbb{E}\big[\{\widetilde{X}^\top(\beta-\beta^*)\}^2\big|\widetilde{W}=0\big]f_{\widetilde{W}}(0)\Big| \\
=& \Big|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{h}K\Big(\frac{w}{h}\Big)v^2 f_{\widetilde{W}|\widetilde{X}^\top(\beta-\beta^*)}(w,v)\,dw\,dF_{\widetilde{X}^\top(\beta-\beta^*)}(v) \\
& - \int_{-\infty}^{\infty} v^2 f_{\widetilde{W}|\widetilde{X}^\top(\beta-\beta^*)}(0,v)\,dF_{\widetilde{X}^\top(\beta-\beta^*)}(v)\Big| \\
=& \Big|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K(w)v^2\big\{f_{\widetilde{W}|\widetilde{X}^\top(\beta-\beta^*)}(wh,v) - f_{\widetilde{W}|\widetilde{X}^\top(\beta-\beta^*)}(0,v)\big\}\,dw\,dF_{\widetilde{X}^\top(\beta-\beta^*)}(v)\Big| \\
=& \Big|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K(w)v^2\Big\{\frac{\partial f_{\widetilde{W}|\widetilde{X}^\top(\beta-\beta^*)}(w,v)}{\partial w}\Big|_{(0,v)}wh \\
& + \frac{\partial^2 f_{\widetilde{W}|\widetilde{X}^\top(\beta-\beta^*)}(w,v)}{\partial w^2}\Big|_{(\tau wh,v)}w^2h^2\Big\}\,dw\,dF_{\widetilde{X}^\top(\beta-\beta^*)}(v)\Big|,
\end{aligned}$$

where because $(\widetilde{W}, \widetilde{X}^\top(\beta-\beta^*))$ and $(-\widetilde{W}, -\widetilde{X}^\top(\beta-\beta^*))$ are identically distributed, we have

$$\begin{aligned}
& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K(w)v^2\Big\{\frac{\partial f_{\widetilde{W}|\widetilde{X}^\top(\beta-\beta^*)}(w,v)}{\partial w}\Big|_{(0,v)}wh\Big\}\,dw\,dF_{\widetilde{X}^\top(\beta-\beta^*)}(v) \\
=& \int_{0}^{\infty}\int_{-\infty}^{\infty}K(w)v^2\Big\{\frac{\partial f_{\widetilde{W}|\widetilde{X}^\top(\beta-\beta^*)}(w,v)}{\partial w}\Big|_{(0,v)} + \frac{\partial f_{\widetilde{W}|\widetilde{X}^\top(\beta-\beta^*)}(w,v)}{\partial w}\Big|_{(0,-v)}\Big\}wh\,dw\,dF_{\widetilde{X}^\top(\beta-\beta^*)}(v) \\
=& 0.
\end{aligned}$$

Therefore, using Assumptions 7, 8 (Lemmas A4.15 and A4.16), and 13, we further have

$$\begin{aligned}
& \Big|\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}}{h}\Big)\big\{\widetilde{X}^\top(\beta-\beta^*)\big\}^2\Big] - \mathbb{E}\big[\big\{\widetilde{X}^\top(\beta-\beta^*)\big\}^2|\widetilde{W}=0\big]f_{\widetilde{W}}(0)\Big| \\
=& \Big|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K(w)v^2\Big\{\frac{\partial^2 f_{\widetilde{W}|\widetilde{X}^\top(\beta-\beta^*)}(w,v)}{\partial w^2}\big|_{(\tau wh,v)}\Big\}w^2h^2\,dw\,dF_{\widetilde{X}^\top(\beta-\beta^*)}(v)\Big| \\
\leq & MM_K\mathbb{E}\big[\{\widetilde{X}^\top(\beta-\beta^*)\}^2\big]h^2 \leq 2MM_K\kappa_x^2\|\beta-\beta^*\|_2^2h^2.
\end{aligned} \tag{A4.14}$$

Using an identical argument, by Assumptions 7, 8 (Lemmas A4.15 and A4.16), and finite second moment assumption $\mathbb{E}[\widetilde{u}^2] < \infty$, we have

$$\Big|\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}}{h}\Big)\widetilde{u}^2\Big] - \mathbb{E}[\widetilde{u}^2|\widetilde{W}=0]f_{\widetilde{W}}(0)\Big| \le MM_K\mathbb{E}[\widetilde{u}^2]h^2. \tag{A4.15}$$

By Assumption 5, we have

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}_{ij}}{h}\Big)\big\{g(W_i)-g(W_j)\big\}^2\Big]\\
\le& 2M_g^2\,\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}}{h}\Big)|\widetilde{W}|^{2\alpha}\Big] + 2M_d^2\,\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}_{ij}}{h}\Big)\,\mathbb{1}\big\{(W_i,W_j)\in A\big\}\Big]\\
\le& 2M_g^2\,\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}}{h}\Big)|\widetilde{W}|^{2\alpha}\Big] + 2M_d^2M_ah,
\end{aligned}$$

where

$$\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}}{h}\Big)|\widetilde{W}|^{2\alpha}\Big] = \int_{-\infty}^{\infty}K(w)|w|^{2\alpha}h^{2\alpha}f_{\widetilde{W}}(wh)\,dw \le MM_Kh^{2\alpha}.$$

Therefore, we have

$$\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}_{ij}}{h}\Big)\big\{g(W_i)-g(W_j)\big\}^2\Big] \le 2M_g^2MM_Kh^{2\alpha} + 2M_d^2M_ah. \tag{A4.16}$$

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

$$\begin{aligned}
&\Big|\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}_{ij}}{h}\Big)\widetilde{X}_{ij}^{\intercal}(\beta-\beta^*)\big\{g(W_i)-g(W_j)\big\}\Big]\Big|\\
\le&\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}_{ij}}{h}\Big)\big\{\widetilde{X}_{ij}^{\intercal}(\beta-\beta^*)\big\}^2\Big]^{1/2}\cdot\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{\widetilde{W}_{ij}}{h}\Big)\big\{g(W_i)-g(W_j)\big\}^2\Big]^{1/2}\\
\le&\big(2MM_K\kappa_x^2\|\beta-\beta^*\|_2^2h^2 + 2\kappa_x^2\|\beta-\beta^*\|_2^2M\big)^{1/2}\times\big(2M_g^2MM_Kh^{2\alpha}+2M_d^2M_ah\big)^{1/2}\\
\le& a_1\|\beta-\beta^*\|_2h^{\gamma},
\end{aligned}\tag{A4.17}$$

where $\gamma=\alpha$ if $M_dM_a=0$, and $\gamma=\min\{\alpha,1/2\}$ if otherwise, and $a_1 = 2\kappa_x(M+MM_KC_0^2)^{1/2}\cdot(M_g^2MM_KC_0^{2\alpha-2\gamma}+M_d^2M_aC_0^{1-2\gamma})^{1/2}$.

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

$$|L_h(\beta)-L_0(\beta)| \le 2a_1\|\beta-\beta^*\|_2h^{\gamma} + a_2h^{2\gamma} + a_3\|\beta-\beta^*\|_2^2h^2,$$

where $a_2 = 2M_g^2MM_KC_0^{2\alpha-2\gamma}+2M_d^2M_aC_0^{1-2\gamma}+MM_K\mathbb{E}\widetilde{u}^2C_0^{2-2\gamma}$, and $a_3 = 2MM_K\kappa_x^2$.

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

$$\begin{aligned}
\kappa_\ell M_\ell\|\beta_h^*-\beta^*\|_2^2 &\le L_0(\beta_h^*)-L_0(\beta^*)\\
&= L_0(\beta_h^*)-L_h(\beta_h^*)+L_h(\beta^*)-L_0(\beta^*)+L_h(\beta_h^*)-L_h(\beta^*)\\
&\le |L_0(\beta_h^*)-L_h(\beta_h^*)|+|L_h(\beta^*)-L_0(\beta^*)|\\
&\le 2a_1\|\beta_h^*-\beta^*\|_2h^{\gamma}+2a_2h^{2\gamma}+a_3\|\beta_h^*-\beta^*\|_2^2h^2.
\end{aligned}$$

When $h^2\le\kappa_\ell M_\ell/(2a_3)$, we have

$$\kappa_\ell M_\ell\|\beta_h^*-\beta^*\|_2^2 \le 4a_1\|\beta_h^*-\beta^*\|_2h^{\gamma}+4a_2h^{2\gamma},$$

which further implies that

$$\|\beta_h^* - \beta^*\|_2 \le \max\Big\{\Big(\frac{8a_2}{\kappa_\ell M_\ell}\Big)^{1/2}, \frac{8a_1}{\kappa_\ell M_\ell}\Big\} \cdot h^\gamma.$$

This completes the proof. □

A4.7 Proof of Theorem 2.3

Theorem A4.7 (Theorem 2.3)**.** Assume $h_n \ge K_1\{\log(np)/n\}^{1/2}$ for positive absolute constant K_1, and assume $h_n \le C_0$ for positive constant C_0. We denote c to be some positive absolute constant, $c' = (\kappa_\ell^2 M_\ell^2 \wedge 64\kappa_\ell M_\ell)/[2^{16}\{(3M^2\kappa_x^2 + 2M^2M_K^2C_0^2\kappa_x^2) \vee 2M\}\kappa_x^2]$, $M_f = M + MM_KC_0$, and $C_1, \ldots, C_4$ as defined in (A3.2) Also denote

$$\begin{aligned}
&\tau_1 = \sqrt{2}(2+c)^{1/2}\kappa_x K_1^{-1}(BM_KC_0^a + DM_K),\\
&\tau_2 = \sqrt{2}(2+c)^{1/2}\kappa_x\{BM_KM(1+C_0)C_0^a + DM_f\},\\
&\tau_3 = 4M_K^2M^2\cdot(BC_0^a + D)^2\cdot(1+C_0^2)\cdot\kappa_x^2,\\
&\tau_4 = \big\{4B^2MM_K\kappa_x^2(1+C_0)C_0^{2a-\gamma_1} + 2D^2\cdot(12M_f\kappa_x^4)^{1/2}\cdot E^{1/2}C_0^{-1/2-\gamma_1}\big\}\cdot M_KK_1^{\gamma_1},\\
&\tau_5 = 4(2+c)\kappa_x^2\{BMM_K(1+C_0)C_0^{2a} + D^2M_f\}M_KK_1^{-1},
\end{aligned}$$

and

$$\begin{aligned}
A' =&\{16\sqrt{3}M_f(1+c)^{\frac12} + 4\sqrt{3}C_1M_f^{\frac12}K_1^{-\frac12}(1+c)^{\frac12} + 8C_2(1+c) + 8C_3M_K^{\frac12}M_f^{\frac12}K_1^{-\frac12}(1+c)^{\frac32}\\
&+ 8C_4M_KK_1^{-1}(1+c)^2 + 8M_f(c+2)\}\cdot(\kappa_x\kappa_u + C\kappa_x^2)\\
A'' =&4\tau_3^{1/2}(1+c)^{1/2} + 2C_1\tau_4^{1/2}(1+c)^{1/2} + 2C_2\tau_2(1+c) + 2C_3\tau_5^{1/2}(1+c)^{3/2}\\
&+ 2C_4\tau_1(1+c)^2 + 4M_f\cdot(BC_0^a + D)\cdot(c+2)\kappa_x,
\end{aligned}$$

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

$$
\begin{aligned}
n > \max\Big\{ & 64(c+2)^2(c+1)\{\log(np)\}^3/3,\ 64(c+2)^2(c+1)\tau_2^2\tau_3^{-1}\{\log(np)\}^4,\ \{\log(np)\}^{5/3},\ 3, \\
& \frac{48\sqrt{6}M_K\kappa_x^2 q}{K_1 p\{\log(np)\}^{1/2}}, \Big(\frac{2^{10}\cdot 6\cdot\sqrt{6}M_f\kappa_x^2 q}{\kappa_\ell M_\ell p}\Big)^{2/3}, \frac{144\kappa_x^4}{K_1^2p^2\log(np)}, \\
& \Big[\frac{2^{11}\cdot 6\cdot\sqrt{3}(2+c)^{1/2}C_1M_K^{1/2}M_f^{1/2}\kappa_x^2}{K_1^{1/2}\kappa_\ell M_\ell}\Big]^{4/3}\cdot q^{4/3}\{\log(np)\}^{1/3}, \\
& \Big[\frac{2^8\cdot 6\cdot(20+7.5c)(2+c)C_2M_f\kappa_x^2}{\kappa_\ell M_\ell}\Big]^{1/2}\cdot q^{1/2}\log(np), \\
& \Big[\frac{2^8\cdot 6(c+2)^{\frac{3}{2}}C_3\{144(2+c)^2M_KM_f\kappa_x^4K_1^{-1}+192M_f^2\kappa_x^4+8M_f\kappa_x^4\}^{\frac{1}{2}}}{\kappa_\ell M_\ell}\Big]^{\frac{4}{5}}q^{\frac{4}{5}}\{\log(np)\}^{\frac{9}{5}}, \\
& \Big[\frac{2^{10}\cdot 6\cdot\sqrt{6}(2+c)^3C_4\kappa_x^2}{K_1\kappa_\ell M_\ell}\Big]^{2/3}q^{2/3}\{\log(np)\}^{5/3}, \\
& \frac{2^{11}\cdot 6\cdot(20+7.5c)(c+2)M_f\kappa_x^2}{\kappa_\ell M_\ell}q\{\log(np)\}^2,\ \frac{2^6\cdot 3q}{(20+7.5c)M_f\kappa_x^2\kappa_\ell M_\ell\log(np)}, \\
& \frac{2^{20}\{(3M^2\kappa_x^2+2M^2M_K^2C_0^2\kappa_x^2)\vee 2M\}\kappa_x^2}{(\kappa_\ell M_\ell)^2\wedge(16\kappa_\ell M_\ell)^2}q\log\Big(\frac{6ep}{q}\Big), \\
& \frac{2^{24}K_1^2M^2M_K^2\kappa_x^2\log(np)}{(\kappa_\ell M_\ell)^2}\Big\}.
\end{aligned}
\tag{A4.18}
$$

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,α)-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_α is the Lipschitz constant for g ($L_\alpha = L$ when $= 1$), $D = E = 0$, $a = 1$, and assume $\lambda_n \geq 4(A'' + A')\{\log(np)/n\}^{1/2} + 8\kappa_x^2 M_f\zeta h_n$, where

$$\zeta = \max\Big\{4\cdot\Big(\frac{L_\alpha^2MM_K + MM_K\mathbb{E}\widetilde{u}^2/2}{\kappa_\ell M_\ell}\Big)^{1/2}, \frac{16\kappa_x(M+MM_KC_0^2)^{1/2}\cdot L_\alpha^2MM_K}{\kappa_\ell M_\ell}\Big\}.$$

Then we have

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

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 $B = M_g$, $D = M_d$, $E = M_a$ and $a = \alpha$. Further assume that

$\lambda_n \geq 4(A'' + A')\{\log(np)/n\}^{1/2} + 8\kappa_x^2 M_f \zeta h_n^\gamma$, where

$$\zeta = \max\Big\{4 \cdot \Big(\frac{M_g^2 M M_K C_0^{2\alpha-2\gamma} + M_d^2 M_a C_0^{1-2\gamma} + M M_K \mathbb{E}\widetilde{u}^2 C_0^{2-2\gamma}/2}{\kappa_\ell M_\ell}\Big)^{1/2},$$
$$\frac{16\kappa_x (M + M M_K C_0^2)^{1/2} \cdot (M_g^2 M M_K C_0^{2\alpha-2\gamma} + M_d^2 M_a C_0^{1-2\gamma})^{1/2}}{\kappa_\ell M_\ell}\Big\},$$

where $\gamma = \alpha$ if $M_d M_a = 0$, and $\gamma = \min\{\alpha,\, 1/2\}$ if otherwise. Then we have

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

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 $q = 2305\{s + \zeta^2 n h_n^{2\gamma}/\log(np)\}$, and take $B = M_g$, $D = M_d$, $E = M_a$ and $a = \alpha$. Denote C to be some positive absolute constant $C > \zeta^2 C_0^{2\gamma}$, and suppose $n \geq (C - \zeta^2 C_0^{2\gamma}) s\log(np)$, where

$$\zeta = \max\Big\{4 \cdot \Big(\frac{M_g^2 M M_K C_0^{2\alpha-2\gamma} + M_d^2 M_a C_0^{1-2\gamma} + M M_K \mathbb{E}\widetilde{u}^2 C_0^{2-2\gamma}/2}{\kappa_\ell M_\ell}\Big)^{1/2},$$
$$\frac{16\kappa_x (M + M M_K C_0^2)^{1/2} \cdot (M_g^2 M M_K C_0^{2\alpha-2\gamma} + M_d^2 M_a C_0^{1-2\gamma})^{1/2}}{\kappa_\ell M_\ell}\Big\},$$

where $\gamma = \alpha$ if $M_d M_a = 0$, and $\gamma = \min\{\alpha,\, 1/2\}$ if otherwise. Further assume $\lambda_n \geq 4(A' + A'' + M\eta_n)\{\log(np)/n\}^{1/2} + 8 M M_K C^{1/2} \kappa_x^2 h_n$. Then we have

$$\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \leq \frac{288 s\lambda_n^2}{M_\ell^2 \kappa_\ell^2} + \frac{2s\log(np)}{n} + \Big\{\frac{288 n\lambda_n^2}{M_\ell^2 \kappa_\ell^2 \log(np)} + 2\Big\} \cdot \zeta^2 h_n^{2\gamma},$$

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 $\lambda_n \geq 4(A+A')\cdot\{\log(np)/n\}^{1/2} + 4\sqrt{2} M_g M_K M \kappa_x (1 + C_0) h_n$, where

$$\begin{aligned}
A =&\{16\sqrt{3} M_f (1+c)^{1/2} + 4\sqrt{3} C_1 M_f^{1/2} K_1^{-1/2} (1+c)^{1/2} + 8C_2(1+c) \\
&+ 8C_3 M_K^{1/2} M_f^{1/2} K_1^{-1/2} (1+c)^{3/2} + 8C_4 M_K K_1^{-1}(1+c)^2 + 8M_f(c+2)\}\kappa_x\kappa_u, \\
A' =& 8 M M_K M_g C_0 (1+C_0)\kappa_x (1+c)^{1/2} + 2C_1 M_g M^{1/2} M_K^{3/2} \kappa_x^{1/2} (1+C_0)^{1/2} C_0^{5/4} K_1^{-1/4} (1+c)^{1/2} \\
&+ 2\sqrt{2} C_2 M M_K M_g (1+C_0)\kappa_x K_1 (1+c)^{3/2} + 4C_3 M M_K^{3/2} M_g^{1/2} (1+C_0)^{1/2} C_0^{1/2} \kappa_x (1+c)^2 \\
&+ 2\sqrt{2} C_4 M_K M_g C_0 \kappa_x K_1^{-1} (1+c)^{5/2} + 2\sqrt{2} M M_K M_g (1+C_0) C_0,
\end{aligned}$$

for positive absolute constant c, $M_f = M + MM_KC_0$, and $C_1, \ldots, C_4$ as defined in (A3.2). Suppose we have

$$
\begin{aligned}
n > \max\Big\{ & 64(c+2)^2(c+1)\{\log(np)\}^3/3,\ 64(c+2)^3(c+1)\{\log(np)\}^4,\ \{\log(np)\}^{5/3},\ 3,\\
& \frac{48\sqrt{6}M_K\kappa_x^2 q}{K_1 p\{\log(np)\}^{1/2}}, \Big(\frac{2^{10}\cdot 6\cdot\sqrt{6}M_f\kappa_x^2 q}{\kappa_\ell M_\ell p}\Big)^{2/3}, \frac{144\kappa_x^4}{K_1^2p^2\log(np)},\\
& \Big[\frac{2^{11}\cdot 6\cdot\sqrt{3}(2+c)^{1/2}C_1M_K^{1/2}M_f^{1/2}\kappa_x^2}{K_1^{1/2}\kappa_\ell M_\ell}\Big]^{4/3}\cdot q^{4/3}\{\log(np)\}^{1/3},\\
& \Big[\frac{2^8\cdot 6\cdot(20+7.5c)(2+c)C_2M_f\kappa_x^2}{\kappa_\ell M_\ell}\Big]^{1/2}\cdot q^{1/2}\log(np),\\
& \Big[\frac{2^8\cdot 6(c+2)^{\frac{3}{2}}C_3\{144(2+c)^2M_KM_f\kappa_x^4K_1^{-1}+192M_f^2\kappa_x^4+8M_f\kappa_x^4\}^{\frac{1}{2}}}{\kappa_\ell M_\ell}\Big]^{\frac{4}{5}}q^{\frac{4}{5}}\{\log(np)\}^{\frac{9}{5}},\\
& \Big[\frac{2^{10}\cdot 6\cdot\sqrt{6}(2+c)^3C_4\kappa_x^2}{K_1\kappa_\ell M_\ell}\Big]^{2/3}q^{2/3}\{\log(np)\}^{5/3},\\
& \frac{2^{11}\cdot 6\cdot(20+7.5c)(c+2)M_f\kappa_x^2}{\kappa_\ell M_\ell}q\{\log(np)\}^2,\ \frac{2^6\cdot 3q}{(20+7.5c)M_f\kappa_x^2\kappa_\ell M_\ell\log(np)},\\
& \frac{2^{20}\{(3M^2\kappa_x^2+2M^2M_K^2C_0^2\kappa_x^2)\vee 2M\}\kappa_x^2}{(\kappa_\ell M_\ell)^2\wedge(16\kappa_\ell M_\ell)^2}q\log\Big(\frac{6ep}{q}\Big),\\
& \frac{2^{24}K_1^2M^2M_K^2\kappa_x^2\log(np)}{(\kappa_\ell M_\ell)^2}\Big\}.
\end{aligned}
\tag{A4.19}
$$

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

$$
\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \le \frac{288s\lambda_n^2}{M_\ell^2\kappa_\ell^2},
$$

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

$$
c' = (\kappa_\ell^2M_\ell^2 \wedge 64\kappa_\ell M_\ell)/[2^{16}\{(3M^2\kappa_x^2 + 2M^2M_K^2C_0^2\kappa_x^2)\vee 2M\}\kappa_x^2].
$$

Proof of Theorem 2.2. We adopt the framework as described in Section 2.1 for $\theta^* = \beta^*$, $\Gamma_0(\theta) = L_0(\beta)$, $\widehat{\Gamma}_n(\theta, h) = \widehat{L}_n(\beta, h)$, $\Gamma_h(\theta) = \mathbb{E}\widehat{L}_n(\beta, h)$, and take $\widetilde{\theta}^*_{h_n} = \beta^*$, which yields $s_n \le s$ 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. $\square$

A4.9 Proof of Theorem A3.1

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

$$
\begin{aligned}
n > \max\Big\{ & \frac{48\sqrt{6}M_K\kappa_x^2 q}{K_1 p\{\log(np)\}^{1/2}}, \Big(\frac{384\sqrt{6}M_f\kappa_x^2 q}{tp}\Big)^{2/3}, \frac{144\kappa_x^4}{K_1^2p^2\log(np)}, \\
& \Big[\frac{768\sqrt{3}(2+c)^{1/2}C_1M_K^{1/2}M_f^{1/2}\kappa_x^2}{K_1^{1/2}t}\Big]^{4/3} \cdot q^{4/3}\{\log(np)\}^{1/3}, \\
& \Big[\frac{96(20+7.5c)(2+c)C_2M_f\kappa_x^2}{t}\Big]^{1/2} \cdot q^{1/2}\log(np), \\
& \Big[\frac{96(c+2)^{\frac{3}{2}}C_3\{144(2+c)^2M_KM_f\kappa_x^4K_1^{-1}+192M_f^2\kappa_x^4+8M_f\kappa_x^4\}^{\frac{1}{2}}}{t}\Big]^{4/5} q^{\frac{4}{5}}\{\log(np)\}^{\frac{9}{5}}, \\
& \Big[\frac{384\sqrt{6}(2+c)^3C_4\kappa_x^2}{K_1t}\Big]^{2/3} q^{2/3}\{\log(np)\}^{5/3}, \\
& \frac{768(20+7.5c)(c+2)M_f\kappa_x^2}{t} q\{\log(np)\}^2, \frac{12q}{(20+7.5c)M_f\kappa_x^2t\log(np)}, \\
& \frac{2^{12}\{(3M^2\kappa_x^2+2M^2M_K^2C_0^2\kappa_x^2)\vee 2M\}\kappa_x^2}{t^2\wedge(16t)} q\log\Big(\frac{6ep}{q}\Big), \\
& \frac{2^{16}K_1^2M^2M_K^2\kappa_x^2\log(np)}{t^2}\Big\},
\end{aligned}
\tag{A4.20}
$$

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

$$
\|\widehat{T}_n - \mathbb{E}\widehat{T}_n\|_{2,q} \le t
$$

with probability at least $1 - 5.77\exp(-c\log p) - 2\exp(-c'n)$, where $c' = (t^2 \wedge 4t)/[2^8\{(3M^2\kappa_x^2 + 2M^2M_K^2C_0^2\kappa_x^2) \vee 2M\}\kappa_x^2]$.

Proof. We denote

$$
\begin{aligned}
X_{h_n} &= \Big(\frac{1}{h_n^{1/2}}K^{1/2}\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ij}^{\mathsf{T}}\Big)_{\binom{n}{2}\times p} \text{ to be a } \binom{n}{2} \times p \text{ matrix,} \\
\Sigma_{h_n} &= \mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}}{h_n}\Big)\widetilde{X}\widetilde{X}^{\mathsf{T}}\Big].
\end{aligned}
$$

And we aim to show that with high probability

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

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

Step I. For set $\mathcal{J} \subset [p]$, consider $E_{\mathcal{J}} \cap \mathcal{S}_2^{p-1}$, where $E_{\mathcal{J}} = \text{span}\big\{e_j : j \in J\big\}$. Construct ϵ-net $\Pi_{\mathcal{J}}$, such that $\Pi_{\mathcal{J}} \subset E_{\mathcal{J}} \cap \mathcal{S}_2^{p-1}$ and $|\Pi_{\mathcal{J}}| \le (1+2\epsilon^{-1})^q$. The existence of $\Pi_{\mathcal{J}}$ can be guaranteed by Lemma 23 of Rudelson and Zhou (2013). Define $\Pi = \cup_{|\mathcal{J}|=q}\Pi_{\mathcal{J}}$, then for $0 < \epsilon < 1$ to be determined later, we have

$$
|\Pi| \le \Big(\frac{3}{\epsilon}\Big)^q\binom{p}{q} \le \Big(\frac{3ep}{q\epsilon}\Big)^q = \exp\Big\{q\log\Big(\frac{6ep}{q}\Big)\Big\}.
$$

For any $v \in E_{\mathcal{J}} \cap \mathcal{S}_2^{p-1}$, let $\Pi(v)$ be the closest point in ϵ-net $\Pi_{\mathcal{J}}$. Then we have

$$\frac{v - \Pi(v)}{\|v - \Pi(v)\|_2} \in E_{\mathcal{J}} \cap \mathcal{S}_2^{p-1}, \text{ and } \|v - \Pi(v)\|_2 \leq \epsilon.$$

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

$$\mathbb{P}\Big(\max_{v\in\Pi}\Big\{\Big|\binom{n}{2}^{-1}\sum_{i<j} g_v(D_i, D_j) - \mu_v\Big|\Big\} \geq \theta\Big),$$

for some $\theta > 0$, where

$$g_v(D_i, D_j) = \frac{1}{h_n} K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)(\widetilde{X}_{ij}^{\mathsf{T}} v)^2, \text{ and } \mu_v = \mathbb{E}[g_v(D_i, D_j)].$$

Also, denote $f_v(D_i) = \mathbb{E}\big[g_v(D_i, D_j)\big|D_i\big]$. Observe that

$$\begin{aligned}
&\Big|\binom{n}{2}^{-1}\sum_{i<j} g_v(D_i, D_j) - \mu_v\Big| \\
\leq &\Big|\binom{n}{2}^{-1}\sum_{i<j}\big\{g_v(D_i, D_j) - f_v(D_i) - f_v(D_j) + \mu_v\big\}\Big| + \Big|\frac{2}{n}\sum_{i=1}^{n}\big\{f_v(D_i) - \mu_v\big\}\Big|.
\end{aligned}$$

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

Step II.1. We bound

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}\big\{f_v(D_i) - \mu_v\big\}\Big| \geq t\Big), \tag{A4.21}$$

for $t > 0$ to be determined, and for each $v \in E_{\mathcal{J}} \cap \mathcal{S}_2^{p-1}$. Apply Lemma A3.3 with conditions of lemma satisfied by Assumptions 7 (Lemma A4.15) and 8 (Lemma A4.16), and we have

$$|f_v(D_i) - f_1(D_i)| \leq |MM_K h_n f_2(D_i)|, \tag{A4.22}$$

where $f_1(D_i) = \mathbb{E}\big[(\widetilde{X}_{ij}^{\mathsf{T}} v)^2\big|\widetilde{W}_{ij} = 0, D_i\big] f_W(W_i)$, and $f_2(D_i) = \mathbb{E}\big[(\widetilde{X}_{ij}^{\mathsf{T}} v)^2\big|X_i\big]$. Also, we have

$$|\mu_v - \mu_1| \leq |MM_K h_n \mu_2|, \tag{A4.23}$$

where $\mu_1 = \mathbb{E}[(\widetilde{X}_{ij}^{\mathsf{T}} v)^2|\widetilde{W}_{ij} = 0] f_{\widetilde{W}}(0)$, and $\mu_2 = \mathbb{E}[f_2(D_i)] = \mathbb{E}[(\widetilde{X}_{ij}^{T} v)^2]$. And we bound (A4.21) as below. We have

$$
\begin{aligned}
&\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{n}\big\{f_v(D_i)-\mu_v\big\}\geq t\Big)\\
=&\mathbb{P}\big(e^{a\left\{\sum_{i=1}^{n} f_v(D_i)-\mu_v\right\}}\geq e^{nat}\big)\\
\leq& e^{-nat}\cdot\mathbb{E}\big[e^{a\left\{\sum_{i=1}^{n} f_v(D_i)-\mu_v\right\}}\big]\\
\leq& e^{-nat}\cdot\mathbb{E}\big[e^{a\left(\sum_{i=1}^{n}\left[\{f_1(D_i)-\mu_1\}+MM_Kh_n\{f_2(D_i)-\mu_2\}\right]\right)}\big]\cdot e^{2MM_Knh_n\mu_2 a}\\
\leq& e^{-nat}\cdot\mathbb{E}\big[e^{2a\sum_{i=1}^{n}\{f_1(D_i)-\mu_1\}}\big]^{1/2}\cdot\mathbb{E}\big[e^{2MM_KC_0a\sum_{i=1}^{n}\{f_2(D_i)-\mu_2\}}\big]^{1/2}\cdot e^{4\kappa_x^2MM_Knh_na}\\
\leq& e^{-ant}\cdot\mathbb{E}\big[e^{2Ma\sum_{i1}^{n}\left|\mathbb{E}[(\widetilde{X}_{ij}^{\intercal}v)^2|\widetilde{W}_{ij}=0,D_i]-\mathbb{E}[(\widetilde{X}_{ij}^{\intercal}v)^2|\widetilde{W}_{ij}=0]\right|}\big]^{1/2}\cdot\mathbb{E}\big[e^{2a\cdot2\kappa_x^2\sum_{i=1}^{n}|f_W(W_i)-\mathbb{E}[f_W(W_i)]|}\big]^{1/2}\\
&\mathbb{E}\big[e^{2MM_KC_0a\sum_{i=1}^{n}\{f_2(D_i)-\mu_2\}}\big]^{1/2}\cdot e^{4\kappa_x^2MM_Knh_na}\\
\leq& e^{-ant}\cdot\mathbb{E}\big[e^{2aM\sum_{i=1}^{n}\left[\{(\widetilde{X}_i-\widetilde{X}_i')^{\intercal}v\}^2-\mathbb{E}[(\widetilde{X}_{ij}^{\intercal}v)^2|\widetilde{W}_{ij}=0]\right]}\big|W_i=W_i'\big]^{1/2}\cdot\big(e^{2a^2\kappa_x^4M^2n}\big)^{1/2}\\
&\mathbb{E}\big[e^{2MM_KC_0a\sum_{i=1}^{n}\left|\{(X_i-X_{i'})^{\intercal}v\}^2-\mu_2\right|}\big]^{1/2}\cdot e^{4\kappa_x^2MM_Knh_na}\\
\leq& e^{-ant}\cdot e^{2M^2\kappa_x^4a^2n}\cdot e^{M^2\kappa_x^4a^2n}\cdot e^{2M^2M_K^2C_0^2\kappa_x^4a^2n}\cdot e^{4MM_K\kappa_x^2nh_na},
\end{aligned}
$$

for $0<a\leq(4M\kappa_x^2)^{-1}$, 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 $\mu_2\leq2\kappa_x^2$ (Assumption 11, Lemma A4.17, and Lemma A4.18). The fourth inequality is by noting that $f_{\widetilde{W}}(0)=\mathbb{E}[f_W(W_i)]$, and applying the following inequality

$$|V_1V_2-\mathbb{E}[V_1]\mathbb{E}[V_2]|\leq|V_1-E[V_1]|\cdot|V_2|+|\mathbb{E}[V_1]|\cdot|V_2-\mathbb{E}[V_2]|,$$

where $V_1=\mathbb{E}[(\widetilde{X}_{ij}^{\intercal}v)^2|\widetilde{W}_{ij}=0,D_i]$, $|\mathbb{E}[V_1]|\leq2\kappa_x^2$ by Assumption 11, Lemma A4.17, and Lemma A4.18, and $V_2=f_W(W_i)\in[0,M]$. For the fifth inequality, the second component in product is bounded due to Jensen's inequality, where (X_i',W_i'), $i=1,\ldots,n$ are independent copies of (X_i,W_i); the third is bounded because $f_W(W_i)\in[0,M]$ and $\mathbb{E}[(\widetilde{X}_{ij}^{\intercal}v)^2|\widetilde{W}_{ij}=0]\leq2\kappa_x^2$ 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 $a=(1\wedge t)\cdot(2a_1)^{-1}$, and $h_n\leq t\cdot(4a_2)^{-1}$, where $a_1=(2M^2\kappa_x^4+2M^2M_K^2C_0^2\kappa_x^4+M^2\kappa_x^4)\vee 2M\kappa_x^2$ and $a_2=4MM_K\kappa_x^2$. Then we further have

$$\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{n}\big\{f_v(D_i)-\mu_v\big\}\geq t\Big)\leq\exp\Big\{\frac{-n(t^2\wedge t)}{8a_1}\Big\}.$$

By the same argument, we have

$$\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{n}\big\{f_v(D_i)-\mu_v\big\}\leq -t\Big)\leq\exp\Big\{\frac{-n(t^2\wedge t)}{8a_1}\Big\}.$$

We take $t=\theta/4$, and have

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}\big\{f_v(D_i)-\mu_v\big\}\Big|\geq\frac{\theta}{4}\Big)\leq2\exp\Big\{\frac{-n(\theta^2\wedge4\theta)}{128a_1}\Big\}. \tag{A4.24}$$

Step II.2. Observe that

$$\Big|\binom{n}{2}^{-1}\sum_{i<j}\big\{g_v(D_i,D_j)-f_v(D_i)-f_v(D_j)+\mu_v\big\}\Big|\le\binom{n}{2}^{-1}s\max_{k,l}\big\{|\sum_{i<j}\widetilde{\varphi}_{kl}(D_i,D_j)|\big\},$$

where

$$\begin{aligned}\widetilde{\varphi}_{kl}(D_i,D_j)=&\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}-\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\big|D_i\Big]\\&-\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\big|D_j\Big]+\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\Big].\end{aligned}$$

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

Apply truncation $|X_{ik}-\mathbb{E}[X_{ik}]|\le\tau_n/2$ for each $i\in[n]$, $k\in[p]$, and $\tau_n=\sqrt{6}(2+c)^{\frac12}\kappa_x\{\log(np)\}^{\frac12}$, for positive absolute constant c. Define events

$$\mathcal{A}_i=\big\{|X_{ik}-\mathbb{E}[X_{ik}]|\le\frac{\tau_n}{2},\,k\in[p]\big\},\quad\mathcal{A}_{[n]}=\big\{|X_{ik}-\mathbb{E}[X_{ik}]|\le\frac{\tau_n}{2},\,i\in[n],\,k\in[p]\big\}.$$

Consider truncated U-statistic $\sum_{i<j}\varphi_{kl}(D_i,D_j)$, where

$$\begin{aligned}\varphi_{kl}(D_i,D_j)=&\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\,\mathbb{1}(\mathcal{A}_i\cap\mathcal{A}_j)-\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\big|D_i\Big]\,\mathbb{1}(\mathcal{A}_i)\\&-\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\big|D_j\Big]\,\mathbb{1}(\mathcal{A}_j)+\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\Big].\end{aligned}$$

First, we bound $\big|\mathbb{E}[\varphi_{kl}(D_i,D_j)]\big|$. We have

$$\begin{aligned}&\big|\mathbb{E}[\varphi_{kl}(D_i,D_j)]\big|\\&=\Big|\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\,\mathbb{1}(\mathcal{A}_i^{\mathsf{c}}\cup\mathcal{A}_j^{\mathsf{c}})\Big]-2\mathbb{E}\Big[\mathbb{E}\Big\{\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\big|D_i\Big\}\,\mathbb{1}(\mathcal{A}_i^{\mathsf{c}})\Big]\Big|\\&\le\Big|\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\,\mathbb{1}(\mathcal{A}_i^{\mathsf{c}}\cup\mathcal{A}_j^{\mathsf{c}})\Big]\Big|+2\Big|\mathbb{E}\Big[\mathbb{E}\Big\{\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\big|D_i\Big\}\,\mathbb{1}(\mathcal{A}_i^{\mathsf{c}})\Big]\Big|.\end{aligned}\tag{A4.25}$$

We have

$$\begin{aligned}\Big|\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\,\mathbb{1}(\mathcal{A}_i^{\mathsf{c}}\cup\mathcal{A}_j^{\mathsf{c}})\Big]\Big|\le&M_K\frac{1}{h_n}\mathbb{E}[\widetilde{X}_{ijk}^2\widetilde{X}_{ijl}^2]^{1/2}\mathbb{P}(\mathcal{A}_i^{\mathsf{c}}\cup\mathcal{A}_j^{\mathsf{c}})^{1/2}\\\le&M_K\frac{1}{h_n}\mathbb{E}[\widetilde{X}_{ijk}^4]^{1/4}\mathbb{E}[\widetilde{X}_{ijl}^4]^{1/4}\mathbb{P}(\mathcal{A}_i^{\mathsf{c}}\cup\mathcal{A}_j^{\mathsf{c}})^{1/2}\\\le&M_K\frac{1}{h_n}(12\kappa_x^4)^{1/2}\cdot(2p\frac{1}{n^3p^3})^{1/2}\\\le&\frac{2\sqrt{6}M_K\kappa_x^2}{K_1np\{\log(np)\}^{1/2}}\le\frac{\theta}{24q},\end{aligned}\tag{A4.26}$$

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

$$n\ge\frac{48\sqrt{6}M_K\kappa_x^2q}{K_1\theta\{\log(np)\}^{1/2}p}.$$

We also have

$$\begin{aligned}\Big|\mathbb{E}\Big[\mathbb{E}\Big\{\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\big|D_i\Big\}\,\mathbb{1}(\mathcal{A}_i^{\mathsf{c}})\Big]\Big| \le&\mathbb{E}\Big[\mathbb{E}\Big\{\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\big|D_i\Big\}^2\Big]^{1/2}\cdot\mathbb{P}(\mathcal{A}_i^{\mathsf{c}})^{1/2}\\ \le&\big\{24(M+M_KC_0)^2\kappa_x^4\big\}^{1/2}\cdot\frac{1}{n^{3/2}p}\\ \le&\frac{\theta}{48q},\end{aligned}\tag{A4.27}$$

where the first ineuqlity 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

$$n\ge\Big\{\frac{96\sqrt{6}(M+MM_KC_0)\kappa_x^2q}{\theta p}\Big\}^{2/3}.$$

Combining (A4.25), (A4.26), and (A4.27), we have

$$\big|\mathbb{E}[\varphi_{kl}(D_i,D_j)]\big|\le\frac{\theta}{12q},\tag{A4.28}$$

when we appropriately choose n bounded from below.

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

For bounding B_g, we have $B_g\le 4M_K\tau_n^2\cdot h_n^{-1}\le\{4\sqrt{6}(2+c)M_K\kappa_x^2\cdot K_1^{-1}\}\cdot\{n\log(np)\}^{1/2}$. For bounding B_f, we have

$$\begin{aligned}&\mathbb{E}\big[|\varphi_{kl}(D_i,D_j)|\big|D_j\big]\\ \le&\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)|\widetilde{X}_{ijk}\widetilde{X}_{ijl}|\,\mathbb{1}(\mathcal{A}_i\cap\mathcal{A}_j)\big|D_j\Big]+\mathbb{E}\Big[\mathbb{E}\Big\{\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)|\widetilde{X}_{ijk}\widetilde{X}_{ijl}|\big|D_i\Big\}\,\mathbb{1}(\mathcal{A}_i)\Big]\\ &+\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)|\widetilde{X}_{ijk}\widetilde{X}_{ijl}|\big|D_j\Big]\,\mathbb{1}(\mathcal{A}_j)+\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)|\widetilde{X}_{ijk}\widetilde{X}_{ijl}|\Big],\\ \le&\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)|\widetilde{X}_{ijk}\widetilde{X}_{ijl}|\,\mathbb{1}(\mathcal{A}_i\cap\mathcal{A}_j)\big|D_j\Big]+\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)|\widetilde{X}_{ijk}\widetilde{X}_{ijl}|\big|D_j\Big]\,\mathbb{1}(\mathcal{A}_j)\\ &+2\cdot\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)|\widetilde{X}_{ijk}\widetilde{X}_{ijl}|\Big].\end{aligned}\tag{A4.29}$$

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

$$\begin{aligned}&\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)|\widetilde{X}_{ijk}\widetilde{X}_{ijl}|\,\mathbb{1}(\mathcal{A}_i\cap\mathcal{A}_j)\big|D_i\Big]\\ \le&\tau_n^2t(M+MM_KC_0)=6(c+2)(M+MM_KC_0)\kappa_x^2\log(np).\end{aligned}\tag{A4.30}$$

Apply Lemma A3.3 on $\varphi=|\widetilde{X}_{ijk}\widetilde{X}_{ijl}|\big|$, with $M_1=M$ and $M_2=M_K$ as given by Assumptions

8 (Lemma A4.16) and 7 (Lemma A4.15), we have

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)|\widetilde{X}_{ijk}\widetilde{X}_{ijl}|\big|D_j\Big]\ \mathbb{I}(\mathcal{A}_j)\\
\le& M\cdot\mathbb{E}\big[|\widetilde{X}_{ijk}\widetilde{X}_{ijl}|\big|D_j,\widetilde{W}_{ij}=0\big]\ \mathbb{I}(\mathcal{A}_j)+MM_KC_0\mathbb{E}\big[|\widetilde{X}_{ijk}\widetilde{X}_{ijl}|\big|D_j\big]\ \mathbb{I}(\mathcal{A}_j)\\
\le& M\mathbb{E}[\widetilde{X}_{ijk}^2\big|D_j,\widetilde{W}_{ij}=0]^{1/2}\mathbb{E}[\widetilde{X}_{ijl}^2\big|D_j,\widetilde{W}_{ij}=0]^{1/2}\ \mathbb{I}(\mathcal{A}_j)\\
&+MM_KC_0\mathbb{E}[\widetilde{X}_{ijk}^2\big|D_j]^{1/2}\mathbb{E}[\widetilde{X}_{ijl}^2\big|D_j]^{1/2}\ \mathbb{I}(\mathcal{A}_j)\\
\le&(1.5c+4)\cdot(M+MM_KC_0)\cdot\kappa_x^2\log(np),
\end{aligned}
\tag{A4.31}
$$

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

$$
\begin{aligned}
\mathbb{E}[\widetilde{X}_{ijk}^2\big|D_j]\ \mathbb{I}(\mathcal{A}_j)=&\big\{\mathbb{E}[(X_{ik}-\mathbb{E}[X_{ik}])^2]+(X_{ik}-\mathbb{E}[X_{jk}])^2\big\}\ \mathbb{I}(\mathcal{A}_j)\\
\le&\ \kappa_x^2+\tau_n^2/4\le(1.5c+4)\kappa_x^2\log(np),
\end{aligned}
$$

and based on an identical argument

$$
\mathbb{E}[\widetilde{X}_{ijk}^2\big|D_j,\widetilde{W}_{ij}=0]\ \mathbb{I}(\mathcal{A}_j)\le(1.5c+4)\kappa_x^2\log(np),
$$

for any $k\in[p]$.

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

$$
\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)|\widetilde{X}_{ijk}\widetilde{X}_{ijl}|\Big]\le 2(M+MM_KC_0)\kappa_x^2
\tag{A4.32}
$$

Combining (A4.29)-(A4.32), we have $B_f\le(20+7.5c)\cdot(M+MM_KC_0)\cdot\kappa_x^2\cdot\log(np)$.

For bounding $\mathbb{E}\big[\mathbb{E}\{\varphi_{kl}(D_i,D_j)\big|D_j\}^2\big]$, we observe that

$$
\begin{aligned}
\varphi_{kl}(D_i,D_j)=&\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\ \mathbb{I}(\mathcal{A}_i\cap\mathcal{A}_j)-\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\ \mathbb{I}(\mathcal{A}_i\cap\mathcal{A}_j)\big|D_i\Big]\\
&-\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\ \mathbb{I}(\mathcal{A}_i\cap\mathcal{A}_j)\big|D_j\Big]+\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\ \mathbb{I}(\mathcal{A}_i\cap\mathcal{A}_j)\Big]\\
&+\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\ \mathbb{I}(\mathcal{A}_i\cap\mathcal{A}_j)\big|D_i\Big]-\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\big|D_i\Big]\ \mathbb{I}(\mathcal{A}_i)\\
&+\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\ \mathbb{I}(\mathcal{A}_i\cap\mathcal{A}_j)\big|D_j\Big]-\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\big|D_j\Big]\ \mathbb{I}(\mathcal{A}_j)\\
&+\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\Big]-\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\ \mathbb{I}(\mathcal{A}_i\cap\mathcal{A}_j)\Big],
\end{aligned}
$$

which further implies that

$$
\begin{aligned}
&\big|\mathbb{E}\big[\varphi_{kl}(D_i,D_j)\big|D_j\big]\big|\\
\le&\Big|\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\ \mathbb{I}(\mathcal{A}_j^{\complement})\Big]\Big|+\Big|\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\ \mathbb{I}(\mathcal{A}_i^{\complement})\big|D_j\Big]\Big|\\
&+\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\ \mathbb{I}(\mathcal{A}_i^{\complement}\cup\mathcal{A}_j^{\complement})\Big]
\end{aligned}
$$

Therefore we have

$$
\begin{aligned}
&\mathbb{E}\big[\mathbb{E}\big\{\varphi_{kl}(D_i, D_j)\big|D_j\big\}^2\big] \\
\leq&\frac{3}{h_n^2}\{\mathbb{E}[\widetilde{X}_{ijk}\widetilde{X}_{ijl}\,\mathbb{1}(\mathcal{A}_j^{\mathsf{c}})]^2 + \mathbb{E}[\widetilde{X}_{ijk}\widetilde{X}_{ijl}\,\mathbb{1}(\mathcal{A}_i^{\mathsf{c}})]^2 + \mathbb{E}[\widetilde{X}_{ijk}\widetilde{X}_{ijl}\,\mathbb{1}(\mathcal{A}_i^{\mathsf{c}}\cup\mathcal{A}_j^{\mathsf{c}})]^2\} \\
\leq&\frac{3n}{K_1^2\log(np)}\{2\mathbb{E}[\widetilde{X}_{ijk}^4]^{1/2}\mathbb{E}[\widetilde{X}_{ijl}^4]^{1/2}\mathbb{P}(\mathcal{A}_j^{\mathsf{c}}) + \mathbb{E}[\widetilde{X}_{ijk}^4]^{1/2}\mathbb{E}[\widetilde{X}_{ijl}^4]\mathbb{P}(\mathcal{A}_i^{\mathsf{c}}\cup\mathcal{A}_j^{\mathsf{c}})\} \\
\leq&\frac{3n}{K_1^2\log(np)}(2\cdot 12\kappa_x^4\frac{1}{n^3p^2} + 12\kappa_x^4\frac{2}{n^3p^2}) \leq \frac{1}{n},
\end{aligned}
$$

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 $\widetilde{X}_{ij}$, and last holds true when we have

$$
n \geq \frac{144\kappa_x^4}{K_1^2p^2\log(np)}.
$$

For bounding σ^2, apply Lemma A3.2 on $Z = \widetilde{X}_{ijk}^2\widetilde{X}_{ijl}^2$ with $M_1 = M$ and $M_2 = M_K$ as given by Assumptions 8 (Lemma A4.16) and 7 (Lemma A4.15), we have

$$
\begin{aligned}
\sigma^2 &\leq \frac{16M_K}{h_n}\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}^2\widetilde{X}_{ijl}^2\Big] \\
&\leq \frac{16M_K}{h_n}\big\{M\cdot\mathbb{E}\big[\widetilde{X}_{ijk}^2\widetilde{X}_{ijl}^2\big|\widetilde{W}_{ij}=0\big] + MM_KC_0\mathbb{E}\big[\widetilde{X}_{ijk}^2\widetilde{X}_{ijl}^2\big]\big\} \\
&\leq \frac{16M_K}{h_n}\Big\{M\mathbb{E}[\widetilde{X}_{ijk}^4\big|\widetilde{W}_{ij}=0]^{1/2}\mathbb{E}[\widetilde{X}_{ijl}^4\big|\widetilde{W}_{ij}=0]^{1/2} + MM_KC_0\mathbb{E}[\widetilde{X}_{ijk}^4]^{1/2}\mathbb{E}[\widetilde{X}_{ijl}^4]^{1/2}\Big\} \\
&\leq \frac{192M_K(M+MM_KC_0)\kappa_x^4}{K_1}\Big\{\frac{n}{\log(np)}\Big\}^{1/2},
\end{aligned}
$$

where the third inequality is by Cauchy-Schwarz, and the last is by subgaussianity of $\widetilde{X}$ and choice of h_n.

For bounding B^2, we have

$$
\begin{aligned}
B^2 =&n\sup_{D_j}\mathbb{E}\big[\varphi_{kl}^2(D_i, D_j)\big|D_j\big] \\
\leq&4M_Knh_n^{-1}\sup_{D_j}\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}^2\widetilde{X}_{ijl}^2\,\mathbb{1}(\mathcal{A}_i\cap\mathcal{A}_j)\big|D_j\Big] \\
&+ 4n\sup_{D_j}\mathbb{E}\Big[\mathbb{E}\Big\{\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\big|D_i\Big\}^2\,\mathbb{1}(\mathcal{A}_i)\Big] \\
&+ 4n\sup_{D_j}\mathbb{E}\Big[\mathbb{E}\Big\{\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\big|D_j\Big\}^2\,\mathbb{1}(\mathcal{A}_j)\Big] \\
&+ 4n\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{X}_{ijl}\Big]^2 \\
\leq&\frac{4M_KM_fn\tau_n^4}{h_n} + 192M_f^2\kappa_x^4n + 8M_f\kappa_x^2n \\
\leq&\{144(2+c)^2M_KM_f\kappa_x^4K_1^{-1} + 192M_f^2\kappa_x^4 + 8M_f\kappa_x^4\}\cdot\{n\log(np)\}^{3/2},
\end{aligned}
$$

where $M_f = M + MM_KC_0$.

We take

$$t = \binom{n}{2}\frac{\theta}{12q},$$
$$u = (2+c)\log p,$$

and require that

$$n > \max\Big\{\frac{48\sqrt{6}M_K\kappa_x^2 q}{K_1 p\{\log(np)\}^{1/2}}, \Big(\frac{96\sqrt{6}M_f\kappa_x^2 q}{\theta p}\Big)^{2/3}, \frac{144\kappa_x^4}{K_1^2p^2\log(np)},$$
$$\Big[\frac{192\sqrt{3}(2+c)^{1/2}C_1M_K^{1/2}M_f^{1/2}\kappa_x^2}{K_1^{1/2}\theta}\Big]^{4/3}\cdot q^{4/3}\{\log(np)\}^{1/3},$$
$$\Big[\frac{24(20+7.5c)(2+c)C_2M_f\kappa_x^2}{\theta}\Big]^{1/2}\cdot q^{1/2}\log(np),$$
$$\Big[\frac{24(c+2)^{\frac{3}{2}}C_3\{144(2+c)^2M_KM_f\kappa_x^4K_1^{-1}+192M_f^2\kappa_x^4+8M_f\kappa_x^4\}^{1/2}}{\theta}\Big]^{\frac{4}{5}}q^{\frac{3}{2}}\{\log(np)\}^{\frac{9}{5}},$$
$$\Big[\frac{96\sqrt{6}(2+c)^3C_4\kappa_x^2}{K_1\theta}\Big]^{2/3}q^{2/3}\{\log(np)\}^{5/3},$$
$$\frac{192(20+7.5c)(c+2)M_f\kappa_x^2}{\theta}q\{\log(np)\}^2, \frac{12q}{(20+7.5c)M_f\kappa_x^2\theta\log(np)}\Big\}\Big\} \tag{A4.33}$$

for some positive absolute constant c, and $C_1,\ldots,C_4$ as defined in (A3.2). Then by Lemma A3.4, we have

$$\mathbb{P}\Big(\Big|\binom{n}{2}^{-1}\sum_{i<j}\varphi_{kl}(D_i,D_j) - \mathbb{E}[\varphi_{kl}](D_i,D_j)\Big| \geq \frac{5\theta}{12q}\Big)$$
$$\leq 2\exp\{-(2+c)\log p\} + 2.77\exp\{-(2+c)\log p\}$$

Combined with (A4.25), the last display further implies that

$$\mathbb{P}\Big(\Big|\binom{n}{2}^{-1}\sum_{i<j}\widetilde{\varphi}_{kl}(D_i,D_j)\Big| \geq \frac{\theta}{2q}\Big)$$
$$\leq \mathbb{P}\Big(\Big|\binom{n}{2}^{-1}\sum_{i<j}\widetilde{\varphi}_{kl}(D_i,D_j)\Big| \geq \frac{\theta}{2q}\cap\mathcal{A}_{[n]}\Big|\Big) + \mathbb{P}(\mathcal{A}_{[n]}^{\mathsf{c}})$$
$$\leq \mathbb{P}\Big(\Big|\binom{n}{2}^{-1}\sum_{i<j}\varphi_{kl}(D_i,D_j) - \mathbb{E}[\varphi_{kl}(D_i,D_j)]\Big| \geq \frac{5\theta}{12m}\Big) + \mathbb{P}(\mathcal{A}_{[n]}^{\mathsf{c}})$$
$$\leq 2\exp\{-(2+c)\log p\} + 2.77\exp\{-(2+c)\log p\} + np\exp\{-(2+c)\log(np)\}$$
$$\leq 5.77\exp\{-(1+c)\log p\},$$

for positive absolute constant c.

Step II.3 Combining results of Step II.1, Step II.2 and Step I, when we have (A4.33), and that

$$n > \max\Big\{\frac{256\{(3M^2\kappa_x^2 + 2M^2M_K^2C_0^2\kappa_x^2) \vee 2M\}\kappa_x^2}{\theta^2 \wedge 4\theta} q\log\Big(\frac{3ep}{q\epsilon}\Big), \frac{4096K_1^2M^2M_K^2\kappa_x^2\log(np)}{\theta^2}\Big\},$$

we have

$$\mathbb{P}\Big(\max_{v\in\Pi}\Big\{\Big|\binom{n}{2}^{-1}\sum_{i<j} g_v(D_i, D_j) - \mu_v\Big|\Big\} \geq \theta\Big) \leq 5.77\exp\{-(c+1)\log p\} + 2\exp(-c'n),$$

where $c' = (\theta^2 \wedge 4\theta)/[256\{(3M^2\kappa_x^2 + 2M^2M_K^2C_0^2\kappa_x^2) \vee 2M\}\kappa_x^2]$.

Step III. Denote

$$\Gamma = \binom{n}{2}^{-1/2} X_{h_n} - \Sigma_{h_n}^{1/2}.$$

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 $E_{\mathcal{J}} \cap \mathcal{S}_2^{p-1}$ by approximation.

For any $v \in E_{\mathcal{J}} \cap \mathcal{S}_2^{p-1}$ for some $|\mathcal{J}| = q$, denote $v_0 = \Pi(v)$. We have

$$\|\Gamma v\|_2 \leq \|\Gamma\Pi(v)\|_2 + \|\Gamma\{v - \Pi(v)\}\|_2. \tag{A4.34}$$

Define $\|\Gamma\|_{2,E_{\mathcal{J}}} = \sup_{y\in E_{\mathcal{J}}\cap\mathcal{S}_2^{p-1}} \|\Gamma y\|_2$. Then by (A4.34), we have

$$\|\Gamma\|_{2,E_{\mathcal{J}}} \leq \theta^{1/2} + \epsilon\|\Gamma\|_{2,E_{\mathcal{J}}},$$

which further implies that

$$\|\Gamma\|_{2,E_{\mathcal{J}}}^2 \leq \frac{\theta}{(1-\epsilon)^2}.$$

Take $\epsilon = 1/2$, then we have

$$\|\Gamma\|_{2,E_{\mathcal{J}}}^2 \leq 4\theta.$$

We take $\theta' = 4\theta$. This completes the proof. ∎

A4.10 Proof of Lemma A3.4

Proof. Denote $\mu = \mathbb{E}\big[g(Z_1, Z_2)\big]$, $\widetilde{f}(z) = f(z) - \mu$, $\widetilde{g}(Z_i, Z_j) = g(Z_i, Z_j) - f(Z_i) - f(Z_j) + \mu$, and $D_n(\widetilde{g}) = \sum_{i<j}\widetilde{g}(Z_i, Z_j)$. Also, denote $\|\widetilde{g}\|_\infty = \widetilde{B}_g$, $\|\widetilde{f}\|_\infty = \widetilde{B}_f$, $\widetilde{\sigma}^2 = \mathbb{E}\big[\widetilde{g}(Z_1, Z_2)^2\big]$, and

$$\widetilde{B}^2 = n\sup_z \mathbb{E}\big[\widetilde{g}(Z, z)^2\big],$$

$$\widetilde{D} = \sup\Big\{\mathbb{E}\big[\sum_{i<j}|\widetilde{g}(Z_i, Z_j)|a_i(Z_i)b_j(Z_j)\big] : \mathbb{E}\big[\sum_{i=2}^n a_i(Z_i)^2\big] \leq 1, \mathbb{E}\big[\sum_{j=1}^{n-1} b_j(Z_j)^2\big] \leq 1\Big\}.$$

Hoeffding decomposition gives us

$$U_n(g) - \mathbb{E}[U_n(g)] = 2(n-1)\sum_{i=1}^{n} \widetilde{f}(Z_i) + D_n(\widetilde{g}),$$

where $D_n(\widetilde{g})$ is a degenerate U-statistic of bounded kernel. By Bernstein inequality, we have

$$\begin{aligned}
\mathbb{P}\Big(\big|\sum_{i=1}^{n} \widetilde{f}(Z_i)\big| \geq \frac{t}{2(n-1)}\Big) &\leq 2\exp\Big(\frac{-t^2/8(n-1)^2}{n\mathbb{E}\big[\widetilde{f}(Z_i)^2\big] + \widetilde{B}_f \cdot t/6(n-1)}\Big) \\
&\leq 2\exp\Big(\frac{-t^2/n^2}{8n\mathbb{E}\big[\widetilde{f}(Z_i)^2\big] + 2\widetilde{B}_f \cdot t/n}\Big),
\end{aligned} \tag{A4.35}$$

when $n \geq 3$. By Theorem 3.4 in Houdré and Reynaud-Bouret (2003), for any $u > 0$, we have

$$\mathbb{P}\big(|D_n(\widetilde{g})| \geq C_1 n\widetilde{\sigma}u^{1/2} + C_2\widetilde{D}u/4 + C_3\widetilde{B}u^{3/2} + C_4\widetilde{B}_g u^2/4\big) \leq C_5 e^{-u}, \tag{A4.36}$$

where positive absolute constants $C_1, \ldots, C_5$ are as defined in (A3.2). Combining (A4.35) and (A4.36), we have

$$\begin{aligned}
&\mathbb{P}\big(|U_n(g) - \mathbb{E}[U_n(g)]| \geq t + C_1 n\widetilde{\sigma}u^{1/2} + C_2\widetilde{D}u/4 + C_3\widetilde{B}u^{3/2} + C_4\widetilde{B}_g u^2\big) \\
\leq&\mathbb{P}\Big(\big|\sum_{i=1}^{n} \widetilde{f}(Z_i)\big| \geq \frac{t}{2(n-1)}\Big) + \mathbb{P}\big(|D_n(\widetilde{g})| \geq C_1 n\widetilde{\sigma}u^{1/2} + C_2\widetilde{D}u/4 + C_3\widetilde{B}u^{3/2} + C_4\widetilde{B}_g u^2/4\big) \\
\leq&2\exp\Big(\frac{-t^2/n^2}{8n\mathbb{E}\big[\widetilde{f}(X)^2\big] + 2\widetilde{B}_f \cdot t/n}\Big) + C_5 e^{-u}.
\end{aligned} \tag{A4.37}$$

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

By some algebra, we have

$$\mathbb{E}\big[\widetilde{g}(X_1, X_2)^2\big|X_2\big] \leq \mathbb{E}\big[g(X_1, X_2)^2\big|X_2\big],$$

which implies that

$$\begin{aligned}
\widetilde{\sigma}^2 &= \mathbb{E}\big[\widetilde{g}(X_1, X_2)^2\big] \\
&= \mathbb{E}\big[\mathbb{E}\big\{\widetilde{g}(X_1, X_2)^2\big|X_2\big\}\big] \\
&\leq \mathbb{E}\big[\mathbb{E}\big\{g(X_1, X_2)^2\big|X_2\big\}\big] \\
&= \mathbb{E}\big[g(X_1, X_2)^2\big] = \sigma^2,
\end{aligned}$$

and that

$$\begin{aligned}
\widetilde{B}^2 &\leq n\sup_{X_2}\mathbb{E}\big[\widetilde{g}(X_1, X_2)^2\big|X_2\big] \\
&\leq n\sup_{X_2}\mathbb{E}\big[g(X_1, X_2)^2\big|X_2\big] = B^2.
\end{aligned}$$

Meanwhile, we have

$$\mathbb{E}\big[|\widetilde{g}(X_i, X_j)|\big|X_j\big] \leq 4B_f.$$

By Hölder's inequality, and combining with the last display, we have

$$
\begin{aligned}
&\mathbb{E}\big[|\widetilde{g}(X_i,X_j)|a_i(X_i)b_j(X_j)\big]\\
=&\mathbb{E}\big[b_j(X_j)\mathbb{E}\big\{|\widetilde{g}(X_i,X_j)|a_i(X_i)\big|X_j\big\}\big]\\
\leq&\mathbb{E}\big[b_j(X_j)\mathbb{E}\big\{|\widetilde{g}(X_i,X_j)|\big|X_j\big\}^{1/2}\mathbb{E}\big\{|\widetilde{g}(X_i,X_j)|a_i(X_i)^2\big|X_j\big\}^{1/2}\big]\\
\leq&(4B_f)^{1/2}\mathbb{E}\big[b_j(X_j)\mathbb{E}\big\{|\widetilde{g}(X_i,X_j)|a_i(X_i)^2\big|X_j\big\}^{1/2}\big]\\
\leq&(4B_f)^{1/2}\mathbb{E}\big[b_j(X_j)^2\big]^{1/2}\mathbb{E}\big[|\widetilde{g}(X_i,X_j)|a_i(X_i)^2\big]^{1/2}\\
=&(4B_f)^{1/2}\mathbb{E}\big[b_j(X_j)^2\big]^{1/2}\mathbb{E}\big[a_i(X_i)^2\mathbb{E}\big\{|\widetilde{g}(X_i,X_j)|\big|X_i\big\}\big]^{1/2}\\
\leq&4B_f\mathbb{E}\big[a_i(X_i)^2\big]^{1/2}\mathbb{E}\big[b_j(X_j)^2\big]^{1/2}.
\end{aligned}
$$

Therefore, we further have

$$
\begin{aligned}
\widetilde{D}&\leq 4B_f\sum_{i=2}^{n}\sum_{j=1}^{i-1}\big\{\mathbb{E}\big[a_i(X_i)^2\big]^{1/2}\mathbb{E}\big[b_j(X_j)^2\big]^{1/2}\big\}\\
&\leq 4B_f\sum_{i=2}^{n}\sum_{j=1}^{i-1}\frac{1}{2}\big\{\mathbb{E}\big[a_i(X_i)^2\big]+\mathbb{E}\big[b_j(X_j)^2\big]\big\}\\
&\leq 4B_f.
\end{aligned}
$$

Combining these upper bounds on constants with (A4.37), we complete the proof. □

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

$$
\begin{aligned}
&\mathbb{P}\Big(\delta\widehat{L}_n(\Delta,h_n)\geq\frac{\kappa_\ell M_\ell}{4}\|\Delta\|_2^2\text{ for all }\Delta\in\big\{\Delta'\in\mathbb{R}^p:\|\Delta_{\mathcal{S}^c}\|_1\leq 3\|\Delta_{\mathcal{S}}\|_1\big\}\Big)\\
\geq&1-5.77\exp(-c\log p)-2\exp(-c'n),
\end{aligned}
$$

where $c>1$ is an absolute constant, and $c'=(\kappa_\ell^2M_\ell^2\wedge 64\kappa_\ell M_\ell)/[2^{16}\{(3M^2\kappa_x^2+2M^2M_K^2C_0^2\kappa_x^2)\vee 2M\}\kappa_x^2]$.

(2) Assume Assumption 16 holds, and that (A4.20) holds with $q=2305\{s+\zeta^2nh_n^{2\gamma}/\log(np)\}$ and $t=\kappa_\ell M_\ell/16$. Then we have

$$
\begin{aligned}
&\mathbb{P}\Big(\delta\widehat{L}_n(\Delta,h_n)\geq\frac{\kappa_\ell M_\ell}{4}\|\Delta\|_2^2\text{ for all }\Delta\in\mathcal{C}_{\widetilde{\mathcal{S}}_n'}\Big)\\
\geq&1-5.77\exp(-c\log p)-2\exp(-c'n),
\end{aligned}
$$

where $\mathcal{C}_{\widetilde{\mathcal{S}}_n'}=\big\{v\in\mathbb{R}^p:\|v_{\mathcal{J}^c}\|_1\leq 3\|v_{\mathcal{J}}\|_1\text{ for some }\mathcal{J}\subset[p]\text{ and }|\mathcal{J}|\leq s+\zeta^2nh_n^{2\gamma}/\log(np)\big\}$, $c>1$ is an absolute constant, and $c'=(\kappa_\ell^2M_\ell^2\wedge 64\kappa_\ell M_\ell)/[2^{16}\{(3M^2\kappa_x^2+2M^2M_K^2C_0^2\kappa_x^2)\vee 2M\}\kappa_x^2]$.

Proof. (1) Denote $\mathcal{C}_{\mathcal{S}}=\big\{v\in\mathbb{R}^p:\|v_{\mathcal{S}^c}\|_1\leq 3\|v_{\mathcal{S}}\|_1\big\}$. By Lemma 13 in Rudelson and Zhou (2013), $\mathcal{C}_{\mathcal{S}}\cap\mathcal{S}_2^{p-1}\subset 2\mathrm{conv}\Big(\cup_{|\mathcal{J}|\leq d}E_{\mathcal{J}}\cap\mathcal{S}_2^{p-1}\Big)$, where $\mathrm{conv}(\cdot)$ means convex hull of a set, $E_{\mathcal{J}}=\mathrm{span}\big\{e_j:$

$j \in \mathcal{J}\}$, and $d = 2305s$. Denote

$$\Sigma_{h_n} = \mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}}{h_n}\Big)\widetilde{X}\widetilde{X}^{\mathsf{T}}\Big],$$

$$\Gamma = \binom{n}{2}^{-1}\sum_{i<j}\Big\{\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ij}\widetilde{X}_{ij}^{\mathsf{T}}\Big\} - \Sigma_{h_n},$$

$$\Sigma_0 = \mathbb{E}\Big[\widetilde{X}\widetilde{X}^{\mathsf{T}}\big|\widetilde{W} = 0\Big]\cdot f_{\widetilde{W}}(0).$$

For any $v \in \mathcal{C}_{\mathcal{S}} \cap \mathcal{S}_2^{p-1}$, we have

$$\begin{aligned}|v^{\mathsf{T}}\Gamma v| &\le 4\max_{v' \in \mathrm{conv}(\cup_{|\mathcal{J}|\le d}E_{\mathcal{J}}\cap\mathcal{S}_2^{p-1})} v'^{\mathsf{T}}\Gamma v'\\ &= 4\max_{v' \in \cup_{|\mathcal{J}|\le d}E_{\mathcal{J}}\cap\mathcal{S}_2^{p-1}} v'^{\mathsf{T}}\Gamma v'\\ &= 4\|\Gamma\|_{2,d},\end{aligned}$$

where the second line is because maximum of $v'^{\mathsf{T}}\Gamma v'$ occurs at extreme points of set $\mathrm{conv}\Big(\cup_{|\mathcal{J}|\le d} E_{\mathcal{J}} \cap \mathcal{S}_2^{p-1}\Big)$. Apply Theorem A3.1 with $q = d = 2305s$ and $t = \kappa_\ell M_\ell/16$, when (A4.20) is satisfied, we have

$$|v^{\mathsf{T}}\Gamma v| \le \frac{\kappa_\ell M_\ell}{4} \tag{A4.38}$$

holds simultaneously for all $v \in \mathcal{C}_{\mathcal{S}} \cap \mathcal{S}_2^{p-1}$ with probability at least $1 - 5.77\exp(-c\log p) - 2\exp(-c'n)$, where $c > 1$ is some absolute constant and $c' = (\kappa_\ell^2 M_\ell^2 \wedge 64\kappa_\ell M_\ell)/[65536\{(2M^2\kappa_x^2 + 2M^2M_K^2C_0^2\kappa_x^2 + M^2\kappa_x^2) \vee 2M\}\kappa_x^2]$.

(A4.38) further implies that $\delta\widehat{L}_n(v, h_n) \ge v^{\mathsf{T}}\Sigma_{h_n}v - \kappa_\ell M_\ell/4$, where

$$\begin{aligned}v^{\mathsf{T}}\Sigma_{h_n}v &\ge v^{\mathsf{T}}\Sigma_0 v - MM_K\mathbb{E}\big[(\widetilde{X}^{\mathsf{T}}v)^2\big]h_n\\ &\ge \kappa_\ell M_\ell\|v\|_2^2 - MM_K\cdot 2\kappa_x^2\|v\|_2^2\cdot h_n\\ &\ge \kappa_\ell M_\ell\|v\|_2^2/2 = \kappa_\ell M_\ell/2.\end{aligned} \tag{A4.39}$$

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

(2) Using an identical argument as used in (1), replacing $\mathcal{C}_{\mathcal{S}}$ by set

$$\big\{v \in \mathbb{R}^p : \|v_{\mathcal{J}^c}\|_1 \le 3\|v_{\mathcal{J}}\|_1 \text{ for some } \mathcal{J} \subset [p] \text{ and } |\mathcal{J}| \le s + \zeta^2 n h_n^{2\gamma}/\log(np)\big\},$$

and using $d = 2305\{s + \zeta^2 n h_n^{2\gamma}/\log(np)\}$ instead, we complete the proof for (2). □

A4.12 Proof of Lemma 3.1

Lemma A4.10 (Lemma 3.1)**.** Assume $h_n \ge K_1\{\log(np)/n\}^{1/2}$ for positive absolute constant K_1, and assume $h_n \le C_0$ 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

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

Then under Assumptions 7, 8, 11, and 17, we have

$$\mathbb{P}\Big(\max_{k\in[p]}\big\{|U_k-\mathbb{E}[U_k]|\big\}\geq C\{\log(np)/n\}^{1/2}\Big)\leq 4.77\exp(-c\log p)+\exp(-c'\log n),$$

where $C=\sqrt{2}C_1M_K^{1/2}M_f^{1/2}M_u^{1/(2+\epsilon)}\kappa_x c^{1/2}K_1^{-1/2}+2C_2M_f(c+1/2)^{1/2}c+8M_fM_u^{1/(2+\epsilon)}\kappa_x c^{1/2}+2C_3M_K^{1/2}M_f^{1/2}(c+2)^{1/2}c^{3/2}K_1^{-1/2}+2C_4M_K(c+2)^{1/2}c^2K_1^{-1}$. Here $M_f=M+MM_KC_0$, and $C_1,\ldots,C_4$ are as defined in (A3.2).

Proof. We apply truncation on $\widetilde{X}_{ijk}$ and $\widetilde{u}_i$ at levels τ_n and $\theta_n/2$ respectively, and first focus on U-statistic

$$\widetilde{U}_k=\binom{n}{2}^{-1}\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{u}_{ij}\,\mathbb{1}(\mathcal{A}_{k,ij}\cap\mathcal{B}_i\cap\mathcal{B}_j),$$

where we denote events

$$\mathcal{A}_{k,ij}=\big\{|\widetilde{X}_{ijk}|\leq\tau_n\big\},\quad \mathcal{B}_i=\big\{|u_i-\mathbb{E}[u]|\leq\theta_n/2\big\}.$$

We also denote events

$$\mathcal{A}_{k,[n]}=\big\{|\widetilde{X}_{ijk}|\leq\tau_n,\ i<j\in[n]\big\},\quad \mathcal{B}_{[n]}=\big\{|u_i-\mathbb{E}[u]|\leq\theta_n/2, i\in[n]\big\}.$$

Denote

$$g(D_i,D_j)=\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{u}_{ij}\,\mathbb{1}(\mathcal{A}_{k,ij}\cap\mathcal{B}_i\cap\mathcal{B}_j),\ \text{ and }\ f(D_i)=\mathbb{E}\big[g(D_i,D_j)\big|D_i\big].$$

We complete the proof in two steps.

Step I. We bound B_g, B_f, $\mathbb{E}\big[f(D_2)^2\big]$, σ^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

$$B_f\leq\tau_n\theta_n\Big\|\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1-W_2}{h_n}\Big)\big|W_1\Big]\Big\|_\infty\leq M_f\tau_n\theta_n,$$

where $M_f=M+MM_KC_0$.

For bounding σ^2, we have

$$\begin{aligned}\sigma^2&=\mathbb{E}\big[g(D_1,D_2)^2\big]\\&\leq\frac{M_K}{h_n}\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\widetilde{X}_{ijk}\widetilde{u}_{ij}\Big]\\&\leq\frac{M_K}{h_n}\big(\mathbb{E}\big[\widetilde{X}_{ijk}^2\widetilde{u}_{ij}^2\big|\widetilde{W}_{ij}=0\big]\cdot M+MM_KC_0\mathbb{E}\big[\widetilde{X}_{ijk}^2\widetilde{u}_{ij}^2\big]\big)\\&\leq 2M_KM_fM_u^{2/(2+\epsilon)}\kappa_x^2/h_n,\end{aligned}$$

where the first inequality is due to $K(\cdot)\in[0,1]$, the second inequality is by applying Lemma A3.2 on $Z=\widetilde{X}_{ijk}\widetilde{u}_{ij}$ with lemma assumptions satisfied by Assumptions 7 and 8, and the last inequality is by Assumptions 11, 12, and independence of $\widetilde{X}_{ijk}$ and $\widetilde{u}_{ij}$.

For bounding $\mathbb{E}\big[f(D_2)^2\big]$, apply Lemma A3.3 on $\varphi=\widetilde{X}_{ijk}\widetilde{u}_{ij}\,\mathbb{1}(\mathcal{A}_{k,ij}\cap\mathcal{B}_i\cap\mathcal{B}_j)$, with lemma assumptions satisfied by Assumptions 7 and 8, and we have $|f(D_2)-f_1(D_2)|\leq MM_KC_0f_2(D_2)$, where

$$f_1(D_2) = \mathbb{E}\big[\widetilde{X}_{12k}\widetilde{u}_{12}\, \mathbb{1}(\mathcal{A}_{k,12} \cap \mathcal{B}_1 \cap \mathcal{B}_2)\big| W_1 = W_2, D_2\big] \cdot f_{W_1}(W_2),$$
$$f_2(D_2) = \mathbb{E}\big[|\widetilde{X}_{12k}\widetilde{u}_{12}|\, \mathbb{1}(\mathcal{A}_{k,12} \cap \mathcal{B}_1 \cap \mathcal{B}_2)\big| D_2\big].$$

We have, by Assumptions 11, 12, and independence of $\widetilde{X}_{ijk}$ and $\widetilde{u}_{ij}$,

$$\mathbb{E}\big[f_1(D_2)^2\big] \leq \mathbb{E}\big[\widetilde{X}_{12k}^2\widetilde{u}_{12}^2 \big| W_1 = W_2\big] M^2 \leq 2MM_u^{2/(2+\epsilon)}\kappa_x^2,$$
$$\mathbb{E}\big[f_2(D_2)^2\big] \leq \mathbb{E}[\widetilde{X}_{12k}^2\widetilde{u}_{12}^2] \leq 2M_u^{2/(2+\epsilon)}\kappa_x^2.$$

This further implies that

$$\mathbb{E}[f(D_2)^2] \leq 2\mathbb{E}[f_1(D_2)^2] + 2M^2M_K^2C_0^2\mathbb{E}[f_2(D_2)^2] \leq 4M_f^2M_u^{2/(2+\epsilon)}\kappa_x^2.$$

For bounding B^2, we have

$$\begin{aligned}
B^2 &= n \sup_{D_2} \mathbb{E}\big[g(D_1, D_2)^2 \big| D_2\big] \\
&\leq \frac{nM_K}{h_n} \sup_{D_2} \mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)(X_{1k} - X_{2k})^2(u_1 - u_2)^2\, \mathbb{1}(\mathcal{A}_{k,12} \cap \mathcal{B}_1 \cap \mathcal{B}_2)\big| D_2\Big] \\
&\leq M_K M_f \frac{n\tau_n^2\theta_n^2}{h_n}.
\end{aligned}$$

We take for some positive absolute constant $c > 1$,

$$\begin{aligned}
t &= 8M_fM_u^{1/(2+\epsilon)}\kappa_x c^{1/2} \cdot \binom{n}{2}\{\log(np)/n\}^{1/2}, \\
\tau_n &= \max\big\{c,\, 2\big\}^{1/2} \cdot \{\log(np)\}^{1/2}, \quad \theta_n = n^{\alpha},\ 0 < \alpha < 3/4, \\
c_u &= c\log p,
\end{aligned}$$

and we have that

$$n > \max\Big\{\big[\big\{16c^3(c+1)C_0^2M_u^{2/(2+\epsilon)}\kappa_x^2\big\}^{1/(3-2\alpha)} \vee 1\big] \cdot (\log p)^{2/(3-2\alpha)},\ \{log(np)\}^{5/(3-4\alpha)}\Big\}.$$

Then by Lemma A3.4, we have

$$\begin{aligned}
\mathbb{P}\Big\{\binom{n}{2}^{-1}\big|\widetilde{U}_k - \mathbb{E}[\widetilde{U}_k]\big| \geq A\{\log(np)/n\}^{1/2}\Big\} &\leq 2\exp(-c\log(np)) + 2.77\exp(-c\log p) \\
&\leq 4.77\exp(-c\log p),
\end{aligned}$$

where with $C_1, \ldots, C_4$ defined in (A3.2),

$$\begin{aligned}
A =& 2\sqrt{2}C_1M_K^{1/2}M_f^{1/2}M_u^{1/(2+\epsilon)}\kappa_x c^{1/2}K_1^{-1/2} + 2C_2M_f(c+1/2)^{1/2}c \\
&+ 2C_3M_K^{1/2}M_f^{1/2}(c+2)^{1/2}c^{3/2}K_1^{-1/2} + 2C_4M_K(c+2)^{1/2}c^2K_1^{-1} + 8M_fM_u^{1/(2+\epsilon)}\kappa_x c^{1/2}.
\end{aligned}$$

Step II. We have $\mathbb{E}[\widetilde{U}_k]=0$, and thus we have

$$
\begin{aligned}
&\mathbb{P}\big(\max_{k\in[p]}\{|U_k-\mathbb{E}[U_k]|\}\geq A\{\log(np)/n\}^{1/2}\big)\\
\leq&\mathbb{P}\big(\max_{k\in[p]}\{|U_k-\mathbb{E}[U_k]|\}\geq A\{\log(np)/n\}^{1/2}\cap\mathcal{B}_{[n]}\big)+\mathbb{P}(\mathcal{B}_{[n]}^{\mathsf{c}})\\
\leq&\sum_{k=1}^{p}\big\{\mathbb{P}\big(|U_k|>A\{\log(np)/n\}^{1/2}\cap\mathcal{A}_{k,[n]}\cap\mathcal{B}_{[n]}\big)+\mathbb{P}(\mathcal{A}_{k,[n]}^{\mathsf{c}})\big\}+\mathbb{P}(\mathcal{B}_{[n]}^{\mathsf{c}})\\
\leq&\sum_{k=1}^{p}\big\{\mathbb{P}\big(|\widetilde{U}|>A\{\log(np)/n\}^{1/2}\cap\mathcal{A}_{k,[n]}\cap\mathcal{B}_{[n]}\big)+\mathbb{P}(\mathcal{A}_{k,[n]}^{\mathsf{c}})\big\}+\mathbb{P}(\mathcal{B}_{[n]}^{\mathsf{c}})\\
\leq&4.77\exp(-c\log p+\log p)+n\frac{\mathbb{E}[|\widetilde{u}|^{2+\epsilon}]}{n^{\alpha(2+\epsilon)}}\\
\leq&4.77\exp(-c\log p+\log p)+\exp(-c'\log n).
\end{aligned}
$$

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. □

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

$$
\begin{aligned}
n>\max\Big\{&\big[\{16(c+2)^3(c+1)C_0^2M_u^{2/(2+\epsilon)}\kappa_x^2\}^{1/(3-2\xi)}\vee 1\big]\cdot(\log p)^{2/(3-2\xi)},\ \{\log(np)\}^{5/(3-4\xi)},\\
&64(c+2)^2(c+1)\{\log(np)\}^3/3,\ 3,\\
&\frac{48\sqrt{6}M_K\kappa_x^2q}{K_1p\{\log(np)\}^{1/2}},\Big(\frac{2^{10}\cdot 6\cdot\sqrt{6}M_f\kappa_x^2q}{\kappa_\ell M_\ell p}\Big)^{2/3},\frac{144\kappa_x^4}{K_1^2p^2\log(np)},\\
&\Big[\frac{2^{11}\cdot 6\cdot\sqrt{3}(2+c)^{1/2}C_1M_K^{1/2}M_f^{1/2}\kappa_x^2}{K_1^{1/2}\kappa_\ell M_\ell}\Big]^{4/3}\cdot q^{4/3}\{\log(np)\}^{1/3},\\
&\Big[\frac{2^8\cdot 6\cdot(20+7.5c)(2+c)C_2M_f\kappa_x^2}{\kappa_\ell M_\ell}\Big]^{1/2}\cdot q^{1/2}\log(np),\\
&\Big[\frac{2^8\cdot 6(c+2)^{3/2}C_3\{144(2+c)^2M_KM_f\kappa_x^4K_1^{-1}+192M_f^2\kappa_x^4+8M_f\kappa_x^4\}^{1/2}}{\kappa_\ell M_\ell}\Big]^{\frac{4}{5}}q^{\frac{4}{5}}\{\log(np)\}^{\frac{9}{5}},\\
&\Big[\frac{2^{10}\cdot 6\cdot\sqrt{6}(2+c)^3C_4\kappa_x^2}{K_1\kappa_\ell M_\ell}\Big]^{2/3}q^{2/3}\{\log(np)\}^{5/3},\\
&\frac{2^{11}\cdot 6\cdot(20+7.5c)(c+2)M_f\kappa_x^2}{\kappa_\ell M_\ell}q\{\log(np)\}^2,\ \frac{2^6\cdot 3q}{(20+7.5c)M_f\kappa_x^2\kappa_\ell M_\ell\log(np)},\\
&\frac{2^{20}\{(3M^2\kappa_x^2+2M^2M_K^2C_0^2\kappa_x^2)\vee 2M\}\kappa_x^2}{(\kappa_\ell M_\ell)^2\wedge(16\kappa_\ell M_\ell)^2}q\log\Big(\frac{6ep}{q}\Big),\\
&\frac{2^{24}K_1^2M^2M_K^2\kappa_x^2\log(np)}{(\kappa_\ell M_\ell)^2}\Big\},
\end{aligned}
\tag{A4.40}
$$

where q is to be determined in specific cases. Denote $M_f = M + MM_K C_0$, and $C_1, \ldots, C_4$ are as defined in (A3.2). Also denote c to be some positive absolute constant, and

$$\begin{aligned} A' =& \sqrt{2} C_1 M_K^{1/2} M_f^{1/2} M_u^{1/(2+\epsilon)} \kappa_x c^{1/2} K_1^{-1/2} + 2C_2 M_f (c+1/2)^{1/2} c + \\ & 2C_3 M_K^{1/2} M_f^{1/2} (c+2)^{1/2} c^{3/2} K_1^{-1/2} + 2C_4 M_K (c+2)^{1/2} c^2 K_1^{-1} + 8 M_f M_u^{1/(2+\epsilon)} \kappa_x c^{1/2}, \\ c'' =& (\kappa_\ell^2 M_\ell^2 \wedge 64 \kappa_\ell M_\ell)/[2^{16}\{(3M^2\kappa_x^2 + 2M^2 M_K^2 C_0^2 \kappa_x^2) \vee 2M\}\kappa_x^2]. \end{aligned}$$

Theorem A4.11 (Corollary A2.1(1))**.** Assume $\lambda_n \geq 4(A+A')\{\log(np)/n\}^{1/2} + 8\kappa_x^2 M_f \zeta h_n$. Further assume (A4.40) holds with $q = 2305s$. Then under Assumptions 6-11, 14, 15, and 17, we have

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

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. □

Theorem A4.12. [Corollary A2.1(2)] Assume that $\lambda_n \geq 4(A + A)\{\log(np)/n\}^{1/2} + 8\kappa_x^2 M_f \zeta h_n^\gamma$. Further assume (A4.40) holds with $q = 2305s$. Then under Assumptions 6-11, 14, 15, and 17, we have

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

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 $\theta^* = \beta^*$, $\Gamma_0(\theta) = L_0(\beta)$, $\widehat{\Gamma}_n(\theta, h) = \widehat{L}_n(\beta, h)$, $\Gamma_h(\theta) = \mathbb{E}\widehat{L}_n(\beta, h)$, and take $\widetilde{\theta}_{h_n}^* = \beta^*$, 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. □

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 that $\lambda_n \geq 4(A' + A + M\eta_n)\{\log(np)/n\}^{1/2} + 8MM_K C^{1/2} \kappa_x^2 h_n$. 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

$$\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \leq \frac{288 s \lambda_n^2}{M_\ell^2 \kappa_\ell^2} + \frac{2s\log(np)}{n} + \left\{\frac{288 n \lambda_n^2}{M_\ell^2 \kappa_\ell^2 \log(np)} + 2\right\} \cdot \zeta^2 h_n^{2\gamma},$$

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 $\theta^* = \beta^*$, $\Gamma_0(\theta) = L_0(\beta)$, $\widehat{\Gamma}_n(\theta, h) = \widehat{L}_n(\beta, h)$, $\Gamma_h(\theta) = \mathbb{E}\widehat{L}_n(\beta, h)$, 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. □

Corollary A4.2 (Corollary A2.1(4)). Denote

$$\tau_1 = \sqrt{2}(2+c)^{1/2}\kappa_x K_1^{-1}(BM_K C_0^a + DM_K),$$
$$\tau_2 = \sqrt{2}(2+c)^{1/2}\kappa_x \{BM_K M(1+C_0)C_0^a + DM_f\},$$
$$\tau_3 = 4M_K^2 M^2 \cdot (BC_0^a + D)^2 \cdot (1+C_0^2) \cdot \kappa_x^2,$$
$$\tau_4 = \left\{4B^2 MM_K \kappa_x^2 (1+C_0)C_0^{2a-\gamma_1} + 2D^2 \cdot (12M_f\kappa_x^4)^{1/2} \cdot E^{1/2} C_0^{-1/2-\gamma_1}\right\} \cdot M_K K_1^{\gamma_1},$$
$$\tau_5 = 4(2+c)\kappa_x^2 \{BMM_K(1+C_0)C_0^{2a} + D^2 M_f\} M_K K_1^{-1},$$

and

$$\begin{aligned} A'' =& 4\tau_3^{1/2}(1+c)^{1/2} + 2C_1\tau_4^{1/2}(1+c)^{1/2} + 2C_2\tau_2(1+c) + 2C_3\tau_5^{1/2}(1+c)^{3/2} \\ &+ 2C_4\tau_1(1+c)^2 + 4M_f \cdot (BC_0^a + D) \cdot (c+2)\kappa_x, \end{aligned}$$

where $\gamma_1 = \min\left\{2a-1, -1/2\right\}$. Consider lower bound on n,

$$n > \max\left\{64(c+2)^2(c+1)\tau_2^2\tau_3^{-1}\{\log(np)\}^4,\ \{\log(np)\}^{5/3}\right\}. \tag{A4.41}$$

Here, B, D, E and a are to be specified in different cases.

(1) Assume that g is (L,α)-Hölder for $\alpha \geq 1$, and g has bounded support when $\alpha > 1$. Suppose (A4.40) holds with $q = 2305s$, and that (A4.41) holds with $B = L_\alpha$, where L_α is the Lipschitz constant for g ($L_\alpha = L$ 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

$$\zeta = \max\left\{4 \cdot \Big(\frac{L_\alpha^2 MM_K + MM_K\mathbb{E}\widetilde{u}^2/2}{\kappa_\ell M_\ell}\Big)^{1/2},\ \frac{16\kappa_x(M + MM_K C_0^2)^{1/2} \cdot L_\alpha^2 MM_K}{\kappa_\ell M_\ell}\right\}.$$

Then under Assumptions 6-8, 9′, 10-11, 13, and 17, we have

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

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

$$\begin{aligned}\zeta = \max\Bigg\{& 4 \cdot \Big(\frac{M_g^2 MM_K C_0^{2\alpha-2\gamma} + M_d^2 M_a C_0^{1-2\gamma} + MM_K\mathbb{E}\widetilde{u}^2 C_0^{2-2\gamma}/2}{\kappa_\ell M_\ell}\Big)^{1/2}, \\ & \frac{16\kappa_x(M + MM_K C_0^2)^{1/2} \cdot (M_g^2 MM_K C_0^{2\alpha-2\gamma} + M_d^2 M_a C_0^{1-2\gamma})^{1/2}}{\kappa_\ell M_\ell}\Bigg\},\end{aligned}$$

$\gamma = \alpha$ if $M_d M_a = 0$, and $\gamma = \min\left\{\alpha,\ 1/2\right\}$ if otherwise. Then under Assumptions 6-8, 9′, 10-11, 13, and 17, we have

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

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

$q = 2305\{s + \zeta^2 n h_n^{2\gamma} / \log(np)\}$, and that (A4.41) holds with $B = M_g$, $D = M_d$, $E = M_a$ and $a = \alpha$. Further assume $\lambda_n \geq 4(A' + A'' + M\eta_n)\{\log(np)/n\}^{1/2} + 8MM_K C\kappa_x^2 h_n$, where

$$\zeta = \max\Big\{4 \cdot \Big(\frac{M_g^2 M M_K C_0^{2\alpha-2\gamma} + M_d^2 M_a C_0^{1-2\gamma} + MM_K \mathbb{E}\widetilde{u}^2 C_0^{2-2\gamma}/2}{\kappa_\ell M_\ell}\Big)^{1/2}, \frac{16\kappa_x(M + MM_K C_0^2)^{1/2} \cdot (M_g^2 M M_K C_0^{2\alpha-2\gamma} + M_d^2 M_a C_0^{1-2\gamma})^{1/2}}{\kappa_\ell M_\ell}\Big\},$$

$\gamma = \alpha$ if $M_d M_a = 0$, and $\gamma = \min\{\alpha, 1/2\}$ if otherwise. Then under Assumptions 6-8, 9′, 10-11, 13, and 17, we have

$$\|\widehat{\beta}_{h_n} - \beta^*\|_2^2 \leq \frac{288 s\lambda_n^2}{M_\ell^2 \kappa_\ell^2} + \frac{2s\log(np)}{n} + \Big\{\frac{288 n\lambda_n^2}{M_\ell^2\kappa_\ell^2 \log(np)} + 2\Big\} \cdot \zeta^2 h_n^{2\gamma},$$

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). ☐

Theorem A4.14 (Corollary A2.1(5))**.** Assume that (A4.40) holds with $q = 2305s$. Assume further that $n > 64(c+2)^2(c+1)\{\log(np)\}^4$ and $\lambda_n \geq 4(A' + A''')\{\log(np)/n\}^{1/2} + 4\sqrt{2}M_g M_K M\kappa_x(1 + C_0)h_n$, where

$$\begin{aligned} A''' =& 8MM_K M_g C_0(1+C_0)\kappa_x(1+c)^{1/2} + 2C_1 M_g M^{1/2} M_K^{3/2}\kappa_x^{1/2}(1+C_0)^{1/2} C_0^{5/4} K_1^{-1/4}(1+c)^{1/2} \\ &+ 2\sqrt{2}C_2 MM_K M_g(1+C_0)\kappa_x K_1(1+c)^{3/2} + 4C_3 MM_K^{3/2} M_g^{1/2}(1+C_0)^{1/2} C_0^{1/2}\kappa_x(1+c)^2 \\ &+ 2\sqrt{2}C_4 M_K M_g C_0\kappa_x K_1^{-1}(1+c)^{5/2} + 2\sqrt{2}MM_K M_g(1+C_0)C_0, \end{aligned}$$

Then under Assumptions 6-11, 4, and 17 we have

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

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 $\theta^* = \beta^*$, $\Gamma_0(\theta) = L_0(\beta)$, $\widehat{\Gamma}_n(\theta, h) = \widehat{L}_n(\beta, h)$, $\Gamma_h(\theta) = \mathbb{E}\widehat{L}_n(\beta, h)$, and take $\widetilde{\theta}^*_{h_n} = \beta^*$, 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. ☐

A4.14 Supporting lemmas

Lemma A4.15. Assumption 7 implies that, for any $0 < a < 3$ and $0 < b < 1$, we have

$$\int_{-\infty}^{+\infty} |w|^a K(w)\, dw \leq M_K \quad \text{and} \quad \sup_{w\in\mathbb{R}} |w|^b K(w) \leq M_K.$$

Proof of Lemma A4.15. For any $0 < a < 3$, we have

$$\int_{-\infty}^{+\infty} |w|^a K(w)\, dw \leq \Big\{\int_{-\infty}^{+\infty} |w|^3 K(w)\, dw\Big\}^{a/3} \leq M_K^{a/3} \leq M_K,$$

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 $\widetilde{X}$-measurable function $\psi(\cdot) : \mathbb{R}^p \to \mathbb{R}^m$ mapping to a m-dimensional real space, we have

$$\sup_{w,z} \Big\{ \Big| \frac{\partial f_{\widetilde{W}|\psi(\widetilde{X})}(w,z)}{\partial w} \Big|, \, f_{\widetilde{W}|\psi(\widetilde{X})}(w,z), \, \Big| \frac{\partial f_{\widetilde{W}}(w)}{\partial w} \Big|, \, f_{\widetilde{W}}(w) \Big\} \leq M. \tag{A4.42}$$

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

$$\sup_{w,x} \Big\{ \Big| \frac{\partial f_{\widetilde{W}|\widetilde{X}}(w,x)}{\partial w} \Big|, \, f_{\widetilde{W}|\widetilde{X}}(w,x) \Big\} \leq M.$$

We have

$$F_{\widetilde{W}|\widetilde{X}=x}(w) = \frac{\int\int F_{W_1|X_1=x_2+x}(w_2+w) \, \frac{dF_{X_1}(x')}{dx'}|_{x'=x_2+x} \, dF_{W_2|X_2=x_2}(w_2) \, dF_{X_2}(x_2)}{\int \frac{dF_{X_1}(x')}{dx'}|_{x'=x_2+x} \, dF_{X_2}(x_2)}.$$

By dominated convergence theorem, we have

$$f_{\widetilde{W}|\widetilde{X}}(w,x) = \frac{\int\int f_{W_1|X_1}(w_2+w, x_2+x) \, \frac{dF_{X_1}(x')}{dx'}|_{x'=x_2+x} \, dF_{W_2|X_2=x_2}(w_2) \, dF_{X_2}(x_2)}{\int \frac{dF_{X_1}(x')}{dx'}|_{x'=x_2+x} \, dF_{X_2}(x_2)} \leq M,$$

and

$$\Big| \frac{\partial f_{\widetilde{W}|\widetilde{X}}(w,x)}{\partial w} \Big| = \Big| \frac{\int\int \frac{\partial f_{W_1|X_1}(w_2+w, x_2+x)}{\partial w} \, \frac{dF_{X_1}(x')}{dx'}|_{x'=x_2+x} \, dF_{W_2|X_2=x_2}(w_2) \, dF_{X_2}(x_2)}{\int \frac{dF_{X_1}(x')}{dx'}|_{x'=x_2+x} \, dF_{X_2}(x_2)} \Big| \leq M.$$

Based on the same argument, we have

$$F_{\widetilde{W}}(w) = \int F_{\widetilde{W}|\widetilde{X}=x}(w) \, dF_{\widetilde{X}}(x),$$

which, by dominated convergence theorem, implies that

$$f_{\widetilde{W}}(w) = \int f_{\widetilde{W}|\widetilde{X}}(w,x) \, dF_{\widetilde{X}}(x) \leq M,$$

and

$$\Big| \frac{\partial f_{\widetilde{W}}(w)}{\partial w} \Big| = \Big| \int \frac{\partial f_{\widetilde{W}|\widetilde{X}}(w,x)}{\partial w} \, dF_{\widetilde{X}}(x) \Big| \leq M$$

Also, for any $\widetilde{X}$-measurable function $\psi(\cdot)$, we have

$$F_{\widetilde{W}|\psi(\widetilde{X})=z}(w) = \frac{\frac{\partial}{\partial v} \int \mathbb{I}\{\psi(x) \leq v\} F_{\widetilde{W}|\widetilde{X}=x}(w) \, dF_{\widetilde{X}}(x)|_{v=z}}{\frac{\partial}{\partial v} \int \mathbb{I}\{\psi(x) \leq v\} \, dF_{\widetilde{X}}(x)|_{v=z}}.$$

By dominated convergence theorem, we have

$$f_{\widetilde{W}|\psi(\widetilde{X})}(w,z) = \frac{\frac{\partial}{\partial v}\int \mathbb{I}\{\psi(x)\le v\} f_{\widetilde{W}|\widetilde{X}}(w,x)\, dF_{\widetilde{X}}(x)|_{v=z}}{\frac{\partial}{\partial v}\int \mathbb{I}\{\psi(x)\le v\}\, dF_{\widetilde{X}}(x)|_{v=z}} \le M,$$

and

$$\Big|\frac{\partial f_{\widetilde{W}|\psi(\widetilde{X})}(w,z)}{\partial w}\Big| = \Big|\frac{\frac{\partial}{\partial v}\int \mathbb{I}\{\psi(x)\le v\} \frac{\partial f_{\widetilde{W}|\widetilde{X}}(w,x)}{\partial w}\, dF_{\widetilde{X}}(x)|_{v=z}}{\frac{\partial}{\partial v}\int \mathbb{I}\{\psi(x)\le v\}\, dF_{\widetilde{X}}(x)|_{v=z}}\Big| \le M.$$

Therefore, Assumption 8 implies (A4.42) □

Lemma A4.17. Assumption 11 implies, conditional on $\widetilde{W}=0$ and unconditionally, $\langle \widetilde{X}, v\rangle$ is mean-zero subgaussian with parameter at most $2\kappa_x^2\|v\|_2^2$, for any $v\in\mathbb{R}^p$. Assumption 12 implies that $\widetilde{u}$ is mean-zero subgaussian with parameter at most $2\kappa_u^2$.

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

$$\mathbb{E}\big[e^{t\widetilde{X}^\intercal v}\big] = \mathbb{E}\big[e^{t\left(X_1^\intercal v - \mathbb{E}[X_1^\intercal v]\right)}\big]\cdot \mathbb{E}\big[e^{t\left(-X_2^\intercal v + \mathbb{E}[X_2^\intercal v]\right)}\big] \le e^{t^2\kappa_x^2\|v\|_2^2},$$

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, $\widetilde{X}^\intercal v$ is mean-zero subgaussian with parameter at most $2\kappa_x^2\|v\|_2^2$.

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

$$\begin{aligned}\mathbb{E}\big[e^{t\widetilde{X}^\intercal v}\big|\widetilde{W}=0\big] &= \mathbb{E}\big(\mathbb{E}\big[e^{t\widetilde{X}^\intercal v}\big|W_1=W_2,W_2\big]\big)\\ &= \mathbb{E}\big[\mathbb{E}\big\{e^{t\left(X_1^\intercal v - \mathbb{E}[X_1^\intercal v|W_1=W_2]\right)}\big|W_1=W_2\big\}\cdot\mathbb{E}\big\{e^{t\left(-X_2^\intercal v+\mathbb{E}[X_2^\intercal v|W_2]\right)}\big|W_2\big\}\big]\\ &\le e^{t^2\kappa_x^2\|v\|_2^2},\end{aligned}$$

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 $\widetilde{W}=0$, $\widetilde{X}^\intercal v$ is mean-zero subgaussian with parameter at most $2\kappa_x^2\|v\|_2^2$. 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 κ_v^2, we have $\mathbb{E}[V^2]\le\kappa_v^2$, $\mathbb{E}[V^4]\le 3\kappa_v^4$, $\mathbb{P}(V^2-\mathbb{E}[V^2]\le v)\ge 1-\exp\{-v/(2\kappa_v^2)\}$ for any $v\ge 2\kappa_v^2$, and that $\mathbb{E}[e^{sV^2-s\mathbb{E}[V^2]}]\le e^{2s^2\kappa_v^4}$ for $|s|\le(2\kappa_v^2)^{-1}$.

Lemma A4.19. Let Z be some subgaussian random variable, with parameter at most κ_z^2. Suppose $\kappa_z^2\le a/4$ for some $a>0$. Then we have

$$\int_a^\infty z\, dF_{Z^2}(z) \le (a+4\kappa_z^2)\exp\{-a/(4\kappa_z^2)\}.$$

Proof of Lemma A4.19. We have $F_{Z^2}(z)\ge\mathbb{P}(Z^2-\mathbb{E}[Z^2]\le z/2)\ge 1-\exp\{-z/(4\kappa_z^2)\}$ for any

$z \geq a \geq 4\kappa_z^2$ (Lemma A4.18). By integration by parts, we have

$$\begin{aligned}\int_a^\infty z\, dF_{Z^2}(z) &= \int_a^\infty (-z)\, d\big\{1 - F_{Z^2}(z)\big\} \\ &= (-z)\big\{1 - F_{Z^2}(z)\big\}\big|_a^\infty + \int_a^\infty 1 - F_{Z^2}(z)\, dz \\ &\leq a \exp\{-a/(4\kappa_z^2)\} + \int_a^\infty \exp\{-z/(4\kappa_z^2)\}\, dz \\ &= (a + 4\kappa_z^2)\exp\{-a/(4\kappa_z^2)\}.\end{aligned}$$

This completes the proof. □

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 $\lambda_n \geq 4(A + A') \cdot \{\log(np)/n\}^{1/2} + 4\sqrt{2}M_g M_K M \kappa_x (1 + C_0) h_n$. Here, A' is as specified in (A4.48), and A'' as in (A4.53). Suppose we have

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

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

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

Proof of Lemma A4.20. Denote

$$\begin{aligned}U_{1k} &= \binom{n}{2}^{-1} \sum_{i<j} \frac{1}{h_n} K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big) \widetilde{X}_{ijk} \widetilde{u}_{ij} \\ U_{2k} &= \binom{n}{2}^{-1} \sum_{i<j} \frac{1}{h_n} K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big) \widetilde{X}_{ijk} \big\{g(W_i) - g(W_j)\big\},\end{aligned}$$

and observe that

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

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

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

for positive absolute constant A and c, and when assuming $n > \max\big\{64(c+2)^2(c+1)\{\log(np)\}^3/3,\ 3\big\}$. 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

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

for positive constants A' and c, and when assuming $n > \max\big\{64(c+2)^3(c+1)\{\log(np)\}^4,\ \{\log(np)\}^{\frac{5}{3}}\big\}$. Here A' is as specified in (A4.53).

By independence of u and (X, W), we have

$$\mathbb{E}[U_{1k}] = 0. \tag{A4.46}$$

We also have

$$\begin{aligned}
|\mathbb{E}[U_{2k}]| \leq & M_g \mathbb{E}\Big[\frac{1}{h_n} K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big) \big|\widetilde{X}_{ijk}\widetilde{W}_{ij}\big|\Big] \\
= & M_g \int\int K(w)|xwh_n| f_{\widetilde{W}_{ij}|\widetilde{X}_{ijk}}(w,x)\, dw\, dF_{\widetilde{X}_{ijk}}(x) \\
= & M_g \int\int K(w)|xwh_n| \Big\{ f_{\widetilde{W}_{ij}|\widetilde{X}_{ijk}}(0,x) + \frac{\partial f_{\widetilde{W}_{ij}|\widetilde{X}_{ijk}}(w,x)}{\partial w}\Big|_{(twh_n,x)} \cdot wh_n \Big\}\, dw\, dF_{\widetilde{X}_{ijk}}(x) \\
\leq & M_g M_K M \mathbb{E}\big[|\widetilde{X}_{ijk}|\big|\widetilde{W}_{ij} = 0\big] h_n + M_g M_K M \mathbb{E}[|\widetilde{X}_{ijk}|] h_n^2 \\
\leq & \sqrt{2} M_g M_K M \kappa_x (1 + C_0) h_n,
\end{aligned} \tag{A4.47}$$

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

$$\begin{aligned}
&\mathbb{P}\Big\{\text{for any } k \in [p],\ \big|\nabla_k \widehat{L}_n(\beta^*, h_n)\big| \leq 2(A + A') \cdot \{\log(np)/n\}^{1/2} + 2\sqrt{2} M_g M_K M \kappa_x (1 + C_0) \cdot h_n \Big\} \\
\geq & 1 - 12.04 \exp(-c \log p),
\end{aligned}$$

for positive absolute constant c, and when we appropriately take n bounded from below. Thus we have completed the proof by noting that $\lambda_n \geq 4(A + A') \cdot \{\log(np)/n\}^{1/2} + 4\sqrt{2} M_g M_K M \kappa_x (1 + C_0) h_n$. □

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

$$\begin{aligned}
\mathbb{E}\Big[\frac{1}{h} K\Big(\frac{W}{h}\Big) Z\Big] &= \int\int K(w) z f_{W|Z}(wh, z)\, dw\, dF_Z(z) \\
&= \int\int K(w) z \Big\{ f_{W|Z}(0, z) + \frac{\partial f_{W|Z}(w, z)}{\partial w}\Big|_{t_{w,h} wh} wh \Big\}\, dw\, dF_Z(z),
\end{aligned}$$

which implies that

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

This completes the proof. □

Proof of Lemma A3.3. By Taylor's expansion, for some $t_{w,h} \in [0,1]$, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{h}K(\frac{W_1 - W_2}{h})\varphi(Z_1, Z_2)\big|W_2, Z_2\Big] \\
=&\int\int \frac{1}{h}K\Big(\frac{w - W_2}{h}\Big)\varphi(z, Z_2)f_{W_1|Z_1}(w, z)\, dw\, dF_{Z_1}(z) \\
=&\int\int K(w)\varphi(z, Z_2)f_{W_1|Z_1}(W_2 + wh, z)\, dw\, dF_{Z_1}(z) \\
=&\int\int K(w)\varphi(z, Z_2)\Big\{f_{W_1|Z_1}(W_2, z) + \frac{\partial f_{W_1|Z_1}(w, z)}{\partial w}\Big|_{W_2 + t_{w,h}wh} wh\Big\}\, dw\, dF_{Z_1}(z),
\end{aligned}
$$

which implies that

$$
\begin{aligned}
&\Big|\mathbb{E}\Big[\frac{1}{h}K\Big(\frac{W_1 - W_2}{h}\Big)\varphi(Z_1, Z_2)\big|W_2, Z_2\Big] - \mathbb{E}\big[\varphi(Z_1, Z_2)\big|W_2, Z_2, W_1 = W_2\big]f_{W_1}(W_2)\Big| \\
\leq& M_1 M_2 \mathbb{E}\big[|\varphi(Z_1, Z_2)|\big|Z_2\big]h.
\end{aligned}
$$

This completes the proof. □

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 $\max\big\{\int_{-\infty}^{+\infty} |w|K(w)\, dw, \sup_{w\in\mathbb{R}} K(w)\big\} \leq M_K$, 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 κ_x^2 and κ_v^2 respectively, for positive absolute constants κ_x and κ_v. Assume that there exists positive absolute constant M, such that

$$
\max\Big\{\Big|\frac{\partial f_{W|(X,V)}(w, x, v)}{\partial w}\Big|,\, f_{W|(X,V)}(w, x, v)\Big\} \leq M,
$$

for any $w, x, v \in R$ 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 $n > \max\big\{64(c+2)^2(c+1)\{\log(np)\}^3/3, 3\big\}$ for positive absolute constant c. Consider U-statistic

$$
U = \sum_{i<j}\Big\{\frac{1}{h_n}K\Big(\frac{W_i - W_j}{h_n}\Big)(X_i - X_j)(V_i - V_j)\Big\}.
$$

Then we have

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

where

$$
\begin{aligned}
C =&\{16\sqrt{3}(1+c)^{1/2}M_f + 4\sqrt{3}C_1(1+c)^{1/2}M_f^{1/2}K_1^{-1/2} + 8C_2(1+c) \\
&+ 8C_3(1+c)^{3/2}M_K^{1/2}M_f^{1/2}K_1^{-1/2} + 8C_4(1+c)^2 M_K K_1^{-1} + 8M_f(c+2)\}\kappa_x\kappa_v,
\end{aligned}
\tag{A4.48}
$$

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

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 $C_y^2\log(np)$, for some positive absolute constants C_x and C_v. Denote $\mathcal{A}_{[n]} = \big\{(X_i - X_j)^2 \leq C_x^2\log(np),\, (V_i - V_j)^2 \leq C_v\log(np),\ i, j \in [n],\ i < j\big\}$, and first focus on U-statistic

$$\widetilde{U} = \sum_{i<j} \Big[\frac{1}{h_n} K\Big(\frac{W_i - W_j}{h_n}\Big) Z_{ij} \, \mathbb{1}\{(X_i - X_j)^2 \le C_x^2 \log(np), \, (V_i - V_j)^2 \le C_v \log(np)\}\Big].$$

Denote

$$g(D_i, D_j) = \frac{1}{h_n} K\Big(\frac{W_i - W_j}{h_n}\Big) Z_{ij} \, \mathbb{1}\{(X_i - X_j)^2 \le C_x^2 \log(np), \, (V_i - V_j)^2 \le C_v \log(np)\},$$

and

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

Assume $h_n \ge K_1\{\log(np)/n\}^{1/2}$ for some positive absolute constant K_1. Denote $\widetilde{X} = X_1 - X_2$, $\widetilde{V} = V_1 - V_2$, and $\widetilde{W} = W_1 - W_2$. 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 $(X_i - X_j)^2 \le C_x^2 \log(np)$, $(V_i - V_j)^2 \le C_v \log(np)$ implies that $|Z_{ij}| \le C \log(np)$.

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

We have $B_g \le CM_K \log(np)/h_n \le (CM_K/K_1) \cdot \{n \log(np)\}^{1/2}$. For B_f, apply Lemma A3.3 on $\varphi = 1$ and with $M_1 = M$, $M_2 = M_K$, and we have

$$\begin{aligned} B_f &\le C \log(np) \cdot \mathbb{E}\Big[\frac{1}{h_n} K\Big(\frac{W_i - W_j}{h_n}\Big)\Big|W_j\Big] \\ &\le C \log(np)\{f_W(W_j) + MM_K C_0\} \\ &\le CM_f \log(np), \end{aligned}$$

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

For bounding $\mathbb{E}\big[f(D_2)^2\big]$, apply Lemma A3.3 on $\varphi = Z_{ij} \, \mathbb{1}\big\{(X_i - X_j)^2 \le C_x^2 \log(np), \, (V_i - V_j)^2 \le C_v \log(np)\big\}$ and with $M_1 = M$, $M_2 = M_K$, and then we have

$$|f(D_2) - f_1(D_2)| \le M_K M f_2(D_2) h_n,$$

where

$$\begin{aligned} f_1(D_2) &\le \mathbb{E}\big[Z_{12} \, \mathbb{1}(|Z_{12}| \le C \log(np))\big|W_1 = W_2, D_2\big] f_{W_1}(W_2) \\ f_2(D_2) &\le \mathbb{E}\big[|Z_{12}| \, \mathbb{1}(|Z_{12}| \le C \log(np))\big|D_2\big]. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \mathbb{E}\big[f(D_2)^2\big] &= \mathbb{E}\big[\{f(D_2) - f_1(D_2) + f_1(D_2)\}^2\big] \\ &\le 2M_K^2 M^2 C_0^2 \mathbb{E}\big[f_2(D_2)^2\big] + 2\mathbb{E}\big[f_1(D_2)^2\big], \end{aligned} \tag{A4.49}$$

and meanwhile,

$$\begin{aligned} &\mathbb{E}\big[f_1(D_2)^2\big] \le M^2 \mathbb{E}[Z_{12}^2] \le M^2 \mathbb{E}[\widetilde{X}^4]^{1/2} \mathbb{E}[\widetilde{V}^4]^{1/2} \le 12M^2 \kappa_x^2 \kappa_v^2, \text{ and} \\ &\mathbb{E}\big[f_2(D_2)^2\big] \le \mathbb{E}[Z_{12}^2] \le \mathbb{E}[\widetilde{X}^4]^{1/2} \mathbb{E}[\widetilde{V}^4]^{1/2} \le 12\kappa_x^2 \kappa_v^2., \end{aligned} \tag{A4.50}$$

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 $\mathbb{E}[\widetilde{X}^4] \le 12\kappa_x^2$, $\mathbb{E}[\widetilde{V}^4] \le 12\kappa_v^2$ (Lemma A4.18). Combining (A4.49) and (A4.50), we have

$$\mathbb{E}\big[f(D_2)^2\big] \le (M_k^2 M^2 C_0^2 + M^2) \cdot 24\kappa_x^2 \kappa_v^2 < 24M_f^2 \kappa_x^2 \kappa_v^2. \tag{A4.51}$$

For bounding σ^2, apply Lemma A3.2 on $Z = Z_{ij}^2$ and with $M_1 = M$, $M_2 = M_K$, and then we have

$$\begin{aligned}
\mathbb{E}\big[g(D_i,D_j)^2\big] &\le \frac{M_K}{h_n}\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_i-W_j}{h_n}\Big)Z_{ij}^2\Big]\\
&\le \frac{M_K}{h_n}\{\mathbb{E}[Z_{ij}^2|W_i=W_j]M + MM_KC_0\mathbb{E}[Z_{ij}^2]\}\\
&\le \frac{M_K}{h_n}\{\mathbb{E}[\widetilde{X}^4|\widetilde{W}=0]^{1/2}\mathbb{E}[\widetilde{V}^4|\widetilde{W}=0]^{1/2}M + MM_KC_0\mathbb{E}[\widetilde{X}^4]^{1/2}\mathbb{E}[\widetilde{V}^4]^{1/2}\}\\
&\le \frac{M_K}{h_n}\{12\kappa_x^2\kappa_v^2M + 12\kappa_x^2\kappa_v^2MM_KC_0\}\\
&\le \frac{12\kappa_x^2\kappa_v^2M_KM_f}{K_1}\Big(\frac{n}{\log(np)}\Big)^{1/2},
\end{aligned}$$

where the third inequality is by Cauchy-Schwarz inequality, and the fourth is due to subgaussianity of $\widetilde{X}$ and $\widetilde{V}$, both conditional on $\widetilde{W}=0$ and unconditionally.

For bounding B^2, we have

$$\begin{aligned}
B^2 &= n\sup_{D_2}\mathbb{E}\big[g(D_1,D_2)^2\big|D_2\big]\\
&\le \frac{nM_K}{h_n}\sup_{D_2}\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1-W_2}{h_n}\Big)Z_{12}^2\,\mathbb{1}\left\{|Z_{12}|\le C\log(np)\right\}\Big|D_2\Big]\\
&\le \frac{nM_K}{h_n}\big(C\log(np)\big)^2\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1-W_2}{h_n}\Big)\Big|D_2\Big]\\
&\le \frac{C^2M_fM_K}{K_1}\big\{n\log(np)\big\}^{3/2},
\end{aligned}$$

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

$$\begin{aligned}
C_x^2 = C_Z\cdot 2\kappa_x^2,\ C_v^2 &= C_Z\cdot 2\kappa_v^2,\ \text{for } C_Z\ge 4,\\
t &= C_t\cdot 16\sqrt{3}M_f\kappa_x\kappa_v\binom{n}{2}\{\log(np)/n\}^{1/2},\\
u &= C_u\log p,\ \text{for } C_u>1,
\end{aligned}$$

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

$$P\Big\{\binom{n}{2}^{-1}\big|\widetilde{U}-\mathbb{E}[\widetilde{U}]\big|\ge A_1\Big\{\log(np)/n\Big\}^{1/2}\Big\}\le 2\exp(-C_t^2\log(np))+C_5\exp(-C_u\log p),$$

where

$$A_1=(16\sqrt{3}C_tM_f+4\sqrt{3}C_1C_u^{\frac12}M_f^{\frac12}K_1^{-\frac12}+8C_2C_u+8C_3C_u^{3/2}M_K^{\frac12}M_f^{\frac12}K_1^{-\frac12}+8C_4C_u^2M_KK_1^{-1})\kappa_x\kappa_v.$$

Here, $C_1,\ldots,C_5$ are as defined in (A3.2).

Step II. We bound $|\mathbb{E}[\widetilde{U}]-\mathbb{E}[\widetilde{U}]|$, and complete the proof. We have

$$\begin{aligned}\binom{n}{2}^{-1}\big|\mathbb{E}[\widetilde{U}]-\mathbb{E}[U]\big| =&\Big|\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W_{ij}}}{h_n}\Big)Z_{ij}\,\mathbb{1}\{|Z_{ij}|>C\log(np)\}\Big]\Big|\\ \le&\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}}{h_n}\Big)\widetilde{X}^2\,\mathbb{1}\{\widetilde{X}^2>2C_Z\kappa_x^2\log(np)\}\Big]^{1/2}\times\\ &\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}}{h_n}\Big)\widetilde{V}^2\,\mathbb{1}\{\widetilde{V}^2>2C_Z\kappa_v^2\log(np)\}\Big]^{1/2},\end{aligned}$$

where

$$\begin{aligned}&\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}}{h_n}\Big)\widetilde{X}^2\,\mathbb{1}\{\widetilde{X}^2>2C_Z\kappa_x^2\log(np)\}\Big]\\ \le&\mathbb{E}[\widetilde{X}^2\,\mathbb{1}\{\widetilde{X}^2>2C_Z\kappa_x^2\log(np)\}\big|\widetilde{W}=0]M+MM_KC_0\mathbb{E}[\widetilde{X}^2\,\mathbb{1}\{\widetilde{X}^2>2C_Z\kappa_x^2\log(np)\}]\\ \le&M_f\{2C_Z\kappa_x^2\log(np)+8\kappa_x^2\}\exp\{-2C_Z\log(np)/(8\kappa_x^2)\}\le 4M_fC_Z\kappa_x^2\{\log(np)/np\}^{1/2},\end{aligned}$$

where the first inequality is by applying Lemma A3.2 on $Z=\widetilde{X}^2\,\mathbb{1}\{\widetilde{X}^2>2C_Z\kappa_x^2\log(np)\}$ 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 κ_x^2 (Lemma A4.18), both conditional on $W_i=W_j$ and unconditionally, and by applying Lemma A4.19 with $a=2C_Z\kappa_x^2\log(np)\ge 4\kappa_x^2$. By an identical argument, we have

$$\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W}}{h_n}\Big)\widetilde{X}^2\,\mathbb{1}\{\widetilde{X}^2>2C_Z\kappa_x^2\log(np)\}\Big]\le 4M_fC_Z\kappa_v^2\{\log(np)/np\}^{1/2}.$$

Combining the last three displays, we have

$$\binom{n}{2}^{-1}\big|\mathbb{E}[\widetilde{U}]-\mathbb{E}[U]\big| =\Big|\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{\widetilde{W_{ij}}}{h_n}\Big)Z_{ij}\,\mathbb{1}\{|Z_{ij}|>C\log(np)\}\Big]\Big|\le A_2\{\log(np)/np\}^{1/2},\quad\text{(A4.52)}$$

where $A_2=4M_fC_Z\kappa_x\kappa_v$.

We have

$$\begin{aligned}&\mathbb{P}\Big\{\binom{n}{2}^{-1}\big|U-\mathbb{E}[U]\big|\ge(A_1+A_2)\cdot\Big\{\frac{\log(np)}{n}\Big\}^{1/2}\Big\}\\ \le&\mathbb{P}\Big\{\binom{n}{2}^{-1}\big|U-\mathbb{E}[U]\big|\ge(A_1+A_2)\cdot\Big\{\frac{\log(np)}{n}\Big\}^{1/2}\cap\mathcal{A}_{[n]}\Big\}+\mathbb{P}(\mathcal{A}_{[n]}^{\mathsf{c}})\\ \le&\mathbb{P}\Big\{\binom{n}{2}^{-1}\big|\widetilde{U}-\mathbb{E}[U]\big|\ge(A_1+A_2)\cdot\Big\{\frac{\log(np)}{n}\Big\}^{1/2}\cap\mathcal{A}_{[n]}\Big\}+\mathbb{P}(\mathcal{A}_{[n]}^{\mathsf{c}})\\ \le&\mathbb{P}\Big\{\binom{n}{2}^{-1}\big|\widetilde{U}-\mathbb{E}[U]\big|\ge(A_1+A_2)\cdot\Big\{\frac{\log(np)}{n}\Big\}^{1/2}\Big\}+\mathbb{P}(\mathcal{A}_{[n]}^{\mathsf{c}})\\ \overset{(i)}{\le}&\mathbb{P}\Big\{\binom{n}{2}^{-1}\big|\widetilde{U}-\mathbb{E}[\widetilde{U}]\big|\ge A_1\cdot\Big\{\frac{\log(np)}{n}\Big\}^{1/2}\Big\}+\mathbb{P}(\mathcal{A}_{[n]}^{\mathsf{c}})\\ \le&2\exp(-C_t^2\log(np))+C_5\exp(-C_u\log p)+2n^2\exp(-C_Z\log(np)/2)\\ \le&2\exp(-C_t^2\log(np))+C_5\exp(-C_u\log p)+2\exp(-C_Z\log p/2),\end{aligned}$$

where (i) is by (A4.52). We take $C_t^2=C_u=c>1$, and $C_Z=\max\big\{2c,4\big\}\le 2c+2$, for positive absolute constant c. This completes the proof. $\square$

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

$$\max\Big\{\int_{-\infty}^{+\infty} |w|^{2\alpha+1} K(w)\, dw,\ \sup_w |w|^\alpha K(w)\Big\} \le M_K,$$

for positive absolute constant $M_K \ge 1$. Let $V_i = v(W_i)$ for function $v(\cdot)$, such that

$$|v(w_1) - v(w_2)| \le M_v |w_1 - w_2|^\alpha + M_d\, \mathbb{1}\big\{(w_1, w_2) \in A\big\},$$

for positive absolute constant M_v, M_d, $0 < \alpha \le 1$, and set A such that $(w_1, w_2) \in A$ implies $(w_2, w_1) \in A$, and that

$$\mathbb{E}\Big[\frac{1}{h_n} K\Big(\frac{\widetilde{W}_{ij}}{h_n}\Big)\, \mathbb{1}\big\{(W_i, W_j) \in A\big\}\Big] \le M_a h_n,$$

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 κ_x^2, for positive absolute constant κ_x. Assume that there exists positive absolute constant M such that

$$\max\Big\{\Big|\frac{\partial f_{W|X}(w, x)}{\partial w}\Big|,\ f_{W|X}(w, x)\Big\} \le M,$$

for any $w, x \in \mathbb{R}$ such that the densities are defined. Take $h_n \ge K_1\{\log(np)/n\}^{1/2}$ for positive absolute constant K_1, and assume that $h_n \le C_0$ for positive constant C_0. Suppose $n > \max\big\{64(c+2)^2(c+1)\tau_2^2\tau_3^{-1}\{\log(np)\}^4,\ \{\log(np)\}^{5/3}\big\}$, for positive absolute constant c. Consider U-statistics

$$U = \sum_{i<j} \Big\{\frac{1}{h_n} K\Big(\frac{W_i - W_j}{h_n}\Big)(X_i - X_j)(V_i - V_j)\Big\}.$$

Then there exists positive absolute constants C, such that

$$\mathbb{P}\Big\{\binom{n}{2}^{-1}\big|U - \mathbb{E}[U]\big| \ge C\{\log(np)/n\}^{1/2}\Big\} \le 5.27\exp\{-(c+1)\log p\},$$

where

$$\begin{aligned} C =& 4\tau_3^{1/2}(1+c)^{1/2} + 2C_1\tau_4^{1/2}(1+c)^{1/2} + 2C_2\tau_2(1+c) + 2C_3\tau_5^{1/2}(1+c)^{3/2} \\ &+ 2C_4\tau_1(1+c)^2 + 4(M + MM_KC_0)\cdot(M_vC_0^\alpha + M_d)\cdot(c+2)\kappa_x \end{aligned} \tag{A4.53}$$

Here $C_1, \ldots, C_5$ are as defined in (A3.2), and

$\tau_1 = \sqrt{2}(2+c)^{1/2}\kappa_x K_1^{-1}(M_vM_KC_0^\alpha + M_dM_K)$,

$\tau_2 = \sqrt{2}(2+c)^{1/2}\kappa_x\{M_vM_KM(1+C_0)C_0^\alpha + M_d(M + MM_KC_0)\}$,

$\tau_3 = 4M_K^2M^2\cdot(M_vC_0^\alpha + M_d)^2\cdot(1+C_0^2)\cdot\kappa_x^2$,

$\tau_4 = \big\{4M_v^2MM_K\kappa_x^2(1+C_0)C_0^{2\alpha-\gamma_1} + 2M_d^2\cdot(12M\kappa_x^4 + 12MM_KC_0\kappa_x^4)^{1/2}\cdot M_a^{1/2}C_0^{-1/2-\gamma_1}\big\}\cdot M_KK_1^{\gamma_1}$,

$\tau_5 = 4(2+c)\kappa_x^2\{M_vMM_K(1+C_0)C_0^{2\alpha} + M_d^2(M + MM_KC_0)\}M_KK_1^{-1}$,

where $\gamma_1 = \min\big\{2\alpha - 1, -1/2\big\}$.

Proof of Lemma A4.22. We apply truncation to $(X_i - X_j)^2$ at level $C\log(np)$ for some positive absolute constant C, and first focus on U-statistic

$$\widetilde{U} = \sum_{i<j} \Big[\frac{1}{h_n} K\Big(\frac{W_i - W_j}{h_n}\Big)(X_i - X_j)(V_i - V_j)\, \mathbb{1}\big\{(X_i - X_j)^2 \le C\log(np)\big\}\Big].$$

Denote

$$g(D_i, D_j) = \frac{1}{h_n} K\Big(\frac{W_i - W_j}{h_n}\Big)(X_i - X_j)(V_i - V_j)\, \mathbb{1}\big\{(X_i - X_j)^2 \le C\log(np)\big\},$$

and

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

Assume $h_n \ge K_1\{\log(np)/n\}^{1/2}$ 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, $\mathbb{E}\big[f(D_2)^2\big]$, σ^2, and B^2 as in Lemma A3.4, and apply Lemma A3.4.

For bounding B_g, we have

$$\begin{aligned} B_g &\le h_n^{-1}(C\log(np))^{1/2}\Big\|K\Big(\frac{W_i - W_j}{h_n}\Big)\big[M_v(W_i - W_j)^\alpha + M_d\, \mathbb{1}\big\{(W_i, W_j) \in A\big\}\big]\Big\|_\infty \\ &\le C^{1/2}K_1^{-1}(M_vM_KC_0^\alpha + M_dM_K)\cdot n^{1/2} = \tau_1 n^{1/2}, \end{aligned}$$

where the second inequality is by $|w|^\alpha K(w) \le M_K$ and $K(w) \le M_k$.

For bounding B_f, we have

$$\begin{aligned} B_f &\le (C\log(np))^{1/2}\Big\|\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_i - W_j}{h_n}\Big)(V_i - V_j)\big|D_j\Big]\Big\|_\infty \\ &\le (C\log(np))^{1/2}\Big\|\Big\{M_v\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_i - W_j}{h_n}\Big)|W_i - W_j|^\alpha\big|D_j\Big] \\ &\quad + M_d\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_i - W_j}{h_n}\Big)\, \mathbb{1}\big\{(W_i, W_j) \in A\big\}\big|D_j\Big]\Big\}\Big\|_\infty, \end{aligned}$$

where

$$\begin{aligned} &\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_i - W_j}{h_n}\Big)|W_i - W_j|^\alpha\big|D_j\Big] \\ =&\int\int K(w)|wh_n|^\alpha f_{W_1}(W_2 + wh_n)\, dw \\ =&\int\int K(w)|wh_n|^\alpha\Big\{f_{W_1}(W_2) + \frac{\partial f_{W_1}(w)}{\partial w}\Big|_{W_2+wh_n}\cdot wh_n\Big\}\, dw \\ \le& MM_K(1 + C_0)h_n^\alpha. \end{aligned}$$

Therefore $B_f \le C^{1/2}\big\{M_vM_KM(1 + C_0)C_0^\alpha + M_d(M + MM_KC_0)\big\}\cdot\{\log(np)\}^{1/2} = \tau_2\{\log(np)\}^{1/2}$.

For bounding $\mathbb{E}\big[f(D_2)^2\big]$, we have

$$\begin{aligned} |f(D_2)| \le& M_v\mathbb{E}\Big[\frac{|W_1 - W_2|^\alpha}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)|X_1 - X_2|\big|D_2\Big] \\ &+ M_d\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)|X_1 - X_2|\big|D_2\Big] \end{aligned} \tag{A4.54}$$

Apply Lemma A3.3 on $\varphi = |X_1 - X_2|$ and with $M_1 = M$, $M_2 = M_K$, we have

$$\begin{aligned} &\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)|X_1 - X_2|\big|D_2\Big] \\ &\le \mathbb{E}\big[|X_1 - X_2|\big|W_1 = W_2, D_2\big]f_{W_1}(W_2) + MM_KC_0\mathbb{E}\big[|X_1 - X_2|\big|D_2\big], \end{aligned} \tag{A4.55}$$

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

$$\begin{aligned} &\mathbb{E}\Big[\frac{|W_1 - W_2|^\alpha}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)|X_1 - X_2|\big|D_2\Big] \\ =&\int\int K(w)|x - X_2|\cdot|wh_n|^\alpha f_{W|X}(W_2 + wh_n, x)\,dw\,dF_X(x) \\ =&\int\int K(w)|x - X_2|\cdot|wh_n|^\alpha\Big\{f_{W|X}(W_2, x) + \frac{\partial f_{W|X}(w,x)}{\partial w}\Big|_{(W_2 + twh_n, x)}\cdot wh_n\Big\}\,dw\,dF_X(x) \\ \le& M_KMC_0^\alpha\cdot\mathbb{E}\big[|X_1 - X_2|\big|W_1 = W_2, D_2\big] + M_KMC_0^{\alpha+1}\cdot\mathbb{E}\big[|X_1 - X_2|\big|D_2\big]. \end{aligned} \tag{A4.56}$$

Combining (A4.54)-(A4.56), and by Jensen's inequality, we have

$$\begin{aligned} \mathbb{E}\big[f(D_2)^2\big] \le& \mathbb{E}\Big(\big\{(M_vM_KMC_0^\alpha + M_dM)\mathbb{E}\big[|X_1 - X_2|\big|W_1 = W_2, D_2\big] \\ &+ (M_vM_KMC_0^{\alpha+1} + M_dMM_KC_0)\mathbb{E}\big[|X_1 - X_2|\big|D_2\big]\big\}^2\Big) \\ \le& 2(M_vM_KMC_0^\alpha + M_dM)^2\mathbb{E}\big[\mathbb{E}\big\{|X_1 - X_2|\big|W_1 = W_2, D_2\big\}^2\big] \\ &+ 2(M_vM_KMC_0^{\alpha+1} + M_dMM_KC_0)^2\mathbb{E}\big[\mathbb{E}\big\{|X_1 - X_2|\big|D_2\big\}^2\big] \\ \le& 2M_K^2M^2(M_vC_0^\alpha + M_d)^2(1 + C_0^2)\mathbb{E}\big[(X_1 - X_2)^2\big] \\ \le& 4M_K^2M^2\cdot(M_vC_0^\alpha + M_d)^2\cdot(1 + C_0^2)\cdot\kappa_x^2 = \tau_3 \end{aligned}$$

For bounding σ^2, we have

$$\begin{aligned} \sigma^2 =& \mathbb{E}\big[g(D_1, D_2)^2\big] \\ \le& \mathbb{E}\Big[\frac{2M_v^2|W_1 - W_2|^{2\alpha} + 2M_d^2\,\mathbb{I}\,\{(W_1, W_2) \in A\}}{h_n^2}K^2\Big(\frac{W_1 - W_2}{h_n}\Big)(X_1 - X_2)^2\Big] \\ \le& \frac{2M_v^2M_K}{h_n}\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)|W_1 - W_2|^{2\alpha}(X_1 - X_2)^2\Big] \\ &+ \frac{2M_d^2M_K}{h_n}\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)\,\mathbb{I}\,\big\{(W_1, W_2) \in A\big\}(X_1 - X_2)^2\Big]. \end{aligned} \tag{A4.57}$$

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

$$\begin{aligned} &\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)|W_1 - W_2|^{2\alpha}(X_1 - X_2)^2\Big] \\ =&\int\int K(w)|wh_n|^{2\alpha}\cdot x^2\cdot f_{\widetilde{W}|\widetilde{X}}(wh_n, x)\,dw\,dF_{\widetilde{X}}(x) \\ =&\int\int K(w)|wh_n|^{2\alpha}\cdot x^2\cdot\Big\{f_{\widetilde{W}|\widetilde{X}}(0, x) + \frac{\partial f_{\widetilde{W}|\widetilde{X}}(w,x)}{\partial w}\Big|_{(twh_n, x)}\cdot wh_n\Big\}\,dw\,dF_{\widetilde{X}}(x) \\ \le& MM_Kh_n^{2\alpha}\mathbb{E}\big[\widetilde{X}^2\big|\widetilde{W} = 0\big] + MM_Kh_n^{2\alpha+1}\mathbb{E}\big[\widetilde{X}^2\big] \\ \le& 2MM_K\kappa_x^2(1 + C_0)h_n^{2\alpha}. \end{aligned} \tag{A4.58}$$

We also have

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1-W_2}{h_n}\Big)\,\mathbb{I}\left\{(W_1,W_2)\in A\right\}(X_1-X_2)^2\Big]\\
\le&\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1-W_2}{h_n}\Big)\,\mathbb{I}\left\{(W_1,W_2)\in A\right\}\Big]^{1/2}\cdot\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1-W_2}{h_n}\Big)(X_1-X_2)^4\Big]^{1/2}\\
\le&(M_a h_n)^{1/2}\cdot\big(\mathbb{E}\big[\widetilde{X}^4\big|\widetilde{W}=0\big]M+MM_KC_0\mathbb{E}\big[\widetilde{X}^4\big]\big)^{1/2}\\
\le&(12M\kappa_x^4+12MM_KC_0\kappa_x^4)^{1/2}\cdot M_a^{1/2}h_n^{1/2},
\end{aligned}
\tag{A4.59}
$$

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 $\widetilde{X}$ conditional on $\widetilde{W}=0$ and unconditionally. Combining (A4.57)-(A4.59), we have

$$
\begin{aligned}
\sigma^2 &\le \left\{4M_v^2MM_K\kappa_x^2(1+C_0)C_0^{2\alpha-\gamma_1}+2M_d^2\cdot(12M\kappa_x^4+12MM_KC_0\kappa_x^4)^{1/2}\cdot M_a^{1/2}C_0^{-1/2-\gamma_1}\right\}M_Kh_n^{\gamma_1}\\
&=\tau_4 n^{-\gamma_1/2}\{\log(np)\}^{\gamma_1/2},
\end{aligned}
$$

where $\gamma_1=\min\left\{2\alpha-1,-1/2\right\}$.

For bounding B^2, we have

$$
\begin{aligned}
B^2 =&n\sup_{D_2}\mathbb{E}\big[g(D_1,D_2)^2\big|D_2\big]\\
\le&n\sup_{D_2}\mathbb{E}\Big[\frac{2M_v^2|W_1-W_2|^{2\alpha}+2M_d^2\,\mathbb{I}\left\{(W_1,W_2)\in A\right\}}{h_n^2}K^2\Big(\frac{W_1-W_2}{h_n}\Big)\\
&\cdot(X_1-X_2)^2\,\mathbb{I}\left\{(X_1-X_2)^2\le C\log(np)\right\}\big|D_2\Big]\\
\le&\frac{2CM_Kn\log(np)}{h_n}\Big\{M_v^2\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1-W_2}{h_n}\Big)|W_1-W_2|^{2\alpha}\big|D_2\Big]\\
&+M_d^2M_K\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1-W_2}{h_n}\Big)\,\mathbb{I}\left\{(W_1,W_2)\in A\right\}\big|D_2\Big]\Big\}.
\end{aligned}
\tag{A4.60}
$$

By a similar argument as used in (A4.56), for some $t\in[0,1]$, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1-W_2}{h_n}\Big)|W_1-W_2|^{2\alpha}\big|D_2\Big]\\
=&\int\int K(w)|wh_n|^{2\alpha}f_{W|X}(W_2+wh_n,x)\,dw\,dF_X(x)\\
=&\int\int K(w)|wh_n|^{2\alpha}\Big\{f_{W|X}(W_w,x)+\frac{\partial f_{W|X}(w,x)}{\partial w}\Big|_{(W_2+twh_n,x)}\cdot wh_n\Big\}\,dw\,dF_X(x)\\
\le&MM_K(1+C_0)h_n^{2\alpha}
\end{aligned}
\tag{A4.61}
$$

Combining (A4.60) and (A4.61), we have

$$
\begin{aligned}
B^2 &\le \frac{2CM_Kn\log(np)}{h_n}\left\{M_v^2MM_K(1+C_0)h_n^{2\alpha}+M_d^2(M+MM_KC_0)\right\}\\
&\le 2CM_K\left\{M_v^2MM_K(1+C_0)h_n^{2\alpha}+M_d^2(M+MM_KC_0)\right\}K_1^{-1}n^{3/2}\{\log(np)\}^{1/2}\\
&=\tau_5 n^{3/2}\{\log(np)\}^{1/2}
\end{aligned}
$$

We take

$$
\begin{aligned}
C &= C_Z \cdot 2\kappa_x^2, \\
t &= C_t 4\tau_3^{1/2}\binom{n}{2}\{\log(np)/n\}^{1/2}, \\
u &= C_u \log p, \text{ for } C_u > 1,
\end{aligned}
$$

and require $n > \max\{64c(c+1)^2\tau_2^2\tau_3^{-1}\{\log(np)\}^4, \{\log(np)\}^{5/3}\}$. For simplicity, we further take $C_t^2 = C_u = c > 1$, and $C_Z = \max\{2c, 4\} \le 2c + 2$. Then by Lemma A3.4, we ha ve

$$
\mathbb{P}\Big\{\binom{n}{2}^{-1}|\widetilde{U} - \mathbb{E}[\widetilde{U}]| \ge A_1\{\log(np)/n\}^{1/2}\Big\} \le 2\exp(-C_t^2\log(np)) + C_5\exp(-C_u \log p),
$$

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 $|\mathbb{E}[\widetilde{U}] - \mathbb{E}[U]|$, and complete the proof.

We have

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)(X_1 - X_2)^2\,\mathbb{1}\big\{(X_1 - X_2)^2 > C\log(np)\big\}\Big] \\
\le&\mathbb{E}\big[(X_1 - X_2)^2\,\mathbb{1}\big\{(X_1 - X_2)^2 > C\log(np)\big\}\big|W_1 = W_2\big]M \\
&+ MM_KC_0\mathbb{E}[(X_1 - X_2)^2\,\mathbb{1}\big\{(X_1 - X_2)^2 > C\log(np)\big\}] \\
\le&4(M + MM_KC_0)C_Z\kappa_x^2 \cdot \{\log(np)/n\},
\end{aligned}
$$

where the first inequality is by applying Lemma A3.2 on $(X_1 - X_2)^2\,\mathbb{1}\big\{(X_1 - X_2)^2 > C\log(np)\big\}$ 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 $a = C\log(np) \ge 4\kappa_x^2$.

Based on earlier arguments, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)(V_1 - V_2)^2\Big]^{1/2} \\
\le&M_v\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)|W_1 - W_2|^{2\alpha}\Big]^{1/2} + M_d\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)\Big]^{1/2} \\
\le&(M + MM_KC_0)^{1/2}(M_vC_0^\alpha + M_d),
\end{aligned}
$$

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

$$
\begin{aligned}
&\binom{n}{2}^{-1}\big|\mathbb{E}[\widetilde{U}] - \mathbb{E}[U]\big| \\
=&\Big|\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)(X_1 - X_2)(V_1 - V_2)\,\mathbb{1}\big\{(X_1 - X_2)^2 > C\log(np)\big\}\Big]\Big| \\
\le&\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)(X_1 - X_2)^2\,\mathbb{1}\big\{(X_1 - X_2)^2 > C\log(np)\big\}\Big]^{\frac{1}{2}}\mathbb{E}\Big[\frac{1}{h_n}K\Big(\frac{W_1 - W_2}{h_n}\Big)(V_1 - V_2)^2\Big]^{\frac{1}{2}} \\
\le&A_2 \cdot \{\log(np)/n\}^{1/2},
\end{aligned}
\tag{A4.62}
$$

where $A_2 = 2(M + MM_KC_0) \cdot (M_vC_0^\alpha + M_d)C_Z\kappa_x$.

Denote $\mathcal{A}_{[n]} = \big\{(X_i - X_j)^2 \le C\log(np),\ i,j \in [n],\ i<j\big\}$, and we have

$$
\begin{aligned}
&\mathbb{P}\Big\{\binom{n}{2}^{-1}\big|U-\mathbb{E}[U]\big| \ge (A_1+A_2)\cdot\Big(\frac{\log(np)}{n}\Big)^{1/2}\Big\}\\
\le&\mathbb{P}\Big\{\binom{n}{2}^{-1}\big|U-\mathbb{E}[U]\big| \ge (A_1+A_2)\cdot\Big(\frac{\log(np)}{n}\Big)^{1/2}\cap\mathcal{A}_{[n]}\Big\} + \mathbb{P}(\mathcal{A}_{[n]}^{\mathsf{c}})\\
\le&\mathbb{P}\Big\{\binom{n}{2}^{-1}\big|\widetilde{U}-\mathbb{E}[U]\big| \ge (A_1+A_2)\cdot\Big(\frac{\log(np)}{n}\Big)^{1/2}\cap\mathcal{A}_{[n]}\Big\} + \mathbb{P}(\mathcal{A}_{[n]}^{\mathsf{c}})\\
\le&\mathbb{P}\Big\{\binom{n}{2}^{-1}\big|\widetilde{U}-\mathbb{E}[U]\big| \ge (A_1+A_2)\cdot\Big(\frac{\log(np)}{n}\Big)^{1/2}\Big\} + \mathbb{P}(\mathcal{A}_{[n]}^{\mathsf{c}})\\
\overset{(i)}{\le}&\mathbb{P}\Big\{\binom{n}{2}^{-1}\big|\widetilde{U}-\mathbb{E}[\widetilde{U}]\big| \ge A_1\cdot\Big(\frac{\log(np)}{n}\Big)^{1/2}\Big\} + \mathbb{P}(\mathcal{A}_{[n]}^{\mathsf{c}})\\
\le&2\exp(-C_t^2\log(np)) + C_5\exp(-C_u\log p) + \frac{n^2}{2}\exp\{-2C_Z\log(np)/2\}\\
\le&2\exp(-C_t^2\log(np)) + C_5\exp(-C_u\log p) + \frac{1}{2}\exp\{-C_Z\log p/2\},
\end{aligned}
$$

where (i) is by (A4.62). This completes the proof. $\square$

References

Arcones, M. A. and Gine, E. (1993). Limit theorems for U-processes. *The Annals of Probability*, 21(3):1494–1542.

de la Peña, V. and Giné, E. (2012). *Decoupling: From Dependence to Independence*. Springer.

Giné, E., Latała, R., and Zinn, J. (2000). Exponential and moment inequalities for U-statistics. In *High Dimensional Probability II*, pages 13–38. Springer.

Han, F. and Liu, H. (2016). ECA: High dimensional elliptical component analysis in non-Gaussian distributions. *Journal of the American Statistical Association*, (to appear).

Houdré, C. and Reynaud-Bouret, P. (2003). Exponential inequalities, with constants, for U-statistics of order two. In *Stochastic Inequalities and Applications*, pages 55–69. Springer.

Mitra, R. and Zhang, C.-H. (2014). Multivariate analysis of nonparametric estimates of large correlation matrices. *arXiv:1403.6195*.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, 56(4):931–954.

Rudelson, M. and Zhou, S. (2013). Reconstruction from anisotropic random measurements. *IEEE Transactions on Information Theory*, 59(6):3434–3447.

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing*, pages 210–268. Cambridge University Press.