ON A NADARAYA-WATSON ESTIMATOR WITH TWO BANDWIDTHS

FABIENNE COMTE* AND NICOLAS MARIE**

Abstract. In a regression model, we write the Nadaraya-Watson estimator of the regression function as the quotient of two kernel estimators, and propose a bandwidth selection method for both the numerator and the denominator. We prove risk bounds for both data driven estimators and for the resulting ratio. The simulation study confirms that both estimators have good performances, compared to the ones obtained by cross-validation selection of the bandwidth. However, unexpectedly, the single-bandwidth cross-validation estimator is found to be much better while choosing very small bandwidths. It is even better than the ratio of the two best estimators of the numerator and the denominator of the collection, for which larger bandwidth are to be chosen.
AMS (2010) classification. 62G08–62G05.
Keywords. Bandwidth selection, Nonparametric kernel estimator, Qotient estimator, Regression model.

1. Introduction

Consider $n \in \mathbb{N}^*$ independent random variables $X_1, \dots, X_n$ having the same probability distribution of density f with respect to Lebesgue's measure. Consider also the random variables $Y_1, \dots, Y_n$ defined by

$$Y_k = b(X_k) + \varepsilon_k \ ; \ k \in \{1, \dots, n\},$$

where b is a continuous function from $\mathbb{R}$ into itself and $\varepsilon_1, \dots, \varepsilon_n$ are n i.i.d. centered random variables of variance $\sigma^2 > 0$ and respectively independent of $X_1, \dots, X_n$.
Since Nadaraya (1964) and Watson (1964), a lot of consideration has been given to the estimator of b defined by

$$\widehat{b}_{n,h}(x) := \frac{\sum_{k=1}^{n} K\left(\frac{X_k - x}{h}\right) Y_k}{\sum_{k=1}^{n} K\left(\frac{X_k - x}{h}\right)} \ ; \ x \in \mathbb{R},$$

where $K : \mathbb{R} \to \mathbb{R}$ is a kernel, and $h > 0$ is the bandwidth.

This estimator has been dealt with as a weighted estimator, for $K \geq 0$:

$$\widehat{b}_{n,h}(x) = \sum_{i=1}^{n} w_{i,h,n}(x) Y_i, \quad w_{i,h,n}(x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{X_k - x}{h}\right)},$$

and is often called "local average regression". It is studied e.g. in Wand and Jones (1995), Györfi *et al.* (2002) or defined in Tsybakov (2009); recent papers still propose methods to improve the estimation, see Chang *et al.* (2017). Several strategies have been proposed to select the bandwidth in a data driven way. Cross-validation based on leave-one-out principle is one of the most standard methods to perform this choice (see Györfi *et al.* (2002)), even if a lot of refinements have been proposed. Optimal rates depend on the regularity of the function $b(.)$ and have been first established by Stone (1982), $n^{-p/(2p+1)}$ for b admitting p derivatives. From theoretical point of view, the rates of the adaptive final estimator are not always given, nor proved.

In this paper, we re-write the Nadaraya-Watson as the quotient of two estimators, an estimator of bf divided by an estimator of f:

$$\widehat{bf}_{n,h}(x) := \frac{1}{nh} \sum_{k=1}^{n} K\left(\frac{X_k - x}{h}\right) Y_k$$

and

$$\widehat{f}_{n,h'}(x) := \frac{1}{nh'} \sum_{k=1}^{n} K\left(\frac{X_k - x}{h'}\right).$$

Clearly, $\widehat{f}_{n,h'}$ is the Parzen-Rosenblatt estimator of f (see Rosenblatt [13] and Parzen [12]). The question we are interested in is the following: can we choose separately the two bandwidths in an adaptive way and obtain good performance for each, and then for the ratio? This is why we study the following estimator

$$\widehat{b}_{n,h,h'}(x) := \frac{\widehat{bf}_{n,h}(x)}{\widehat{f}_{n,h'}(x)} \ ; \ x \in \mathbb{R}$$

as an estimator of the regression function b, where $h, h' > 0$, $K : \mathbb{R} \to \mathbb{R}_+$ is a kernel. Thus, $\widehat{b}_{n,h,h} = \widehat{b}_{n,h}$ is the initial Nadaraya-Watson estimator of b with single bandwidth h. For this reason, the estimator studied in this paper is called the *two bandwidths Nadaraya-Watson* (2bNW) estimator.

Adaptive estimation of the density has been widely studied recently. A bandwidth selection method has been proposed by Goldenschluger and Lepski (2011), and proved to reach the adequate bias-variance compromise. Implementation of this method revealed to be difficult due to the choice of two constants involved in the procedure, the intuition of which is not obvious. This is why the question was further investigated by Lacour *et al.* (2017): they improve and modify the strategy, and, relying on specific theoretical tools for their proofs (precisely, a deviation inequality for U-statistics proved by Houdré and Reynaud-Bouret (2003)), they bound the Mean Integrated Square Error of their final estimator, which they call PCO (Penalised Comparison to Overfitting) estimator. Numerically, the good performance of their proposal has been illustrated in a naive way and for high order kernels in Comte and Marie (2019), and through a systematic numerical study in Varet *et al.* (2019), including multivariate case. These two methods and the associated results are dedicated to the selection of h' for $\widehat{f}_{n,h'}(x)$, and we can use them. Unfortunately, the theoretical results do not apply to $\widehat{bf}_{n,h}(x)$, mainly because they hold under a boundedness assumption: in our context, this would lead to assume that the Y_k's are bounded. We do not want to require such an assumption as it would exclude the case of Gaussian errors ε_i, for instance. Thus, we give moment assumptions under which the Goldenshluger and Lepski method on the one hand (see Section 3) and the PCO estimator on the other hand (see Section 4) can be applied to the estimation of bf. When gathering the results for the numerator and the denominator, we can bound the risk of the quotient estimator of b.

Concretely, we implement the PCO method for bf and compare it with a cross-validation (CV) strategy: in our examples, PCO almost always performs slightly better than CV. Therefore, the PCO adaptive estimation strategies for f and for bf are clearly good. However, unexpectedly, the quotient fails systematically to beat the specific regression CV method. Even if we compare the classical single-bandwidth CV regression estimator to the ratio of the oracles estimators of the numerator and the denominator, the former wins, and we obtain a quotient with two bandwidths which is in mean much worse than the CV estimator with single bandwidth. In practice, the bandwidth selected by the CV algorithm in that case is very small, and associated to quite bad estimators of the numerator and the denominator. We believe that both positive but also negative results are of interest, and detailed tables, explanations and discussion are given in Section 5.

Notations:

(1) For every square integrable functions $f, g : \mathbb{R} \to \mathbb{R}$,

$$(f * g)(x) := \int_{-\infty}^{\infty} f(x-y)g(y)dy \ ; \ x \in \mathbb{R}.$$

(2) $K_\varepsilon := 1/\varepsilon K(\cdot/\varepsilon)$ for every $\varepsilon > 0$.

2. Bound on the MISE of the 2bNW estimator

First, we state some simple risk bound results in the case of a fixed bandwidth.

Consider $\beta > 0$ and $\ell := \lfloor \beta \rfloor$, where $\lfloor \beta \rfloor$ denotes the largest integer smaller than β. In the sequel, the kernel K and the density function f fulfill the following assumption.

Assumption 2.1.

(i) *The map* K *belongs to* $\mathbb{L}^2(\mathbb{R}, dy)$, K *is bounded and* $\int_{\mathbb{R}} K(y)dy = 1$.

(ii) *The density function* f *is bounded.*

Under this assumption, a suitable control of the MISE of $\widehat{bf}_{n,h}$ has been established in Comte [3], Proposition 4.2.1.

Proposition 2.2. *Under Assumption 2.1,*

$$\mathbb{E}(\|\widehat{bf}_{n,h} - bf\|_2^2) \leqslant \|bf - (bf)_h\|_2^2 + \frac{\mathfrak{c}_{K,Y}}{nh}$$

where $\mathfrak{c}_{K,Y} := \|K\|_2^2 \mathbb{E}(Y_1^2)$.

In order to provide a suitable control of the MISE of the 2bNW estimator, we assume that b and f fulfill the following assumption.

Assumption 2.3. *The function* $b^2 f$ *is bounded by a constant* $\mathfrak{c}_{b,f} > 0$.

Note that this assumption does not require that b is bounded and is satisfied in most classical examples. Moreover, for any $\mathcal{S} \in \mathcal{B}(\mathbb{R})$, consider the norm $\|.\|_{2,f,\mathcal{S}}$ on $\mathbb{L}^2(\mathcal{S}, f(x)dx)$ defined by

$$\|\varphi\|_{2,f,\mathcal{S}} := \left(\int_{\mathcal{S}} \varphi(x)^2 f(x)dx\right)^{1/2} \quad ; \forall \varphi \in \mathbb{L}^2(\mathcal{S}, f(x)dx).$$

Proposition 2.4. *Let* $(m_k)_{k\in\mathbb{N}}$ *be a decreasing sequence of* $]0, \infty[$ *such that* $\lim_\infty m_k = 0$ *and, for every* $k \in \mathbb{N}$, *consider*

$$\mathcal{S}_k := \{x \in \mathbb{R} : f(x) \geqslant m_k\}.$$

Under Assumptions 2.1 and 2.3,

$$\mathbb{E}(\|\widehat{b}_{n,h,h'} - b\|_{2,f,\mathcal{S}_n}^2) \leqslant \frac{8\mathfrak{c}_f}{m_n^2}\left(\|bf - (bf)_h\|_2^2 + \frac{\mathfrak{c}_{K,Y}}{nh} + 2\mathfrak{c}_{b,f}\left(\|f - f_{h'}\|_2^2 + \frac{\mathfrak{c}_K}{nh'}\right)\right),$$

where $(bf)_h := K_h * (bf)$, $\mathfrak{c}_f := \|f\|_\infty^2 \vee 1$ *and* $\mathfrak{c}_K := \int_{\mathbb{R}} K(y)^2 dy$.

Proposition 2.4 gives a decomposition of the risk of the quotient estimator as the sum of the risks of the estimators of the numerator bf and the denominator f, up to the multiplicative constant $8\mathfrak{c}_f/m_n^2$. Therefore, the rate of the quotient estimator is, in the best case, the worst rate of the two estimators used to define it (see also Remark 2.5 below). The factor $1/m_n^2$ may imply a global loss with respect to this rate. Clearly, the smaller m_n, the larger the loss.

For instance, if f is lower bounded by a known constant f_0 on a given compact set A, then we can take $\mathcal{S}_n = A$ and $m_n = f_0$; in that case, no loss occurs. If f_0 is unknown, we still can bound the risk with $\mathcal{S}_n = A$ and $1/m_n^2 = \log(n)$, for n large enough; a log-loss occurs then in the rate.

Remark 2.5. *We consider, for* $\beta, L > 0$, *the Nikol'ski ball* $\mathcal{H}(\beta, L)$, *defined as the set of functions* $\varphi : \mathbb{R} \to \mathbb{R}$ *such that* $\varphi^{(\ell)}$ *exists and satisfies*

$$\left[\int_{-\infty}^{\infty} \left(\varphi^{(\ell)}(x+t) - \varphi^{(\ell)}(x)\right)^2 dx\right]^{1/2} \leqslant L|t|^{\beta-\ell}, \quad \forall t \in \mathbb{R}.$$

Now, assume that bf *belongs to* $\mathcal{H}(\beta_1, L)$ *and* f *to* $\mathcal{H}(\beta_2, L)$. *We also assume that the kernel* K *satisfies Assumption 2.1 and is of order* $\ell = \lfloor \max(\beta_1, \beta_2) \rfloor$, *that is* $\int u^k K(u)du = 0$ *for* $k = 1, \dots, \ell$. *Then, it follows from Tsybakov (2009, chapter 1) that* $\|bf - (bf)_h\|_2^2 \leq C(\beta_1, L)h^{2\beta_1}$ *and* $\|f - f_{h'}\|_2^2 \leq C'(\beta_2, L)(h')^{2\beta_2}$. *This implies that choosing* $h_{\text{opt}} = c_1 n^{1/(2\beta_1+1)}$ *in Proposition 2.2 yields* $\mathbb{E}(\|\widehat{bf}_{n,h_{\text{opt}}} - bf\|_2^2) \lesssim n^{-2\beta_1/(2\beta_1+1)}$, *which is a standard optimal rate of estimation on Nikols'ski balls. The same rate holds for the estimation of* f *under our assumption, with* β_1 *replaced by* β_2, *and* $h'_{\text{opt}} = c_2 n^{1/(2\beta_2+1)}$. *This implies that*

$$\|bf - (bf)_{h_{\text{opt}}}\|_2^2 + \frac{\mathfrak{c}_{K,Y}}{nh_{\text{opt}}} + 2\mathfrak{c}_{b,f}\left(\|f - f_{h'_{\text{opt}}}\|_2^2 + \frac{\mathfrak{c}_K}{nh'_{\text{opt}}}\right) \lesssim \max(n^{-2\beta_1/(2\beta_1+1)}, n^{-2\beta_2/(2\beta_2+1)}).$$

So the rate is optimal if $\beta := \min(\beta_1, \beta_2)$ is the regularity of $b(.)$. However, such bandwidth choices are not possible as they depend on unknown regularity parameters: data driven bandwidth selection methods are settled to automatically reach a squared bias-variance compromise, inducing the optimal rate if the function under estimation does belong to a regularity space.

3. A bandwidth selection procedure for the 2bNW estimator based on the GL method

The bound on the MISE of $\widehat{b}_{n,h,h'}$ obtained in Proposition 2.4 suggests to select h and h' separately, so that both bounds are minimal. The Goldenshluger-Lepski (2011) method allows to do this for $\widehat{f}_{n,h'}$, but requires to be extended to the estimator of bf. In particular, extensions of the proof are required as we do not wish to assume that the Y_k's are bounded.
Consider the collection of bandwidths $\mathcal{H}_n := \{h_1, \dots, h_{N(n)}\} \subset [0,1]$, where $N(n) \in \{1, \dots, n\}$ and

$$\frac{1}{n} < h_1 < \dots < h_{N(n)}.$$

Moreover, we will need the following constraints.

Assumption 3.1. *There exists $\mathfrak{m} > 0$, not depending on n, such that*

$$\frac{1}{n} \sum_{h \in \mathcal{H}_n} \frac{1}{h} \leqslant \mathfrak{m}$$

and for every $c > 0$, there exists $\mathfrak{m}(c) > 0$, not depending on n, such that

$$\sum_{h \in \mathcal{H}_n} \frac{1}{\sqrt{h}} \exp\left(-\frac{c}{\sqrt{h}}\right) \leqslant \mathfrak{m}(c).$$

Example. Consider dyadic bandwidths, defined by $h_k = 2^{-k}$ for $k = 0, 1, \dots, [\log(n)/\log(2)]$. Then

$$\frac{1}{n} \sum_{k=1}^{[\log(n)/\log(2)]} 2^k \leq \frac{2n-1}{n} \leq 2,$$

and $\sum_{k=1}^{[\log(n)/\log(2)]} 2^{k/2} \exp(-c2^{k/2}) \leq \sum_{k=1}^{n} \sqrt{k} \exp(-c\sqrt{k}) \leq \mathfrak{m}(c) < \infty$. Thus, Assumption 3.1 is fulfilled.

Consider also

$$\begin{aligned}\widehat{bf}_{n,h,\eta}(x) &:= (K_\eta * \widehat{bf}_{n,h})(x)\\ &= \frac{1}{n} \sum_{k=1}^{n} Y_k (K_\eta * K_h)(X_k - x).\end{aligned}$$

A way to extend the Goldenshluger-Lepski method to $\widehat{bf}_{n,h}$ is to solve the minimization problem

$$\min_{h \in \mathcal{H}_n} \{A_n(h) + V_n(h)\}, \tag{1}$$

where

$$A_n(h) := \sup_{\eta \in \mathcal{H}_n} (\|\widehat{bf}_{n,h,\eta} - \widehat{bf}_{n,\eta}\|_2^2 - V_n(\eta))_+ \quad \text{and} \quad V_n(h) := \upsilon \frac{\mathfrak{c}_{K,Y}}{nh} \|K\|_1^2,$$

with $\upsilon > 0$ not depending on n and h, and $\mathfrak{c}_{K,Y} = \|K\|_2^2 \mathbb{E}(Y_1^2)$. In the sequel, the solution to the minimization Problem (1) is denoted by $\widehat{h}_n$.

Theorem 3.2. *Under Assumptions 2.1 and 3.1, and if $\mathbb{E}(Y_1^6) < \infty$, then there exist two deterministic constants $\mathfrak{c}, \overline{\mathfrak{c}} > 0$, not depending on n, such that*

$$\mathbb{E}(\|\widehat{bf}_{n,\widehat{h}_n} - bf\|_2^2) \leqslant \mathfrak{c} \cdot \inf_{h \in \mathcal{H}_n} \{\|(bf)_h - bf\|_2^2 + V_n(h)\} + \overline{\mathfrak{c}} \frac{\log(n)^2}{n}.$$

Theorem 3.2 says that $\widehat{bf}_{n,\widehat{h}_n}$ automatically leads to a squared bias ($\|(bf)_h - bf\|_2^2$)–variance ($V_n(h)$) compromise, up to the multiplicative constant $\mathfrak{c}$ and to the quantity, $\overline{\mathfrak{c}}(\log(n)^2/n)$, which is negligible with respect to possible rate of convergence, see Remark 2.5.
We recall now a version of the result proved by Goldenshluger and Lepski (2011), which is available for the estimator of f (see also a simplified proof in [3], section 4.2). Let us consider

$$\widehat{h}'_n \in \arg\min_{h'\in\mathcal{H}_n}\{A'_n(h') + V'_n(h')\},$$

where

$$A'_n(h') := \sup_{\eta\in\mathcal{H}_n}(\|K_\eta * \widehat{f}_{n,h'} - \widehat{f}_{n,\eta}\|_2^2 - V'_n(\eta))_+ \text{ and } V'_n(h') := \chi\frac{\|K\|_2^2\|K\|_1^2}{nh'}$$

with $\chi > 0$ not depending on n and h'. Under Assumptions 2.1 and 3.1, there exist two deterministic constants $\mathfrak{c}', \overline{\mathfrak{c}}' > 0$, not depending on n, such that

$$\mathbb{E}(\|\widehat{f}_{n,\widehat{h}'_n} - f\|_2^2) \leqslant \mathfrak{c}' \cdot \inf_{h'\in\mathcal{H}_n}\{\|f_{h'} - f\|_2^2 + V'_n(h')\} + \frac{\overline{\mathfrak{c}}'}{n}. \tag{2}$$

Corollary 3.3. *Let* $(m_k)_{k\in\mathbb{N}}$ *be a decreasing sequence of* $]0,\infty[$ *such that* $\lim_\infty m_k = 0$ *and, for every* $k \in \mathbb{N}$*, consider*

$$\mathcal{S}_k := \{x \in \mathbb{R} : f(x) \geqslant m_k\}.$$

Under Assumptions 2.1, 2.3 and 3.1, if $\mathbb{E}(Y_1^6) < \infty$*, then*

$$\mathbb{E}(\|\widehat{b}_{n,\widehat{h}_n,\widehat{h}'_n} - b\|_{2,f,\mathcal{S}_n}^2) \leqslant \mathfrak{C}_n \inf_{(h,h')\in\mathcal{H}_n^2}\{\|(bf)_h - bf\|_2^2 + \|f_{h'} - f\|_2^2 + V_n(h) + V'_n(h')\} + \overline{\mathfrak{C}}_n\frac{\log(n)^2}{n},$$

where

$$\mathfrak{C}_n := \frac{8\mathfrak{c}_f}{m_n^2}(\mathfrak{c} \vee (2\mathfrak{c}_{b,f}\mathfrak{c}')) \text{ and } \overline{\mathfrak{C}}_n := \frac{8\mathfrak{c}_f}{m_n^2}(\overline{\mathfrak{c}} + 2\mathfrak{c}_{b,f}\overline{\mathfrak{c}}').$$

The comments following Proposition 2.4 and in Remark 2.5 apply here.

4. A bandwidths selection procedure for the 2bNW estimator based on the PCO method

The Goldenshluger-Lepski method is mathematically very nice and provides a rigorous risk bound for the adaptive estimator with random bandwidth. However, it has been acknowledged as being difficult to implement, due to the square grid in h, η required to compute intermediate versions of the criterion and to the lack of intuition to guide the choice of the constants υ and χ which should be calibrated from preliminary simulation experiments, see e.g. Comte and Rebafka (2016). This is the reason why Lacour et al. [10] investigated and proposed a simplified criterion (PCO) relying on deviation inequalities for U-statistics due to Houdré and Reynaud-Bouret [8]. This inequality applies in our more complicated context and Lacour-Massart-Rivoirard's result can be extended here as follows.

Let us recall that $K_h(\cdot) := 1/hK(\cdot/h)$ and

$$(bf)_h = \mathbb{E}(\widehat{bf}_{n,h}) = K_h * (bf)$$

(see Lemma 6.1). Let $h_{\min}$ be the minimal proposal in $\mathcal{H}_n$ and consider

$$\mathrm{crit}(h) := \|\widehat{bf}_{n,h} - \widehat{bf}_{n,h_{\min}}\|_2^2 + \mathrm{pen}(h)$$

with

$$\mathrm{pen}(h) := \frac{2\langle K_{h_{\min}}, K_h\rangle_2}{n^2}\sum_{k=1}^n Y_k^2.$$

Then, let us define

$$\widetilde{h}_n \in \arg\min_{h\in\mathcal{H}_n}\mathrm{crit}(h).$$

In the sequel, in addition to Assumption 2.1, the kernel K, the functions b and f, the distribution of Y_1 and $h_{\min}$ fulfill the following assumption.

Assumption 4.1. *The kernel K is symmetric and $K(0) > 0$,*

$$\frac{1}{nh_{\min}} \leqslant 1,$$

bf is bounded and there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|Y_1|)) < \infty$.

Note that, as for Assumption 2.3, we can note that assuming bf bounded does not require b to be bounded, since most densities decrease fast at infinity. Moreover, the moment condition here is $\mathbb{E}(\exp(\alpha|Y_1|)) < \infty$ and stronger than for the Goldenschluger and Lepski method ($\mathbb{E}(|Y_1|^6) < +\infty$).

Theorem 4.2. *Consider $\vartheta \in (0,1)$. Under Assumptions 2.1 and 4.1, there exist two deterministic constants $\mathfrak{a}, \mathfrak{b} > 0$, not depending on n, $h_{\min}$ and ϑ, such that*

$$\mathbb{E}(\|\widehat{bf}_{n,\widetilde{h}_n} - bf\|_2^2) \leqslant (1+\vartheta) \inf_{h\in\mathcal{H}_n} \mathbb{E}(\|\widehat{bf}_{n,h} - bf\|_2^2) + \frac{\mathfrak{a}}{\vartheta}\|(bf)_{h_{\min}} - bf\|_2^2 + \frac{\mathfrak{b}}{\vartheta}\cdot\frac{\log(n)^5}{n}.$$

Theorem 4.2 says that the estimator $\widehat{bf}_{n,\widetilde{h}_n}$ has performance of order of the best estimator of the collection $\inf_{h\in\mathcal{H}_n}\mathbb{E}(\|\widehat{bf}_{n,h} - bf\|_2^2)$ up to a factor $(1+\vartheta)$. Indeed the two other terms can be considered as negligible. If bf is in the Nikol'ski ball of Remark 2.5, then the first right-hand-side term has order $n^{-2\beta/(2\beta+1)}$. As, for $h_{\min} = 1/n$, $\|(bf)_{h_{\min}} - bf\|_2^2$ has order $n^{-2\beta}$, both this term and the last residual term $\log^5(n)/N$ are negligible compared to the first one.

Now we state the result that can be deduced from Lacour *et al.* (2017) for the estimator of f. Let us consider

$$\widetilde{h}'_n \in \arg\min_{h'\in\mathcal{H}_n} \mathrm{crit}'(h')$$

where

$$\mathrm{crit}'(h') := \|\widehat{f}_{n,h'} - \widehat{f}_{n,h_{\min}}\|_2^2 + \mathrm{pen}'(h') \text{ and } \mathrm{pen}'(h') := \frac{2\langle K_{h_{\min}}, K_h\rangle_2}{n^2}.$$

By Lacour et al. [10], Theorem 2, there exist two deterministic constants $\mathfrak{a}', \mathfrak{b}' > 0$, not depending on n and $h_{\min}$, such that for every $\vartheta \in (0,1)$,

$$\mathbb{E}(\|\widehat{f}_{n,\widetilde{h}'_n} - f\|_2^2) \leqslant (1+\vartheta)\inf_{h'\in\mathcal{H}_n}\mathbb{E}(\|\widehat{f}_{n,h'} - f\|_2^2) + \frac{\mathfrak{a}'}{\vartheta}\|f_{h_{\min}} - f\|_2^2 + \frac{\mathfrak{b}'}{\vartheta n}.$$

Corollary 4.3. *Let $(m_k)_{k\in\mathbb{N}}$ be a decreasing sequence of $]0,\infty[$ such that $\lim_\infty m_k = 0$ and, for every $k \in \mathbb{N}$, consider*

$$\mathcal{S}_k := \{x \in \mathbb{R} : f(x) \geqslant m_k\}.$$

Consider also $\vartheta \in (0,1)$. Under Assumptions 2.1, 2.3 and 4.1,

$$\begin{aligned}
&\mathbb{E}(\|\widehat{bf}_{n,\widetilde{h}_n,\widetilde{h}'_n} - bf\|_{2,f,\mathcal{S}_n}^2)\\
&\leqslant (1+\vartheta)\mathfrak{C}_n(1,1)\inf_{(h,h')\in\mathcal{H}_n^2}\{\mathbb{E}(\|\widehat{bf}_{n,h} - bf\|_2^2) + \mathbb{E}(\|\widehat{f}_{n,h'} - f\|_2^2)\}\\
&\quad+\frac{\mathfrak{C}_n(\mathfrak{a},\mathfrak{a}')}{\vartheta}(\|(bf)_{h_{\min}} - bf\|_2^2 + \|f_{h_{\min}} - f\|_2^2) + \frac{\mathfrak{C}_n(\mathfrak{b},\mathfrak{b}')}{\vartheta}\cdot\frac{\log(n)^5}{n},
\end{aligned}$$

where

$$\mathfrak{C}_n(u,v) := \frac{8\mathfrak{c}_f}{m_n^2}(u \vee (2\mathfrak{c}_{b,f}v)) \ ; \ \forall u, v \in \mathbb{R}.$$

The proof of Corollary 4.3 relies to the same arguments as the proof of Corollary 3.3 provided in Subsection 3.3, and is therefore omitted.

5. Simulation study

For the noise, we consider $\varepsilon \sim \sigma\mathcal{N}(0,1)$, with $\sigma = 0.1$, and for the signal we take either $X \sim \mathcal{N}(0,1)$ or $X \sim \gamma(3,2)/5$ (where the factor 5 is set to keep the variance of X of order 1, as in the first case).

For the functions b, we took functions with different features and regularities:

- $b_1(x) = \exp(-x^2/2)$,
- $b_2(x) = x^2/4 - 1$,
- $b_3(x) = \sin(\pi x)$,
- $b_4(x) = \exp(-|x|)$.

The PCO method is implemented for f and bf with a kernel of order 7 (i.e. $\int x^k K(x)dx = 0$ for $k = 1$ to 7), defined by $K(x) = 4n_1(x) - 6n_2(x) + 4n_3(x) - n_4(x)$ where $n_j(x)$ is the density of a Gaussian with mean 0 and variance j. Note that, for $n_{i,h}(x) = (1/h)n_i(x/h)$, it holds that

$$\langle n_{i,h_1}, n_{j,h_2}\rangle_2 = \int_{-\infty}^{\infty} n_{i,h_1}(x)n_{j,h_2}(x)dx = \frac{1}{\sqrt{2\pi}} \times \frac{1}{\sqrt{ih_1^2 + jh_2^2}}. \tag{3}$$

The bandwidth is selected among $M = 50$ equispaced values in between 0.01 and 1. All functions (true or estimated) are computed at 100 equispaced points in the interquantile interval corresponding to the 2% and 98% quantiles of X. The bandwidth is selected with the PCO criterion, where $h_{\min} = 0.01$, and

$$\mathrm{crit}(h) := \|\widehat{bf}_{n,h} - \widehat{bf}_{n,h_{\min}}\|_2^2 + 2\mathrm{pen}(h), \quad \mathrm{pen}(h) := \frac{2\langle K_{h_{\min}}, K_h\rangle_2}{n^2}\sum_{k=1}^n Y_k^2.$$

The $\mathbb{L}_2$-norm is computed as a Riemann sum on the interquantile interval, while the penalty is explicit and exact, thanks to formula (3).

The cross-validation criterion for selecting the bandwidth of $\widehat{bf}_{n,h}$ is computed as follows:

$$CV(h) = \int [\widehat{bf}_h(x)]^2 dx - \frac{2}{n(n-1)h}\sum_{i=1}^n \sum_{j=1, j\neq i}^n Y_i N\left(\frac{X_i - X_j}{h}\right),$$

where $N(.)$ is the Gaussian kernel and is also used to compute the estimator $\widehat{bf}_{n,h}$. The chosen bandwidth is the minimizer of $CV(h)$ in the same collection as previously.

For the one-bandwidth Nadaraya-Watson estimator $\widehat{b}_{n,h}$, also computed with the Gaussian kernel $N(.)$, the criterion is

$$CV_{\mathrm{NW}}(h) = \sum_{i=1}^n (Y_i - b_{n,h}^{(-i)}(X_i))^2, \quad b_{n,h}^{(-i)}(x) = \sum_{j=1, j\neq i}^n \frac{N((X_j - x)/h)}{\sum_{k=1, k\neq i} N((X_k - x)/h)} Y_j.$$

n	b_1f PCO	b_1f CV	b_1f Or	b_2f PCO	b_2f CV	b_2f Or	b_3f PCO	b_3f CV	b_3f Or	b_4f PCO	b_4f CV	b_4f Or
250	0.31	0.43	0.14	0.29	0.37	0.14	0.41	0.47	0.28	0.34	0.41	0.17
	(0.29)	(1.25)	(0.15)	(0.26)	(1.10)	(0.14)	(0.23)	(0.79)	(0.15)	(0.27)	(0.85)	(0.14)
500	0.16	0.28	0.09	0.15	0.23	0.08	0.23	0.31	0.16	0.19	0.32	0.11
	(0.13)	(0.70)	(0.07)	(0.12)	(0.56)	(0.07)	(0.13)	(0.67)	(0.09)	(0.13)	(0.67)	(0.07)
1000	0.09	0.21	0.05	0.09	0.21	0.04	0.12	0.24	0.09	0.11	0.19	0.07
	(0.07)	(0.57)	(0.04)	(0.07)	(0.60)	(0.03)	(0.07)	(0.55)	(0.05)	(0.07)	(0.40)	(0.04)

Table 1. 100*MISE (with 100*std in parenthesis below) for the estimation of bf corresponding to the four examples $b_1, \ldots, b_4$, 200 repetitions, $X \sim \mathcal{N}(0,1)$. Columns PCO and CV correspond to the two competing methods. "Or" is for "oracle" and gives the average error of the best possible performance of the collection, computed for each sample.

	b_1f			b_2f			b_3f			b_4f		
n	PCO	CV	Or	PCO	CV	Or	PCO	CV	Or	PCO	CV	Or
250	0.35	0.51	0.18	0.49	0.61	0.23	0.58	0.62	0.39	0.17	0.22	0.09
	(0.30)	(0.96)	(0.19)	(0.37)	(1.06)	(0.22)	(0.33)	(0.59)	(0.26)	(0.14)	(0.34)	(0.09)
500	0.19	0.23	0.09	0.26	0.29	0.12	0.29	0.29	0.20	0.09	0.11	0.05
	(0.15)	(0.28)	(0.09)	(0.19)	(0.29)	(0.11)	(0.18)	(0.35)	(0.14)	(0.08)	(0.13)	(0.04)
1000	0.10	0.17	0.05	0.13	0.15	0.07	0.14	0.18	0.10	0.05	0.08	0.03
	(0.07)	(0.30)	(0.04)	(0.09)	(0.15)	(0.05)	(0.08)	(0.31)	(0.06)	(0.04)	(0.14)	(0.02)

TABLE 2. 100*MISE (with 100*std in parenthesis below) for the estimation of bf, 200 repetitions, $X \sim \gamma(3,2)/5$.

	b_1f			b_2f			b_3f			b_4f		
n	PCO	CV	Or	PCO	CV	Or	PCO	CV	Or	PCO	CV	Or
250	0.68	0.52	0.55	0.76	0.61	0.62	0.35	0.28	0.30	0.56	0.40	0.38
	(0.16)	(0.16)	(0.07)	(0.16)	(0.17)	(0.08)	(0.04)	(0.06)	(0.03)	(0.18)	(0.14)	(0.07)
500	0.60	0.45	0.50	0.68	0.54	0.58	0.32	0.26	0.28	0.45	0.31	0.33
	(0.12)	(0.15)	(0.07)	(0.13)	(0.17)	(0.08)	(0.04)	(0.06)	(0.03)	(0.13)	(0.13)	(0.10)
1000	0.55	0.41	0.47	0.63	0.50	0.54	0.30	0.24	0.26	0.39	0.28	0.28
	(0.11)	(0.14)	(0.07)	(0.12)	(0.17)	(0.07)	(0.03)	(0.07)	(0.03)	(0.11)	(0.11)	(0.06)

TABLE 3. Selected bandwidth (with std in parenthesis below) for the estimation of bf, 200 repetitions, $X \sim \mathcal{N}(0,1)$.

Tables 1 and 2 give the MISE obtained for 200 repetitions and sample sizes 250, 500 and 1000, for the estimation of bf with PCO and CV method. The column "Or" gives the mean of the minimal squared errors for each sample, which requires to use the unknown true function and represents what could be obtained at best (that is if the best possible bandwidth was chosen for each sample). Table 1 corresponds to $X \sim \mathcal{N}(0,1)$ and Table 2 to $X \sim \gamma(3,2)/5$. We can see that PCO is globally better than the CV method, with not an important difference, and the oracle shows that we are in the right orders, even if not at best.

Table 3 presents the mean of the selected bandwidths in each case PCO and CV, and allows to compare it with the oracle bandwidth, for the same paths and configurations as previously. The conclusion here is that, in mean, the PCO method over-estimates the oracle bandwidth, while the CV method slightly under-evaluates it. Clearly, the too-large choice gives better result. We give only the results for Gaussian X, as those for our Gamma example are quite similar.

Tables 4 presents the results for the estimation of b, either with the $CV_{\rm NW}$ criterion or with ratio of PCO of bf and f, or with the ratio of the best estimators of bf and f in the collection. More precisely, the column "Or" gives here the MISE computed with the estimator of b obtained as a quotient of the two oracles of bf and f in each example and for each sample path. Clearly, the performance of the Nadaraya-Watson Cross-Validation criterion is much better, within a multiplicative factor from 2 and up to 6. The variance of the quotient estimators (oracle and PCO) are large, which shows that the mean performance is probably deteriorated by a few very bad results. Medians would have probably partly hidden the bad performances of the quotient estimators. But the result is puzzling: even the ratio of the two best estimators of the numerator and denominator does not reach the good performance of the single-bandwidth CV method. Table 5 shows in addition that the selected bandwidths are in mean very small. We can check that the ratio of this bad numerator divided by a bad denominator fits well to the b quotient function. It is likely that both imply a compensation resulting in a locally, and thus also globally better estimate.

n	b_1 CV	b_1 PCO	b_1 Or	b_2 CV	b_2 PCO	b_2 Or	b_3 CV	b_3 PCO	b_3 Or	b_4 CV	b_4 PCO	b_4 Or
250	0.34	2.77	1.25	0.49	3.09	1.44	1.53	7.66	6.06	0.40	2.85	1.27
	(0.19)	(2.76)	(1.28)	(0.31)	(3.95)	(2.76)	(1.06)	(4.68)	(4.09)	(0.22)	(1.77)	(0.64)
500	0.19	1.36	0.67	0.23	1.18	0.66	0.62	4.36	3.30	0.22	1.54	0.77
	(0.09)	(1.32)	(0.67)	(0.11)	(1.35)	(0.85)	(0.25)	(2.55)	(2.40)	(0.09)	(0.96)	(0.45)
1000	0.10	0.74	0.37	0.13	0.53	0.26	0.30	2.33	1.66	0.13	0.93	0.48
	(0.04)	(0.51)	(0.31)	(0.05)	(0.58)	(0.28)	(0.08)	(1.29)	(1.02)	(0.04)	(0.55)	(0.21)

TABLE 4. 100*MISE (with 100*std in parenthesis below) for the estimation of b_i, $i = 1,\dots,4$, 200 repetitions, $X \sim \mathcal{N}(0,1)$. CV and PCO are the two competing methods. Colum "Or" gives the average of ISE for the ratio of the two best estimators of bf and f in the collection.

n	b_1	b_2	b_3	b_4
250	0.13	0.13	0.06	0.10
	(0.02)	(0.03)	(0.01)	(0.02)
500	0.12	0.11	0.05	0.09
	(0.02)	(0.02)	(0.01)	(0.01)
1000	0.11	0.09	0.05	0.08
	(0.01)	(0.02)	(0.01)	(0.01)

TABLE 5. Mean of selected bandwidth (with std in parenthesis below) with the CV method for NW-single bandwidth estimator of b.

6. PROOFS

6.1. **Proof of Proposition 2.4.** On the one hand, by Comte [3], Proposition 3.3.1,

$$\mathbb{E}(\|\widehat{f}_{n,h'} - f\|_2^2) \leqslant \|f - f_{h'}\|_2^2 + \frac{\mathfrak{c}_K}{nh'} \tag{4}$$

and, by Proposition 2.2,

$$\mathbb{E}(\|\widehat{bf}_{n,h} - bf\|_2^2) \leqslant \|bf - (bf)_h\|_2^2 + \frac{\mathfrak{c}_{K,Y}}{nh}. \tag{5}$$

For the proof of Inequality (4), the reader can also refer to Tsybakov [15]. On the other hand,

$$\widehat{b}_{n,h,h'} - b = \left(\frac{\widehat{bf}_{n,h} - bf}{\widehat{f}_{n,h'}} + bf\left(\frac{1}{\widehat{f}_{n,h'}} - \frac{1}{f}\right)\right)\mathbf{1}_{\widehat{f}_{n,h'}(.)>m_n/2} - b\mathbf{1}_{\widehat{f}_{n,h'}(.)\leqslant m_n/2}.$$

Then,

$$\begin{aligned}\|\widehat{b}_{n,h,h'} - b\|_{2,f,\mathcal{S}_n}^2 \leqslant & \frac{8\mathfrak{c}_1}{m_n^2}\left(\|\widehat{bf}_{n,h} - bf\|_2^2 + \int_{-\infty}^{\infty} b(x)^2 f(x)|\widehat{f}_{n,h'}(x) - f(x)|^2 dx\right) \\ & +2\int_{\mathcal{S}_n} b(x)^2 f(x)\mathbf{1}_{|\widehat{f}_{n,h'}(x)-f(x)|>m_n/2}dx\end{aligned}$$

with $\mathfrak{c}_1 := \|f\|_\infty \vee \|f\|_\infty^2$.

By Markov's inequality,

$$\begin{aligned}\mathbb{E}(\|\widehat{b}_{n,h,h'}-b\|_{2,f,\mathcal{S}_n}^2) &\leqslant \frac{8\mathfrak{c}_1}{m_n^2}(\mathbb{E}(\|\widehat{bf}_{n,h}-bf\|_2^2)+\mathfrak{c}_{b,f}\mathbb{E}(\|\widehat{f}_{n,h'}-f\|_2^2))\\ &\quad +2\mathfrak{c}_{b,f}\int_{\mathcal{S}_n}\mathbb{P}\left(|\widehat{f}_{n,h'}(x)-f(x)|>\frac{m_n}{2}\right)dx\\ &\leqslant \frac{8(\mathfrak{c}_1\vee 1)}{m_n^2}(\mathbb{E}(\|\widehat{bf}_{n,h}-bf\|_2^2)+2\mathfrak{c}_{b,f}\mathbb{E}(\|\widehat{f}_{n,h'}-f\|_2^2)).\end{aligned}$$

Inequalities (4) and (5) allow to conclude.

6.2. **Proof of Theorem 3.2.** First, let us prove the following lemma.

Lemma 6.1. *Consider*

$$(bf)_h := K_h*(bf) \text{ and } (bf)_{h,\eta} := K_\eta*K_h*(bf).$$

Then,

$$\mathbb{E}(\widehat{bf}_{n,h}(x)) = (bf)_h(x) \text{ and } \mathbb{E}(\widehat{bf}_{n,h,\eta}(x)) = (bf)_{h,\eta}(x).$$

Proof. Since $\mathbb{E}(\varepsilon_k)=0$ and X_k and ε_k are independent for every $k\in\{1,\dots,n\}$,

$$\begin{aligned}\mathbb{E}(\widehat{bf}_{n,h}(x)) &= \frac{1}{n}\sum_{k=1}^n\mathbb{E}(b(X_k)K_h(X_k-x))+\frac{1}{n}\sum_{k=1}^n\mathbb{E}(\varepsilon_k)\mathbb{E}(K_h(X_k-x))\\ &= \int_{-\infty}^{\infty}K_h(y-x)b(y)f(y)dy=(bf)_h(x)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}(\widehat{bf}_{n,h,\eta}(x)) &= \frac{1}{n}\sum_{k=1}^n\mathbb{E}(b(X_k)(K_\eta*K_h)(X_k-x))+\frac{1}{n}\sum_{k=1}^n\mathbb{E}(\varepsilon_k)\mathbb{E}((K_\eta*K_h)(X_k-x))\\ &= \int_{-\infty}^{\infty}(K_\eta*K_h)(y-x)b(y)f(y)dy=(bf)_{h,\eta}(x).\end{aligned}$$

□

Since

$$\widehat{h}_n\in\arg\min_{h\in\mathcal{H}_n}\{A_n(h)+V_n(h)\},$$

for every $h\in\mathcal{H}_n$,

$$\mathbb{E}(\|\widehat{bf}_{n,\widehat{h}_n}-bf\|_2^2)\leqslant 3\mathbb{E}(\|\widehat{bf}_{n,h}-bf\|_2^2)+6V_n(h)+6\mathbb{E}(A_n(h)). \tag{6}$$

Let us find a suitable control of $\mathbb{E}(A_n(h))$. First of all, for any $h,\eta\in\mathcal{H}_n$,

$$\|\widehat{bf}_{n,h,\eta}-\widehat{bf}_{n,\eta}\|_2^2\leqslant 3(\|\widehat{bf}_{n,h,\eta}-(bf)_{h,\eta}\|_2^2+\|\widehat{bf}_{n,\eta}-(bf)_\eta\|_2^2+\|(bf)_{h,\eta}-(bf)_\eta\|_2^2).$$

Then,

$$\begin{aligned}A_n(h)\leqslant 3\Bigg[&\sup_{\eta\in\mathcal{H}_n}\left(\|\widehat{bf}_{n,h,\eta}-(bf)_{h,\eta}\|_2^2-\frac{V_n(\eta)}{6}\right)_+\\ &+\sup_{\eta\in\mathcal{H}_n}\left(\|\widehat{bf}_{n,\eta}-(bf)_\eta\|_2^2-\frac{V_n(\eta)}{6}\right)_+ +\|(bf)_{h,\eta}-(bf)_\eta\|_2^2\Bigg].\end{aligned} \tag{7}$$

On the one hand,

$$\|(bf)_{h,\eta}-(bf)_\eta\|_2=\|K_\eta*(K_h*(bf)-bf)\|_2\leqslant\|K\|_1\|bf-(bf)_h\|_2.$$

On the other hand, let $\mathcal{C}$ be a countable and dense subset of the unit sphere of $\mathbb{L}^2(\mathbb{R}, dx)$ and consider $\mathfrak{m}(n) > 0$. Then, by Lemma 6.1,

$$\mathbb{E}\left[\sup_{\eta\in\mathcal{H}_n}\left(\|\widehat{bf}_{n,\eta} - (bf)_\eta\|_2^2 - \frac{V_n(\eta)}{6}\right)_+\right] \leqslant \sum_{\eta\in\mathcal{H}_n}\mathbb{E}\left(\left(\sup_{\psi\in\mathcal{C}} 2\mathbf{V}_{n,\eta}(\psi)^2 - \frac{V_n(\eta)}{6}\right)_+\right) + 2\sum_{\eta\in\mathcal{H}_n}\mathbb{E}(\mathbf{W}_{n,\eta})$$

where, for any $\psi \in \mathcal{C}$,

$$\mathbf{V}_{n,\eta}(\psi) := \frac{1}{n}\sum_{k=1}^{n}(v_{\psi,n,\eta}(X_k, Y_k) - \mathbb{E}(v_{\psi,n,\eta}(X_k, Y_k)))$$

with

$$v_{\psi,n,\eta}(x, y) := y\mathbf{1}_{|y|\leqslant\mathfrak{m}(n)}\int_{-\infty}^{\infty}\psi(u)K_\eta(x - u)du \ ; \ \forall (x, y) \in \mathbb{R}^2,$$

and

$$\mathbf{W}_{n,\eta} := \frac{1}{n}\sum_{k=1}^{n}\int_{-\infty}^{\infty}|Y_k\mathbf{1}_{|Y_k|>\mathfrak{m}(n)}K_\eta(X_k - u) - \mathbb{E}(Y_k\mathbf{1}_{|Y_k|>\mathfrak{m}(n)}K_\eta(X_k - u))|^2 du.$$

In order to apply Talagrand's inequality (see Klein and Rio [9]):

- For every $\psi \in \mathcal{C}$, $x \in \mathbb{R}$ and $y \in [-\mathfrak{m}(n), \mathfrak{m}(n)]$,

$$\begin{aligned}|v_{\psi,n,\eta}(x, y)| &\leqslant |y|\int_{-\infty}^{\infty}|\psi(u)||K_\eta(u - x)|du\\ &\leqslant |y|\cdot\|K_\eta(\cdot - x)\|_2 \leqslant \frac{\mathfrak{m}(n)\|K\|_2}{\sqrt{\eta}}.\end{aligned}$$

  Then,

$$\sup_{\psi\in\mathcal{C}}\|v_{\psi,n,\eta}\|_\infty \leqslant \mathfrak{m}_1(n, \eta) := \frac{\mathfrak{m}(n)\|K\|_2}{\sqrt{\eta}}.$$

- By Proposition 2.4 and Lemma 6.1,

$$\begin{aligned}\mathbb{E}\left(\sup_{\psi\in\mathcal{C}}\mathbf{V}_{n,\eta}(\psi)^2\right) &\leqslant \int_{-\infty}^{\infty}\mathrm{var}(\widehat{bf}_{n,\eta}(u))du\\ &\leqslant \mathfrak{m}_2(n, \eta) := \frac{\mathfrak{c}_{K,Y}}{n\eta}.\end{aligned}$$

- For any $\psi \in \mathcal{C}$ and $k \in \{1, \dots, n\}$,

$$\begin{aligned}\mathrm{var}(v_{\psi,n,\eta}(X_k, Y_k)) &\leqslant \mathbb{E}\left(\left|Y_k\int_{-\infty}^{\infty}\psi(u)K_\eta(X_k - u)du\right|^2\right)\\ &\leqslant \mathbb{E}((K_\eta * \psi)(X_1)^4)^{1/2}\mathbb{E}(Y_1^4)^{1/2} \leqslant \|f\|_\infty^{1/2}\|K_\eta * \psi\|_4^2\mathbb{E}(Y_1^4)^{1/2}.\end{aligned}$$

  By Young's inequality, $\|K_\eta * \psi\|_4 \leqslant \|\psi\|_2\|K_\eta\|_{4/3}$. So,

$$\mathrm{var}(v_{\psi,n,\eta}(X_k, Y_k)) \leqslant \mathfrak{m}_3 := \frac{\mathfrak{m}_{f,K}}{\sqrt{\eta}}$$

  with $\mathfrak{m}_{f,K} := \|f\|_\infty^{1/2}\|K\|_{4/3}^2\mathbb{E}(Y_1^4)^{1/2}$.

By applying Talagrand's inequality to $(v_{\psi,n,\eta})_{\psi\in\mathcal{C}}$ and to the independent random variables $(X_1,Y_1),\dots,(X_n,Y_n)$, there exist three constants $\mathfrak{c}_1,\mathfrak{c}_2,\mathfrak{c}_3>0$, not depending on n and η, such that

$$\begin{aligned}\mathbb{E}\left(\left(\sup_{\psi\in\mathcal{C}}\mathbf{V}_{n,\eta}(\psi)^2-4\mathfrak{m}_2(n,\eta)\right)_+\right) &\leqslant \mathfrak{c}_1\left(\frac{\mathfrak{m}_3}{n}\exp\left(-\mathfrak{c}_2\frac{n\mathfrak{m}_2(n,\eta)}{\mathfrak{m}_3}\right)\right.\\ &\quad\left.+\frac{\mathfrak{m}_1(n,\eta)^2}{n^2}\exp\left(-\mathfrak{c}_3\frac{n\mathfrak{m}_2(n,\eta)^{1/2}}{\mathfrak{m}_1(n,\eta)}\right)\right)\\ &= \mathfrak{c}_1\left[\frac{\mathfrak{m}_{f,K}}{n\sqrt{\eta}}\exp\left(-\frac{\mathfrak{c}_2\mathfrak{c}_{K,Y}}{\mathfrak{m}_{f,K}\sqrt{\eta}}\right)\right.\\ &\quad\left.+\frac{1}{n^2\eta}\mathfrak{m}(n)^2\|K\|_2^2\exp\left(-\sqrt{n}\frac{\mathfrak{c}_3\mathbb{E}(Y_1^2)^{1/2}}{\mathfrak{m}(n)}\right)\right].\end{aligned}$$

By taking $\mathfrak{m}(n):=\mathfrak{c}_3\mathbb{E}(Y_1^2)^{1/2}n^{1/2}/\log(n)^{1/2}$,

$$\mathbb{E}\left(\left(\sup_{\psi\in\mathcal{C}}\mathbf{V}_{n,\eta}(\psi)^2-4\mathfrak{m}_2(n,\eta)\right)_+\right)\leqslant\frac{\mathfrak{c}_1}{n}\left[\frac{\mathfrak{m}_{f,K}}{\sqrt{\eta}}\exp\left(-\frac{\mathfrak{c}_2\mathfrak{c}_{K,Y}}{\mathfrak{m}_{f,K}\sqrt{\eta}}\right)+\frac{\mathfrak{c}_3^2\mathbb{E}(Y_1^2)\|K\|_2^2}{\eta n\log(n)}\right].$$

By the conditional Markov inequality,

$$\begin{aligned}\mathbb{E}(\mathbf{W}_{n,\eta}) &\leqslant \int_{-\infty}^{\infty}\mathbb{E}(Y_1^2\mathbf{1}_{|Y_1|>\mathfrak{m}(n)}K_\eta(X_1-z)^2)dz\\ &= \frac{\|K\|_2^2}{\eta}\cdot\mathbb{E}(Y_1^2\mathbb{E}(\mathbf{1}_{|Y_1|>\mathfrak{m}(n)}|Y_1))\\ &\leqslant \frac{\|K\|_2^2}{\eta\mathfrak{m}(n)^4}\mathbb{E}(Y_1^6)=\mathfrak{c}_3^{-4}\mathbb{E}(Y_1^2)^{-2}\mathbb{E}(Y_1^6)\|K\|_2^2\frac{\log(n)^2}{n^2\eta}.\end{aligned}$$

Finally, for $\upsilon\geqslant 48$,

$$\frac{V_n(\eta)}{12}\geqslant 4\mathfrak{m}_2(n,\eta).$$

Then, since

$$\frac{1}{n}\sum_{\eta\in\mathcal{H}_n}\frac{1}{\eta}\leqslant\mathfrak{m},\quad \sum_{\eta\in\mathcal{H}_n}\frac{1}{\sqrt{\eta}}\exp\left(-\frac{c}{\sqrt{\eta}}\right)\leqslant\mathfrak{m}(c)\ ;\ \forall c>0,$$

there exists a constant $\mathfrak{c}_4>0$, not depending on n, such that

$$\begin{aligned}\mathbb{E}\left[\sup_{\eta\in\mathcal{H}_n}\left(\|\widehat{bf}_{n,\eta}-(bf)_\eta\|_2^2-\frac{V_n(\eta)}{6}\right)_+\right] &\leqslant 2\sum_{\eta\in\mathcal{H}_n}\mathbb{E}\left[\left(\sup_{\psi\in\mathcal{C}}\mathbf{V}_{n,\eta}(\psi)^2-4\mathfrak{m}_2(n,\eta)\right)_+\right]+2\sum_{\eta\in\mathcal{H}_n}\mathbb{E}(\mathbf{W}_{n,\eta})\\ &\leqslant \mathfrak{c}_4\frac{\log(n)^2}{n}.\end{aligned}\tag{8}$$

The same ideas give that there exists a constant $\mathfrak{c}_5>0$, not depending on n and h, such that

$$\mathbb{E}\left[\sup_{\eta\in\mathcal{H}_n}\left(\|\widehat{bf}_{n,h,\eta}-(bf)_{h,\eta}\|_2^2-\frac{V_n(\eta)}{6}\right)_+\right]\leqslant\mathfrak{c}_5\frac{\log(n)^2}{n}.\tag{9}$$

Therefore, by Inequalities (6)–(9), there exist two deterministic constants $\mathfrak{c},\overline{\mathfrak{c}}>0$, not depending on n, such that

$$\mathbb{E}(\|\widehat{bf}_{n,\widehat{h}_n}-bf\|_2^2)\leqslant\mathfrak{c}\cdot\inf_{h\in\mathcal{H}_n}\{\|(bf)_h-bf\|_2^2+V_n(h)\}+\overline{\mathfrak{c}}\frac{\log(n)^2}{n}.$$

6.3. **Proof of Corollary 3.3.** As established in the proof of Proposition 2.4,

$$\begin{aligned}\|\widehat{b}_{n,\widehat{h}_n,\widehat{h}_n'}-b\|_{2,f,\mathcal{S}_n}^2 &\leqslant \frac{8\mathfrak{c}_1}{m_n^2}\left(\|\widehat{bf}_{n,\widehat{h}_n}-bf\|_2^2+\mathfrak{c}_{b,f}\int_{-\infty}^{\infty}|\widehat{f}_{n,\widehat{h}_n'}(x)-f(x)|^2dx\right)\\ &\quad+2\mathfrak{c}_{b,f}\int_{\mathcal{S}_n}\mathbf{1}_{|\widehat{f}_{n,\widehat{h}_n'}(x)-f(x)|>m_n/2}dx\end{aligned}$$

with $\mathfrak{c}_1 := \|f\|_\infty \vee \|f\|_\infty^2$. By Markov's inequality,

$$\begin{aligned}\mathbb{E}(\|\widehat{b}_{n,\widehat{h}_n,\widehat{h}'_n} - b\|_{2,f,\mathcal{S}_n}^2) &\leqslant \frac{8\mathfrak{c}_1}{m_n^2}(\mathbb{E}(\|\widehat{bf}_{n,\widehat{h}_n} - bf\|_2^2) + \mathfrak{c}_{b,f}\mathbb{E}(\|\widehat{f}_{n,\widehat{h}'_n} - f\|_2^2))\\ &\quad +2\mathfrak{c}_{b,f}\int_{\mathcal{S}_n}\mathbb{P}\left(|\widehat{f}_{n,\widehat{h}'_n}(x) - f(x)| > \frac{m_n}{2}\right)dx\\ &\leqslant \frac{8(\mathfrak{c}_1\vee 1)}{m_n^2}(\mathbb{E}(\|\widehat{bf}_{n,\widehat{h}_n} - bf\|_2^2) + 2\mathfrak{c}_{b,f}\mathbb{E}(\|\widehat{f}_{n,\widehat{h}'_n} - f\|_2^2)).\end{aligned}$$

Theorem 3.2 and Inequality (2) allow to conclude.

6.4. **Proof of Theorem 4.2.** The proof relies on three lemmas, which are stated first.

Lemma 6.2. *Consider the U-statistic*

$$U_n(h,h_{\min}) := \sum_{k\neq l}\langle Y_kK_h(X_k - \cdot) - (bf)_h, Y_lK_{h_{\min}}(X_l - \cdot) - (bf)_{h_{\min}}\rangle_2.$$

Under Assumption 2.1, if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|Y_1|)) < \infty$, then there exists a deterministic constant $\mathfrak{c}_U > 0$, not depending on n and $h_{\min}$, such that for every $\vartheta \in (0,1)$,

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\left\{\frac{|U_n(h,h_{\min})|}{n^2} - \frac{\vartheta\|K\|_2^2}{nh}\mathbb{E}(Y_1^2)\right\}\right) \leqslant \mathfrak{c}_U\frac{\log(n)^5}{\vartheta n}.$$

Lemma 6.3. *For every $\eta,\eta' \in \mathcal{H}_n$, consider*

$$V_n(\eta,\eta') := \langle\widehat{bf}_{n,\eta} - (bf)_{\eta'}, (bf)_{\eta'} - bf\rangle_2.$$

Under Assumption 2.1, if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|Y_1|)) < \infty$ and bf is bounded, then there exists a deterministic constant $\mathfrak{c}_V > 0$, not depending on n and $h_{\min}$, such that for every $\vartheta \in (0,1)$,

$$\mathbb{E}\left(\sup_{\eta,\eta'\in\mathcal{H}_n}\{|V_n(\eta,\eta')| - \vartheta\|(bf)_{\eta'} - bf\|_2^2\}\right) \leqslant \mathfrak{c}_V\frac{\log(n)^3}{\vartheta n}.$$

Lemma 6.4. *Under Assumption 2.1, if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|Y_1|)) < \infty$ and bf is bounded, then there exists a deterministic constant $\mathfrak{c}_L > 0$, not depending on n and $h_{\min}$, such that for every $\vartheta \in (0,1)$,*

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\left\{\|(bf)_h - bf\|_2^2 + \frac{\mathfrak{c}_{K,Y}}{nh} - \frac{1}{1-\vartheta}\|\widehat{bf}_{n,h} - bf\|_2^2\right\}\right) \leqslant \frac{\mathfrak{c}_L}{\vartheta(1-\vartheta)}\cdot\frac{\log(n)^5}{n}.$$

6.4.1. *Steps of the proof.* The proof of Theorem 4.2 is dissected in three steps.

Step 1. In this step, a suitable decomposition of

$$\|\widehat{bf}_{n,\widetilde{h}_n} - bf\|_2^2$$

is provided. On the one hand,

$$\begin{aligned}\|\widehat{bf}_{n,\widetilde{h}_n} - bf\|_2^2 + \mathrm{pen}(\widetilde{h}_n) &= \|\widehat{bf}_{n,\widetilde{h}_n} - \widehat{bf}_{n,h_{\min}}\|_2^2 + \mathrm{pen}(\widetilde{h}_n)\\ &\quad +\|\widehat{bf}_{n,h_{\min}} - bf\|_2^2\\ &\quad -\langle\widehat{bf}_{n,h_{\min}} - \widehat{bf}_{n,\widetilde{h}_n}, \widehat{bf}_{n,h_{\min}} - bf\rangle_2.\end{aligned}$$

Since

$$\widetilde{h}_n \in \arg\min_{h\in\mathcal{H}_n}\mathrm{crit}(h)$$

with

$$\mathrm{crit}(h) = \|\widehat{bf}_{n,h} - \widehat{bf}_{n,h_{\min}}\|_2^2 + \mathrm{pen}(h),$$

for any $h \in \mathcal{H}_n$,

$$\|\widehat{bf}_{n,\widetilde{h}_n} - bf\|_2^2 \leqslant \|\widehat{bf}_{n,h} - bf\|_2^2 + \mathrm{pen}(h) - 2\psi_n(h) - (\mathrm{pen}(\widetilde{h}_n) - 2\psi_n(\widetilde{h}_n)) \quad (10)$$

with

$$\psi_n := \langle \widehat{bf}_{n,h_{\min}} - bf, \widehat{bf}_{n,\cdot} - bf\rangle_2.$$

On the other hand,

$$\psi_n(h) = \psi_{1,n}(h) + \psi_{2,n}(h) + \psi_{3,n}(h),$$

where

$$\begin{aligned}
\psi_{1,n}(h) &:= \frac{\langle K_{h_{\min}}, K_h\rangle_2}{n^2}\sum_{k=1}^{n} Y_k^2 + \frac{U_n(h,h_{\min})}{n^2},\\
\psi_{2,n}(h) &:= -\frac{1}{n^2}\left(\sum_{k=1}^{n} Y_k\langle K_{h_{\min}}(X_k-\cdot),(bf)_h\rangle_2\right.\\
&\quad \left.+\sum_{k=1}^{n} Y_k\langle K_h(X_k-\cdot),(bf)_{h_{\min}}\rangle_2\right) + \frac{1}{n}\langle (bf)_{h_{\min}},(bf)_h\rangle_2,\\
\psi_{3,n}(h) &:= V_n(h,h_{\min}) + V_n(h_{\min},h) + \langle (bf)_h - bf,(bf)_{h_{\min}} - bf\rangle_2.
\end{aligned}$$

Step 2. In this step, let us provide some suitable controls of

$$\mathbb{E}(\psi_{i,n}(h)) \text{ and } \mathbb{E}(\psi_{i,n}(\widetilde{h}_n))\ ;\ i = 1,2,3.$$

(1) Consider

$$\widetilde{\psi}_{1,n}(h) := \psi_{1,n}(h) - \frac{\langle K_{h_{\min}}, K_h\rangle_2}{n^2}\sum_{k=1}^{n} Y_k^2 = \frac{U(h,h_{\min})}{n^2}.$$

By Lemma 6.2,

$$\mathbb{E}(|\widetilde{\psi}_{1,n}(h)|) \leqslant \frac{\theta\|K\|_2^2}{nh}\mathbb{E}(Y_1^2) + \frac{2\mathfrak{c}_U}{\theta}\cdot\frac{\log(n)^5}{n}.$$

and

$$\mathbb{E}(|\widetilde{\psi}_{1,n}(\widetilde{h}_n)|) \leqslant \mathbb{E}\left(\frac{\theta\|K\|_2^2}{n\widetilde{h}_n}\right)\mathbb{E}(Y_1^2) + \frac{2\mathfrak{c}_U}{\theta}\cdot\frac{\log(n)^5}{n}.$$

(2) On the one hand, for every $\eta,\eta'\in\mathcal{H}_n$, consider

$$\Psi_{2,n}(\eta,\eta') := \frac{1}{n}\sum_{k=1}^{n} Y_k\langle K_\eta(X_k-\cdot),(bf)_{\eta'}\rangle_2.$$

Then,

$$\begin{aligned}
\mathbb{E}\left(\sup_{\eta,\eta'\in\mathcal{H}_n}|\Psi_{2,n}(\eta,\eta')|\right) &\leqslant \mathbb{E}\left(|Y_1|\sup_{\eta,\eta'\in\mathcal{H}_n}\int_{-\infty}^{\infty}|K_\eta(X_1-u)(bf)_{\eta'}(u)|du\right)\\
&\leqslant \mathbb{E}(Y_1^2)^{1/2}\|K\|_1^2\|bf\|_\infty.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\sup_{\eta,\eta'\in\mathcal{H}_n}|\langle (bf)_\eta,(bf)_{\eta'}\rangle_2| &\leqslant \sup_{\eta,\eta'\in\mathcal{H}_n}\|K_\eta*(bf)\|_2\|K_{\eta'}*(bf)\|_2\\
&\leqslant \|K\|_1^2\|bf\|_2^2 \leqslant \mathbb{E}(Y_1^2)^{1/2}\|K\|_1^2\|bf\|_\infty.
\end{aligned}$$

Then,

$$\mathbb{E}(|\psi_{2,n}(h)|) \leqslant \frac{3}{n}\mathbb{E}(Y_1^2)^{1/2}\|K\|_1^2\|bf\|_\infty$$

and

$$\mathbb{E}(|\psi_{2,n}(\widetilde{h}_n)|) \leqslant \frac{3}{n}\mathbb{E}(Y_1^2)^{1/2}\|K\|_1^2\|bf\|_\infty.$$

(3) By Lemma 6.3,

$$\begin{aligned}\mathbb{E}(|\psi_{n,3}(h)|) &\leqslant \frac{\theta}{2}(\|(bf)_h - bf\|_2^2 + \|(bf)_{h_{\min}} - bf\|_2^2) + 4\mathfrak{c}_V\frac{\log(n)^3}{\theta n}\\ &+ \left(\frac{\theta}{2}\right)^{1/2}\|(bf)_h - bf\|_2 \times \left(\frac{2}{\theta}\right)^{1/2}\|(bf)_{h_{\min}} - bf\|_2\\ &\leqslant \theta\|(bf)_h - bf\|_2^2 + \left(\frac{\theta}{2} + \frac{2}{\theta}\right)\|(bf)_{h_{\min}} - bf\|_2^2 + 4\mathfrak{c}_V\frac{\log(n)^3}{\theta n}\end{aligned}$$

and

$$\mathbb{E}(|\psi_{n,3}(\widetilde{h}_n)|) \leqslant \theta\mathbb{E}(\|(bf)_{\widetilde{h}_n} - bf\|_2^2) + \left(\frac{\theta}{2} + \frac{2}{\theta}\right)\|(bf)_{h_{\min}} - bf\|_2^2 + 4\mathfrak{c}_V\frac{\log(n)^3}{\theta n}.$$

Step 3. Consider

$$\widetilde{\psi}_n(h) := \psi_n(h) - \frac{\langle K_{h_{\min}}, K_h\rangle_2}{n^2}\sum_{k=1}^{n} Y_k^2.$$

By Step 2, there exists a deterministic constant $\mathfrak{c}_{U,V} > 0$, not depending on n, h and $h_{\min}$, such that

$$\begin{aligned}\mathbb{E}(|\widetilde{\psi}_n(h)|) &\leqslant \theta\left(\|(bf)_h - bf\|_2^2 + \frac{\mathfrak{c}_{K,Y}}{nh}\right)\\ &+\frac{\mathfrak{c}_{U,V}}{\theta}\cdot\frac{\log(n)^5}{n} + \left(\frac{\theta}{2} + \frac{2}{\theta}\right)\|(bf)_{h_{\min}} - bf\|_2^2\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}(|\widetilde{\psi}_n(\widetilde{h}_n)|) &\leqslant \theta\left[\mathbb{E}(\|(bf)_{\widetilde{h}_n} - bf\|_2^2) + \mathbb{E}\left(\frac{\mathfrak{c}_{K,Y}}{n\widetilde{h}_n}\right)\right]\\ &+\frac{\mathfrak{c}_{U,V}}{\theta}\cdot\frac{\log(n)^5}{n} + \left(\frac{\theta}{2} + \frac{2}{\theta}\right)\|(bf)_{h_{\min}} - bf\|_2^2.\end{aligned}$$

Then, by Lemma 6.4,

$$\begin{aligned}\mathbb{E}(|\widetilde{\psi}_n(h)|) &\leqslant \frac{\theta}{1-\theta}\mathbb{E}(\|\widehat{bf}_{n,h} - bf\|_2^2) + \left(\frac{\theta}{2} + \frac{2}{\theta}\right)\|(bf)_{h_{\min}} - bf\|_2^2\\ &+\left(\frac{\mathfrak{c}_{U,V}}{\theta} + \frac{\mathfrak{c}_L}{1-\theta}\right)\frac{\log(n)^5}{n}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}(|\widetilde{\psi}_n(\widetilde{h}_n)|) &\leqslant \frac{\theta}{1-\theta}\mathbb{E}(\|\widehat{bf}_{n,\widetilde{h}_n} - bf\|_2^2) + \left(\frac{\theta}{2} + \frac{2}{\theta}\right)\|(bf)_{h_{\min}} - bf\|_2^2\\ &+\left(\frac{\mathfrak{c}_{U,V}}{\theta} + \frac{\mathfrak{c}_L}{1-\theta}\right)\frac{\log(n)^5}{n}.\end{aligned}$$

By Inequality (10), there exist two deterministic constant $\mathfrak{c}_1, \mathfrak{c}_2 > 0$, not depending on n, h and $h_{\min}$, such that

$$\begin{aligned}\mathbb{E}(\|\widehat{bf}_{n,\widetilde{h}_n} - bf\|_2^2) &\leqslant \mathbb{E}(\|\widehat{bf}_{n,h} - bf\|_2^2) + 2(\mathbb{E}(|\widetilde{\psi}_n(h)|) + \mathbb{E}(|\widetilde{\psi}_n(\widetilde{h}_n)|))\\ &\leqslant \left(1 + \frac{2\theta}{1-\theta}\right)\mathbb{E}(\|\widehat{bf}_{n,h} - bf\|_2^2) + \frac{2\theta}{1-\theta}\mathbb{E}(\|\widehat{bf}_{n,\widetilde{h}_n} - bf\|_2^2)\\ &+\frac{\mathfrak{c}_1}{\theta}\|(bf)_{h_{\min}} - bf\|_2^2 + \frac{\mathfrak{c}_2}{\theta(1-\theta)}\cdot\frac{\log(n)^5}{n}.\end{aligned}$$

This concludes the proof.

6.4.2. *Proof of Lemma 6.2.* Consider

$$\Delta_n := \{(k,l) \in \{1,\dots,n\} : 2 \leqslant k \text{ and } l < k\}$$

and $Z_k := (X_k, Y_k)$ for every $k \in \{1,\dots,n\}$.

On the one hand, consider $n \in \mathbb{N}$ such that $\mathfrak{m}(n) := 4\log(n)/\alpha \geqslant 1$ and

$$U_{1,n}(h, h_{\min}) := \sum_{k=2}^{n}\sum_{l<k}(G_{n,h,h_{\min}}(Z_k, Z_l) + G_{n,h_{\min},h}(Z_k, Z_l))$$

where, for every $\eta, \eta' \in \{h, h_{\min}\}$ and $z, z' \in \mathbb{R}^2$,

$$\begin{aligned} G_{n,\eta,\eta'}(z,z') &:= \langle z_2\mathbf{1}_{|z_2|\leqslant\mathfrak{m}(n)}K_\eta(z_1 - \cdot) - (bf)_{n,\eta}, \\ &\quad z_2'\mathbf{1}_{|z_2'|\leqslant\mathfrak{m}(n)}K_{\eta'}(z_1' - \cdot) - (bf)_{n,\eta'}\rangle_2 \end{aligned}$$

and

$$(bf)_{n,\eta} := \mathbb{E}(Y_1\mathbf{1}_{|Y_1|\leqslant\mathfrak{m}(n)}K_\eta(X_1 - \cdot)).$$

For every $\eta, \eta' \in \{h, h_{\min}\}$ and $(k,l) \in \Delta_n$,

$$\begin{aligned} \mathbb{E}(G_{n,\eta,\eta'}(Z_k, Z_l)|Z_k) &= \int_{-\infty}^{\infty}(Y_k\mathbf{1}_{|Y_k|\leqslant\mathfrak{m}(n)}K_\eta(X_k - z) - (bf)_{n,\eta}(z)) \\ &\quad \times\mathbb{E}(Y_l\mathbf{1}_{|Y_l|\leqslant\mathfrak{m}(n)}K_{\eta'}(X_l - z) - (bf)_{n,\eta'}(z))dz = 0. \end{aligned}$$

So, by Houdré and Reynaud-Bouret [8], Theorem 3.4, there exists a universal constant $\mathfrak{e} > 0$ such that

$$\mathbb{P}(|U_{1,n}(h, h_{\min})| \geqslant \mathfrak{e}(\mathfrak{c}_n\lambda^{1/2} + \mathfrak{d}_n\lambda + \mathfrak{b}_n\lambda^{3/2} + \mathfrak{a}_n\lambda^2)) \leqslant 5.54e^{-\lambda}, \tag{11}$$

where the constants $\mathfrak{a}_n$, $\mathfrak{b}_n$, $\mathfrak{c}_n$ and $\mathfrak{d}_n$ will be defined and controlled in the sequel.

- **The constant $\mathfrak{a}_n$.** Consider

$$\mathfrak{a}_n := \sup_{(z,z')\in\mathbb{R}^2\times\mathbb{R}^2}\mathbf{A}_n(z,z'),$$

where

$$\mathbf{A}_n(z,z') := |G_{n,h,h_{\min}}(z,z') + G_{n,h_{\min},h}(z,z')| \ ; \ \forall z, z' \in \mathbb{R}^2.$$

First, note that for every $\eta \in \mathcal{H}_n$,

$$\|(bf)_{n,\eta}\|_1 \leqslant \mathbb{E}(|Y_1|\mathbf{1}_{|Y_1|\leqslant\mathfrak{m}(n)})\|K\|_1 \leqslant \mathfrak{m}(n)\|K\|_1$$

and

$$\|(bf)_{n,\eta}\|_\infty \leqslant \frac{\mathfrak{m}(n)\|K\|_\infty}{\eta}.$$

For any $z, z' \in \mathbb{R}\times[-\mathfrak{m}(n), \mathfrak{m}(n)]$,

$$\begin{aligned} \mathbf{A}_n(z,z') &\leqslant \langle z_2K_h(z_1 - \cdot) - (bf)_{n,h}, z_2'K_{h_{\min}}(z_1' - \cdot) - (bf)_{n,h_{\min}}\rangle_2 \\ &\quad +\langle z_2K_{h_{\min}}(z_1 - \cdot) - (bf)_{n,h_{\min}}, z_2'K_h(z_1' - \cdot) - (bf)_{n,h}\rangle_2 \\ &\leqslant 2(\mathfrak{m}(n)\|K_{h_{\min}}\|_\infty + \|(bf)_{n,h_{\min}}\|_\infty)(\mathfrak{m}(n)\|K\|_1 + \|(bf)_{n,h}\|_1) \\ &\leqslant \frac{8\|K\|_1\|K\|_\infty}{h_{\min}}\mathfrak{m}(n)^2. \end{aligned}$$

Therefore,

$$\frac{\mathfrak{a}_n\lambda^2}{n^2} \leqslant \frac{8\|K\|_1\|K\|_\infty}{n^2h_{\min}}\mathfrak{m}(n)^2\lambda^2.$$

- **The constant $\mathfrak{b}_n$.** Consider

$$\mathfrak{b}_n^2 := n\max\left\{\sup_{z\in\mathbb{R}^2}\mathbb{E}(G_{n,h,h_{\min}}(z, Z_1)^2) \ ; \ \sup_{z\in\mathbb{R}^2}\mathbb{E}(G_{n,h_{\min},h}(z, Z_1)^2)\right\}.$$

First, note that for every $\eta \in \mathcal{H}_n$,

$$\|(bf)_{n,\eta}\|_2^2 \leqslant \frac{\mathfrak{m}(n)^2\|K\|_2^2}{\eta}.$$

For any $\eta,\eta'\in\{h,h_{\min}\}$ and $z\in\mathbb{R}\times[-\mathfrak{m}(n),\mathfrak{m}(n)]$,

$$\begin{aligned}
\mathbb{E}(G_{n,\eta,\eta'}(z,Z_1)^2) &\leqslant \|z_2K_\eta(z_1-\cdot)-(bf)_{n,\eta}\|_2^2\\
&\quad\times\int_{-\infty}^{\infty}\mathbb{E}(|Y_1\mathbf{1}_{|Y_1|\leqslant\mathfrak{m}(n)}K_{\eta'}(X_1-u)-(bf)_{n,\eta'}(u)|^2)du\\
&\leqslant \frac{2\|K\|_2^2}{\eta}\mathfrak{m}(n)^2\int_{-\infty}^{\infty}\mathrm{var}(Y_1\mathbf{1}_{|Y_1|\leqslant\mathfrak{m}(n)}K_{\eta'}(X_1-u))du\\
&\leqslant \frac{2\|K\|_2^4}{\eta\eta'}\mathbb{E}(Y_1^2)\mathfrak{m}(n)^2.
\end{aligned}$$

Therefore, for any $\theta\in(0,1)$,

$$\begin{aligned}
\frac{\mathfrak{b}_n\lambda^{3/2}}{n^2} &\leqslant \sqrt{2}\cdot\frac{\|K\|_2^2}{n^{3/2}(hh_{\min})^{1/2}}\mathbb{E}(Y_1^2)^{1/2}\mathfrak{m}(n)\lambda^{3/2}\\
&= 2\left(\frac{3\mathfrak{c}}{\theta}\right)^{1/2}\frac{\|K\|_2}{nh_{\min}^{1/2}}\mathfrak{m}(n)\lambda^{3/2}\times\left(\frac{\theta}{3\mathfrak{c}}\right)^{1/2}\frac{\|K\|_2}{n^{1/2}h^{1/2}}\mathbb{E}(Y_1^2)^{1/2}\\
&\leqslant \frac{3\mathfrak{c}\|K\|_2^2}{\theta n^2h_{\min}}\mathfrak{m}(n)^2\lambda^3+\frac{\theta\|K\|_2^2}{3\mathfrak{c}nh}\mathbb{E}(Y_1^2).
\end{aligned}$$

- **The constant $\mathfrak{c}_n$.** Consider

$$\mathfrak{c}_n^2:=\sum_{(k,l)\in\Delta_n}\mathbb{E}(|G_{n,h,h_{\min}}(Z_k,Z_l)+G_{n,h_{\min},h}(Z_k,Z_l)|^2).$$

First, note that for every $\eta\in\mathcal{H}_n$,

$$\|(bf)_{n,\eta}\|_\infty\leqslant\mathfrak{m}(n)\|f\|_\infty\|K\|_1.$$

For any $\eta,\eta'\in\{h,h_{\min}\}$ and $(k,l)\in\Delta_n$,

$$\begin{aligned}
\mathbb{E}(G_{n,\eta,\eta'}(Z_k,Z_l)^2) &\leqslant 4(\mathfrak{m}(n)^2\mathbb{E}(\langle K_\eta(X_k-\cdot),K_{\eta'}(X_l-\cdot)\rangle_2^2Y_l^2)\\
&\quad+\|(bf)_{n,\eta}\|_\infty^2\mathbb{E}(Y_l^2\|K_{\eta'}(X_l-\cdot)\|_1^2)\\
&\quad+\|(bf)_{n,\eta'}\|_\infty^2\mathbb{E}(Y_k^2\|K_\eta(X_k-\cdot)\|_1^2)\\
&\quad+\|(bf)_{n,\eta}\|_\infty^2\|(bf)_{n,\eta'}\|_1^2)\\
&\leqslant 4\mathfrak{m}(n)^2(\mathbb{E}(\langle K_\eta(X_k-\cdot),K_{\eta'}(X_l-\cdot)\rangle_2^2Y_l^2)\\
&\quad+3\|f\|_\infty^2\|K\|_1^4\mathbb{E}(Y_1^2)).
\end{aligned}$$

Moreover,

$$\begin{aligned}
\mathbb{E}(\langle K_\eta(X_k-\cdot),K_{\eta'}(X_l-\cdot)\rangle_2^2Y_l^2) &= \sigma^2\mathbb{E}((K_\eta*K_{\eta'})(X_k-X_l)^2)\\
&\quad+\mathbb{E}((K_\eta*K_{\eta'})(X_k-X_l)^2b(X_l)^2)\\
&\leqslant \sigma^2\|f\|_\infty\|K_\eta*K_{\eta'}\|_2^2\\
&\quad+\|f\|_\infty\mathbb{E}(b(X_1)^2)\|K_\eta*K_{\eta'}\|_2^2\\
&\leqslant \frac{\|f\|_\infty\|K\|_1^2\|K\|_2^2}{\eta}\mathbb{E}(Y_1^2).
\end{aligned}$$

Then, there exists a universal constant $\mathfrak{c}_1>0$ such that

$$\mathfrak{c}_n^2\leqslant\mathfrak{c}_1n^2\|f\|_\infty\|K\|_1^2\mathfrak{m}(n)^2\mathbb{E}(Y_1^2)\left(\frac{\|K\|_2^2}{h}+3\|f\|_\infty\|K\|_1^2\right).$$

Therefore, since $\mathfrak{m}(n)$ is larger than 1, there exists a universal constant $\mathfrak{c}_2>0$ such that

$$\frac{\mathfrak{c}_n\lambda^{1/2}}{n^2}\leqslant\frac{\theta\|K\|_2^2}{3\mathfrak{c}nh}\mathbb{E}(Y_1^2)+\frac{\mathfrak{c}_2}{n\theta}\|f\|_\infty\|K\|_1^2\mathfrak{m}(n)^2(\lambda^{1/2}+\lambda).$$

- **The constant $\mathfrak{d}_n$.** Consider

$$\mathfrak{d}_n := \sup_{(\alpha,\beta)\in\mathcal{S}}\sum_{(k,l)\in\Delta_n}\mathbb{E}((G_{h,h_{\min}}(Z_k,Z_l)+G_{h_{\min},h}(Z_k,Z_l))\alpha_k(Z_k)\beta_l(Z_l)),$$

where

$$\mathcal{S} := \left\{(\alpha,\beta) : \sum_{k=2}^{n}\mathbb{E}(\alpha_k(Z_k)^2)\leqslant 1 \text{ and } \sum_{l=1}^{n-1}\mathbb{E}(\beta_l(Z_l)^2)\leqslant 1\right\}.$$

For any $(\alpha,\beta)\in\mathcal{S}$,

$$\sum_{(k,l)\in\Delta_n}\mathbb{E}(G_{h,h_{\min}}(Z_k,Z_l)\alpha_k(Z_k)\beta_l(Z_l))\leqslant \mathbf{D}_2(\alpha,\beta)\sup_{u\in\mathbb{R}}\mathbf{D}_1(\alpha,\beta,u)$$

with, for every $u\in\mathbb{R}$,

$$\begin{aligned}
\mathbf{D}_1(\alpha,\beta,u) &:= \sum_{k=2}^{n}\mathbb{E}(|\alpha_k(Z_k)(Y_k\mathbf{1}_{|Y_k|\leqslant\mathfrak{m}(n)}K_h(X_k-u)-(bf)_{n,h}(u))|)\\
&\leqslant \mathbb{E}\left[\left(\sum_{k=2}^{n}\alpha_k(Z_k)^2\right)^{1/2}\right.\\
&\qquad\left.\times\left(\sum_{k=2}^{n}|Y_k\mathbf{1}_{|Y_k|\leqslant\mathfrak{m}(n)}K_h(X_k-u)-(bf)_{n,h}(u)|^2\right)^{1/2}\right]\\
&\leqslant \left(\sum_{k=2}^{n}\mathbb{E}(\alpha_k(Z_k)^2)\right)^{1/2}\left(\sum_{k=2}^{n}\mathbb{E}(Y_k^2\mathbf{1}_{|Y_k|\leqslant\mathfrak{m}(n)}K_h(X_k-u)^2)\right)^{1/2}\\
&\leqslant \frac{\|f\|_{\infty}^{1/2}\|K\|_2}{h^{1/2}}n^{1/2}\mathfrak{m}(n)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{D}_2(\alpha,\beta) &:= \sum_{l=1}^{n-1}\mathbb{E}\left(|\beta_l(Z_l)|\int_{-\infty}^{\infty}|Y_l\mathbf{1}_{|Y_l|\leqslant\mathfrak{m}(n)}K_{h_{\min}}(X_l-u)-(bf)_{n,h_{\min}}(u)|du\right)\\
&\leqslant \sqrt{2}\left(\sum_{l=1}^{n-1}\mathbb{E}(\beta_l(Z_l)^2)\right)^{1/2}\\
&\qquad\times\left(\sum_{l=1}^{n-1}[\mathbb{E}(Y_l^2\|K_{h_{\min}}(X_l-\cdot)\|_1^2)+\|(bf)_{n,h_{\min}}\|_1^2]\right)^{1/2}\\
&\leqslant \sqrt{2}\cdot\|K\|_1\mathbb{E}(Y_1^2)^{1/2}n^{1/2}.
\end{aligned}$$

Then,

$$\mathfrak{d}_n\leqslant 2n\frac{\|K\|_2\|K\|_1\|f\|_{\infty}^{1/2}}{h^{1/2}}\mathbb{E}(Y_1^2)^{1/2}\mathfrak{m}(n).$$

Therefore,

$$\begin{aligned}
\frac{\mathfrak{d}_n\lambda}{n^2} &\leqslant 2\times\left(\frac{\theta}{3\mathfrak{c}}\right)^{1/2}\frac{\|K\|_2}{(nh)^{1/2}}\mathbb{E}(Y_1^2)^{1/2}\times\left(\frac{3\mathfrak{c}}{\theta}\right)^{1/2}\frac{\|K\|_1\|f\|_{\infty}^{1/2}}{n^{1/2}}\mathfrak{m}(n)\lambda\\
&\leqslant \frac{\theta\|K\|_2^2}{3\mathfrak{c}nh}\mathbb{E}(Y_1^2)+\frac{3\mathfrak{c}\|K\|_1^2\|f\|_{\infty}}{\theta n}\mathfrak{m}(n)^2\lambda^2.
\end{aligned}$$

So, there exist two universal constants $\mathfrak{c}_3, \mathfrak{c}_4 > 0$ such that with probability larger than $1 - 5.54e^{-\lambda}$,

$$\begin{aligned}\frac{|U_{1,n}(h, h_{\min})|}{n^2} &\leqslant \frac{\theta\|K\|_2^2}{nh}\mathbb{E}(Y_1^2) \\ &+\mathfrak{c}_3\left(\frac{\|K\|_1\|K\|_\infty}{n^2 h_{\min}}\mathfrak{m}(n)^2\left(\frac{\lambda^3}{\theta} + \lambda^2\right)\right. \\ &\left.+\frac{\|f\|_\infty\|K\|_1^2}{n\theta}\mathfrak{m}(n)^2(\lambda^2 + \lambda + \lambda^{1/2})\right) \\ &\leqslant \frac{\theta\|K\|_2^2}{nh}\mathbb{E}(Y_1^2) + \frac{\mathfrak{c}_4}{\theta}\left(\frac{\|K\|_1\|K\|_\infty}{n^2 h_{\min}} + \frac{\|f\|_\infty\|K\|_1^2}{n}\right)\mathfrak{m}(n)^2(1+\lambda)^3.\end{aligned}$$

Then, with probability larger than $1 - 5.54|\mathcal{H}_n|e^{-\lambda}$,

$$S_n(h_{\min}) \leqslant \frac{\mathfrak{c}_4}{\theta}\left(\frac{\|K\|_1\|K\|_\infty}{n^2 h_{\min}} + \frac{\|f\|_\infty\|K\|_1^2}{n}\right)\mathfrak{m}(n)^2(1+\lambda)^3$$

where

$$S_n(h_{\min}) := \sup_{h\in\mathcal{H}_n}\left\{\frac{|U_{1,n}(h, h_{\min})|}{n^2} - \frac{\theta\|K\|_2^2}{nh}\mathbb{E}(Y_1^2)\right\}.$$

For every $s \in \mathbb{R}_+$, consider

$$\lambda(s) := -1 + \left(\frac{s}{\mathfrak{m}(n, h_{\min}, \theta)}\right)^{1/3}$$

where

$$\mathfrak{m}(n, h_{\min}, \theta) := \frac{\mathfrak{c}_4}{\theta}\left(\frac{\|K\|_1\|K\|_\infty}{n^2 h_{\min}} + \frac{\|f\|_\infty\|K\|_1^2}{n}\right)\mathfrak{m}(n)^2.$$

Then, for any $A > 0$,

$$\begin{aligned}\mathbb{E}(S_n(h_{\min})) &\leqslant 2A + \int_A^\infty \mathbb{P}(S_n(h_{\min}) \geqslant s)ds \\ &\leqslant 2A + 5.54\mathfrak{c}_5|\mathcal{H}_n|\mathfrak{m}(n, h_{\min}, \theta)\exp\left(-\frac{A^{1/3}}{2\mathfrak{m}(n, h_{\min}, \theta)^{1/3}}\right)\end{aligned}$$

where

$$\mathfrak{c}_5 := \int_0^\infty e^{1-s^{1/3}/2}ds.$$

Since there exists a deterministic constant $\mathfrak{c}_6 > 0$, not depending on n and $h_{\min}$ such that

$$\mathfrak{m}(n, h_{\min}, \theta) \leqslant \mathfrak{c}_6\frac{\log(n)^2}{n},$$

by taking $A := 2^3\mathfrak{c}_6\log(n)^5/n$,

$$\mathbb{E}(S_n(h_{\min})) \leqslant 2^4\mathfrak{c}_6\frac{\log(n)^5}{n} + 5.54\mathfrak{c}_5\mathfrak{m}(n, h_{\min}, \theta)\frac{|\mathcal{H}_n|}{n}.$$

Therefore, since $|\mathcal{H}_n| \leqslant n$, there exists a deterministic constant $\mathfrak{c}_7 > 0$, not depending on n and $h_{\min}$, such that

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\left\{\frac{|U_{1,n}(h, h_{\min})|}{n^2} - \frac{\theta\|K\|_2^2}{nh}\mathbb{E}(Y_1^2)\right\}\right) \leqslant \frac{\mathfrak{c}_5}{\theta}\cdot\frac{\log(n)^5}{n}.$$

On the other hand,

$$U_n(h, h_{\min}) = \sum_{i=1}^4 U_{i,n}(h, h_{\min})$$

where, for $i = 2, 3, 4$,

$$U_{i,n}(h, h_{\min}) := \sum_{k\neq l} g^i_{n,h,h_{\min}}(Z_k, Z_l)$$

with

$$\begin{aligned} g_{n,h,h_{\min}}^2(z,z') &:= \langle z_2\mathbf{1}_{|z_2|\leqslant\mathfrak{m}(n)}K_h(z_1-\cdot),z_2'\mathbf{1}_{|z_2'|>\mathfrak{m}(n)}K_{h_{\min}}(z_1'-\cdot)\rangle_2,\\ g_{n,h,h_{\min}}^3(z,z') &:= \langle z_2\mathbf{1}_{|z_2|>\mathfrak{m}(n)}K_h(z_1-\cdot),z_2'\mathbf{1}_{|z_2'|\leqslant\mathfrak{m}(n)}K_{h_{\min}}(z_1'-\cdot)\rangle_2 \text{ and}\\ g_{n,h,h_{\min}}^4(z,z') &:= \langle z_2\mathbf{1}_{|z_2|>\mathfrak{m}(n)}K_h(z_1-\cdot),z_2'\mathbf{1}_{|z_2'|>\mathfrak{m}(n)}K_{h_{\min}}(z_1'-\cdot)\rangle_2 \end{aligned}$$

for every $z,z'\in\mathbb{R}^2$. Consider $k,l\in\{1,\dots,n\}$ such that $k\neq l$. By Markov's inequality,

$$\begin{aligned} \mathbb{E}\left(\sup_{h\in\mathcal{H}_n}|g_{n,h,h_{\min}}^2(Z_k,Z_l)|\right) &\leqslant \mathfrak{m}(n)\sum_{h\in\mathcal{H}_n}\int_{-\infty}^{\infty}\mathbb{E}(|K_h(X_k-u)|)\\ &\quad\times\mathbb{E}(|Y_l|\mathbf{1}_{|Y_l|>\mathfrak{m}(n)}|K_{h_{\min}}(X_l-u)|)du\\ &\leqslant \mathfrak{m}(n)|\mathcal{H}_n|\cdot\|f\|_\infty\|K\|_1^2\\ &\quad\times\mathbb{E}(Y_1^2)^{1/2}\mathbb{P}(\exp(\alpha|Y_1|)>n^4)^{1/2}\\ &\leqslant \|f\|_\infty\|K\|_1^2\mathbb{E}(Y_1^2)^{1/2}\mathbb{E}(\exp(\alpha|Y_1|))^{1/2}\frac{\mathfrak{m}(n)}{n^2}|\mathcal{H}_n|. \end{aligned}$$

Then, there exists a deterministic constant $\mathfrak{c}_7>0$, not depending on n and $h_{\min}$, such that

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\frac{|U_{2,n}(h,h_{\min})|}{n^2}\right)\leqslant\mathfrak{c}_7\frac{\log(n)}{n}.$$

The same ideas give that there exists a deterministic constant $\mathfrak{c}_8>0$, not depending on n and $h_{\min}$, such that

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\frac{|U_{3,n}(h,h_{\min})|}{n^2}\right)\leqslant\mathfrak{c}_8\frac{\log(n)}{n}.$$

For $i=4$, by Markov's inequality,

$$\begin{aligned} \mathbb{E}\left(\sup_{h\in\mathcal{H}_n}|g_{n,h,h_{\min}}^4(Z_k,Z_l)|\right) &\leqslant \sum_{h\in\mathcal{H}_n}\int_{-\infty}^{\infty}\mathbb{E}(|Y_k|\mathbf{1}_{|Y_k|>\mathfrak{m}(n)}|K_h(X_k-u)|)\\ &\quad\times\mathbb{E}(|Y_l|\mathbf{1}_{|Y_l|>\mathfrak{m}(n)}|K_{h_{\min}}(X_l-u)|)du\\ &\leqslant \frac{\|K\|_\infty}{h_{\min}}\mathbb{E}(|Y_l|\mathbf{1}_{|Y_l|>\mathfrak{m}(n)})\\ &\quad\times\sum_{h\in\mathcal{H}_n}\int_{-\infty}^{\infty}\mathbb{E}(|Y_k|\mathbf{1}_{|Y_k|>\mathfrak{m}(n)}|K_h(X_k-u)|)du\\ &\leqslant \frac{\|K\|_\infty\|K\|_1}{h_{\min}}|\mathcal{H}_n|\cdot\mathbb{E}(Y_1^2)\mathbb{P}(|Y_1|>\mathfrak{m}(n))\\ &\leqslant \|K\|_\infty\|K\|_1\mathbb{E}(Y_1^2)\mathbb{E}(\exp(\alpha|Y_1|))\frac{1}{n^4h_{\min}}|\mathcal{H}_n|. \end{aligned}$$

Then, there exists a deterministic constant $\mathfrak{c}_9>0$, not depending on n and $h_{\min}$, such that

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\frac{|U_{4,n}(h,h_{\min})|}{n^2}\right)\leqslant\mathfrak{c}_9\frac{\log(n)}{n^3h_{\min}}.$$

Therefore,

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\left\{\frac{|U_n(h,h_{\min})|}{n^2}-\frac{\theta\|K\|_2^2}{nh}\mathbb{E}(Y_1^2)\right\}\right)\leqslant\frac{\mathfrak{c}_U}{\theta}\cdot\frac{\log(n)^5}{n}.$$

6.4.3. *Proof of Lemma 6.3.* Consider $\mathfrak{m}(n):=4\log(n)/\alpha$. For any $\eta,\eta'\in\mathcal{H}_n$,

$$V_n(\eta,\eta')=V_{1,n}(\eta,\eta')+V_{2,n}(\eta,\eta')$$

where

$$V_{i,n}(\eta,\eta'):=\frac{1}{n}\sum_{k=1}^{n}(g_{\eta,\eta'}^i(X_k,Y_k)-\mathbb{E}(g_{\eta,\eta'}^i(X_k,Y_k)))\ ;\ i=1,2$$

with, for every $x, y \in \mathbb{R}$,

$$g^1_{\eta,\eta'}(x,y) := \langle yK_\eta(x-\cdot), (bf)_{\eta'} - bf\rangle_2 \mathbf{1}_{|y|\leqslant \mathfrak{m}(n)}$$

and

$$g^2_{\eta,\eta'}(x,y) := \langle yK_\eta(x-\cdot), (bf)_{\eta'} - bf\rangle_2 \mathbf{1}_{|y|> \mathfrak{m}(n)}.$$

In order to apply Bernstein's inequality to $g^1_{\eta,\eta'}(X_k,Y_k)$, $k=1,\dots,n$, let us find suitable controls of

$$\mathfrak{c}_{\eta,\eta'} := \frac{\|g^1_{\eta,\eta'}\|_\infty}{3} \text{ and } \mathfrak{v}_{\eta,\eta'} := \mathbb{E}(g^1_{\eta,\eta'}(X_1,Y_1)^2).$$

On the one hand, since $\|K\|_1 \geqslant 1$ and bf is bounded,

$$\begin{aligned}
\mathfrak{c}_{\eta,\eta'} &= \frac{1}{3}\sup_{x,y\in\mathbb{R}} |\langle yK_\eta(x-\cdot), (bf)_{\eta'} - bf\rangle_2 \mathbf{1}_{|y|\leqslant\mathfrak{m}(n)}|\\
&\leqslant \frac{\mathfrak{m}(n)}{3}\|(bf)_{\eta'} - bf\|_\infty \sup_{x\in\mathbb{R}}\|K_\eta(x-\cdot)\|_1\\
&\leqslant \frac{\mathfrak{m}(n)}{3}\|K\|_1(\|K\|_1+1)\|bf\|_\infty \leqslant \frac{2}{3}\mathfrak{m}(n)\|K\|_1^2\|bf\|_\infty.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathfrak{v}_{\eta,\eta'} &= \mathbb{E}(\langle Y_1K_\eta(X_1-\cdot), (bf)_{\eta'} - bf\rangle_2^2 \mathbf{1}_{|Y_1|\leqslant\mathfrak{m}(n)})\\
&= \mathbb{E}\left(Y_1^2\mathbf{1}_{|Y_1|\leqslant\mathfrak{m}(n)}\left|\int_{-\infty}^{\infty} K_\eta(X_1-u)((bf)_{\eta'}(u) - (bf)(u))du\right|^2\right)\\
&\leqslant \mathfrak{m}(n)^2\|f\|_\infty\|K\|_1^2\|(bf)_{\eta'} - bf\|_2^2.
\end{aligned}$$

So, by Bernstein's inequality, there exists a universal constant $\mathfrak{c}_1 > 0$ such that with probability larger than $1-2e^{-\lambda}$,

$$\begin{aligned}
|V_{1,n}(\eta,\eta')| &\leqslant \sqrt{\frac{2\lambda}{n}\mathfrak{v}_{\eta,\eta'}} + \frac{\lambda}{n}\mathfrak{c}_{\eta,\eta'}\\
&\leqslant \theta\|(bf)_{\eta'} - bf\|_2^2 + \mathfrak{c}_1\frac{\mathfrak{m}(n)^2}{\theta n}\|K\|_1^2(\|f\|_\infty + \|bf\|_\infty)\lambda.
\end{aligned}$$

Then, with probability larger than $1-2|\mathcal{H}_n|^2e^{-\lambda}$,

$$S_n \leqslant \mathfrak{c}_1\frac{\mathfrak{m}(n)^2}{\theta n}\|K\|_1^2(\|f\|_\infty + \|bf\|_\infty)\lambda$$

where

$$S_n := \sup_{\eta,\eta'\in\mathcal{H}_n}\{|V_{1,n}(\eta,\eta')| - \theta\|(bf)_{\eta'} - bf\|_2^2\}.$$

For every $s\in\mathbb{R}_+$, consider

$$\lambda(s) := \frac{s}{\mathfrak{m}(n,\theta)}$$

where

$$\mathfrak{m}(n,\theta) := \mathfrak{c}_1\frac{\mathfrak{m}(n)^2}{\theta n}\|K\|_1^2(\|f\|_\infty + \|bf\|_\infty).$$

Then, for any $A > 0$,

$$\begin{aligned}
\mathbb{E}(S_n) &\leqslant 2A + \int_A^\infty \mathbb{P}(S_n \geqslant s)ds\\
&\leqslant 2A + 2\mathfrak{c}_2|\mathcal{H}_n|^2\mathfrak{m}(n,\theta)\exp\left(-\frac{A}{2\mathfrak{m}(n,\theta)}\right)
\end{aligned}$$

where

$$\mathfrak{c}_2 := \int_0^\infty e^{-s/2}ds.$$

Since there exists a deterministic constant $\mathfrak{c}_3 > 0$, not depending on n and $h_{\min}$ such that

$$\mathfrak{m}(n,\theta) \leqslant \mathfrak{c}_3 \frac{\log(n)^2}{n},$$

by taking $A := 4\mathfrak{c}_3 \log(n)^3/n$,

$$\mathbb{E}(S_n) \leqslant 8\mathfrak{c}_3 \frac{\log(n)^3}{n} + 2\mathfrak{c}_2 \mathfrak{m}(n,\theta)\frac{|\mathcal{H}_n|}{n^2}.$$

Therefore, since $|\mathcal{H}_n| \leqslant n$, there exists a deterministic constant $\mathfrak{c}_4 > 0$, not depending on n and $h_{\min}$, such that

$$\mathbb{E}\left(\sup_{\eta,\eta'\in\mathcal{H}_n}\{|V_{1,n}(\eta,\eta')| - \theta\|(bf)_{\eta'} - bf\|_2^2\}\right) \leqslant \frac{\mathfrak{c}_4}{\theta}\cdot\frac{\log(n)^3}{n}.$$

Now, let us find a suitable control of

$$\mathfrak{v}_{2,n} := \mathbb{E}\left(\sup_{\eta,\eta'\in\mathcal{H}_n}|V_{2,n}(\eta,\eta')|\right).$$

By Markov's inequality,

$$\begin{aligned}
\mathfrak{v}_{2,n} &\leqslant 2\mathbb{E}\left(\sup_{\eta\in\mathcal{H}_n}|Z_{2,1}(\eta,\eta')|\right)\\
&\leqslant 2\mathbb{E}(Y_1^2\mathbf{1}_{|Y_1|>\mathfrak{m}(n)})^{1/2}\\
&\quad\times\mathbb{E}\left(\sup_{\eta,\eta'\in\mathcal{H}_n}\left|\int_{-\infty}^{\infty}K_\eta(X_1-u)((bf)_{\eta'}(u)-(bf)(u))du\right|^2\right)^{1/2}\\
&\leqslant 2\mathbb{E}(Y_1^4)^{1/4}\mathbb{P}(\exp(\alpha|Y_1|)>n^4)^{1/4}\|K\|_1\sup_{\eta'\in\mathcal{H}_n}\|(bf)_{\eta'}-bf\|_\infty\\
&\leqslant 2\mathbb{E}(Y_1^4)^{1/4}\mathbb{E}(\exp(\alpha|Y_1|))^{1/4}\|K\|_1^2\|bf\|_\infty\frac{1}{n}.
\end{aligned}$$

Therefore,

$$\mathbb{E}\left(\sup_{\eta,\eta'\in\mathcal{H}_n}\{|V_n(\eta,\eta')| - \theta\|(bf)_{\eta'} - bf\|_2^2\}\right) \leqslant \mathfrak{c}_V\frac{\log(n)^3}{\theta n}.$$

6.4.4. *Proof of Lemma 6.4.* First of all,

$$\|(bf)_h - bf\|_2^2 = \|\widehat{bf}_{n,h} - bf\|_2^2 - \|\widehat{bf}_{n,h} - (bf)_h\|_2^2 - 2V_n(h,h).$$

Then, for any $\theta\in(0,1/2)$,

$$\begin{aligned}
(1-2\theta)\left(\|(bf)_h - bf\|_2^2 + \frac{\mathfrak{c}_{K,Y}}{nh}\right) - \|\widehat{bf}_{n,h} - bf\|_2^2 &\leqslant 2(|V_n(h,h)| - \theta\|(bf)_h - bf\|_2^2)\\
&\quad + \Lambda_n(h) - 2\theta\frac{\mathfrak{c}_{K,Y}}{nh}
\end{aligned}\tag{12}$$

where

$$\begin{aligned}
\Lambda_n(h) &:= \left|\|\widehat{bf}_{n,h} - (bf)_h\|_2^2 - \frac{\mathfrak{c}_{K,Y}}{nh}\right|\\
&= \left|\frac{U_n(h,h)}{n^2} + \frac{W_n(h)}{n} - \frac{1}{n}\|(bf)_h\|_2^2\right|
\end{aligned}$$

with

$$W_n(h) := \frac{1}{n}\sum_{k=1}^{n}(Z_k(h) - \mathbb{E}(Z_k(h)))$$

and

$$Z_k(h) := \|Y_kK_h(X_k - \cdot) - (bf)_h\|_2^2\ ;\ \forall k\in\{1,\dots,n\},$$

because

$$\begin{aligned}\mathbb{E}(Z_1(h)) &= \sigma^2\int_{-\infty}^{\infty}\mathbb{E}(K_h(X_1-u)^2)du + \int_{-\infty}^{\infty}\mathbb{E}(b(X_1)^2K_h(X_1-u)^2)du\\ &\quad -2\int_{-\infty}^{\infty}\mathbb{E}(b(X_1)K_h(X_1-u))(bf)_h(u)du + \int_{-\infty}^{\infty}(bf)_h(u)^2du\\ &= \frac{\|K\|_2^2}{h}(\sigma^2+\mathbb{E}(b(X_1)^2)) - \|(bf)_h\|_2^2 = \frac{\mathfrak{c}_{K,Y}}{h} - \|(bf)_h\|_2^2.\end{aligned}$$

Consider $\mathfrak{m}(n) := 2\log(n)/\alpha$ and note that $W_n(h) = W_{1,n}(h) + W_{2,n}(h)$, where

$$W_{i,n}(h) := \frac{1}{n}\sum_{k=1}^{n}(g_h^i(X_k,Y_k) - \mathbb{E}(g_h^i(X_k,Y_k)))\ ;\ i = 1,2$$

with, for every $x, y\in\mathbb{R}$,

$$g_h^1(x,y) := \|yK_h(x-\cdot) - (bf)_h\|_2^2\mathbf{1}_{|y|\leqslant\mathfrak{m}(n)}$$

and

$$g_h^2(x,y) := \|yK_h(x-\cdot) - (bf)_h\|_2^2\mathbf{1}_{|y|>\mathfrak{m}(n)}.$$

Note also that

$$\|(bf)_h\|_2 \leqslant \|K_h\|_2\int_{-\infty}^{\infty}|b(x)|f(x)dx \leqslant \frac{\|K\|_2}{h^{1/2}}\mathbb{E}(|b(X_1)|) \leqslant \left(\frac{\mathfrak{c}_{K,Y}}{h}\right)^{1/2}$$

and

$$\|(bf)_h\|_2 \leqslant \|K_h\|_1\left(\int_{-\infty}^{\infty}b(x)^2f(x)^2dx\right)^{1/2} \leqslant \|K\|_1\|f\|_\infty^{1/2}\mathbb{E}(b(X_1)^2)^{1/2}.$$

In order to apply Bernstein's inequality to $g_h^1(X_k,Y_k)$, $k = 1,\dots,n$, let us find suitable controls of

$$\mathfrak{c}_h := \frac{\|g_h^1\|_\infty}{3} \text{ and } \mathfrak{v}_h := \mathbb{E}(g_h^1(X_1,Y_1)^2).$$

On the one hand,

$$\begin{aligned}\mathfrak{c}_h &= \frac{1}{3}\sup_{x,y\in\mathbb{R}}\|yK_h(x-\cdot) - (bf)_h\|_2^2\mathbf{1}_{|y|\leqslant\mathfrak{m}(n)}\\ &\leqslant \frac{2}{3}\left(\mathfrak{m}(n)^2\frac{\|K\|_2^2}{h} + \frac{\mathfrak{c}_{K,Y}}{h}\right).\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathfrak{v}_h &\leqslant 2\mathbb{E}(Z_1(h)(\|Y_1K_h(X_1-\cdot)\|_2^2\mathbf{1}_{|Y_1|\leqslant\mathfrak{m}(n)} + \|(bf)_h\|_2^2))\\ &\leqslant \frac{2}{h}\mathbb{E}(Z_1(h))(\|K\|_2^2\mathfrak{m}(n)^2 + \mathfrak{c}_{K,Y}) \leqslant 2(\|K\|_2^2 + \mathfrak{c}_{K,Y})\frac{\mathfrak{c}_{K,Y}}{hh_{\min}}\mathfrak{m}(n)^2.\end{aligned}$$

So, by Bernstein's inequality, there exists a universal constant $\mathfrak{c}_1 > 0$ such that with probability larger than $1 - 2e^{-\lambda}$,

$$\begin{aligned}|W_{1,n}(h)| &\leqslant \sqrt{\frac{2\lambda}{n}\mathfrak{v}_h} + \frac{\lambda}{n}\mathfrak{c}_h\\ &\leqslant \theta\frac{\mathfrak{c}_{K,Y}}{h} + \mathfrak{c}_1\frac{\mathfrak{m}(n)^2}{\theta nh_{\min}}(\|K\|_2^2 + \mathfrak{c}_{K,Y})\lambda.\end{aligned}$$

Then, with probability larger than $1 - 2|\mathcal{H}_n|e^{-\lambda}$,

$$S_n(h_{\min}) \leqslant \mathfrak{c}_1\frac{\mathfrak{m}(n)^2}{\theta n^2h_{\min}}(\|K\|_2^2 + \mathfrak{c}_{K,Y})\lambda$$

where

$$S_n(h_{\min}) := \sup_{h\in\mathcal{H}_n}\left\{\frac{|W_{1,n}(h)|}{n} - \theta\frac{\mathfrak{c}_{K,Y}}{nh}\right\}.$$

For every $s\in\mathbb{R}_+$, consider

$$\lambda(s) := \frac{s}{\mathfrak{m}(n,h_{\min},\theta)}$$

where

$$\mathfrak{m}(n,h_{\min},\theta) := \mathfrak{c}_1\frac{\mathfrak{m}(n)^2}{\theta n^2h_{\min}}(\|K\|_2^2+\mathfrak{c}_{K,Y}).$$

Then, for any $A>0$,

$$\begin{aligned}\mathbb{E}(S_n(h_{\min})) &\leqslant 2A+\int_A^{\infty}\mathbb{P}(S_n(h_{\min})\geqslant s)ds\\ &\leqslant 2A+2\mathfrak{c}_2|\mathcal{H}_n|\mathfrak{m}(n,h_{\min},\theta)\exp\left(-\frac{A}{2\mathfrak{m}(n,h_{\min},\theta)}\right)\end{aligned}$$

where

$$\mathfrak{c}_2 := \int_0^{\infty}e^{-s/2}ds.$$

Since there exists a deterministic constant $\mathfrak{c}_3>0$, not depending on n and $h_{\min}$ such that

$$\mathfrak{m}(n,\theta)\leqslant\mathfrak{c}_3\frac{\log(n)^2}{n},$$

by taking $A:=2\mathfrak{c}_3\log(n)^3/n$,

$$\mathbb{E}(S_n(h_{\min}))\leqslant 4\mathfrak{c}_3\frac{\log(n)^3}{n}+2\mathfrak{c}_2\mathfrak{m}(n,h_{\min},\theta)\frac{|\mathcal{H}_n|}{n}.$$

Therefore, since $|\mathcal{H}_n|\leqslant n$, there exists a deterministic constant $\mathfrak{c}_4>0$, not depending on n and $h_{\min}$, such that

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\left\{\frac{|W_{1,n}(h)|}{n}-\theta\frac{\mathfrak{c}_{K,Y}}{nh}\right\}\right)\leqslant\frac{\mathfrak{c}_4}{\theta}\cdot\frac{\log(n)^3}{n}.$$

Now, by Markov's inequality,

$$\begin{aligned}\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\frac{|W_{2,n}(h)|}{n}\right) &\leqslant \frac{2}{n}\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}|Z_1(h)|\mathbf{1}_{|Y_1|>\mathfrak{m}(n)}\right)\\ &\leqslant \frac{4}{n}\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}(\|Y_1K_h(X_1-\cdot)\|_2^2+\|(bf)_h\|_2^2)\mathbf{1}_{|Y_1|>\mathfrak{m}(n)}\right)\\ &\leqslant \frac{4}{nh_{\min}}(\|K\|_2^2\mathbb{E}(Y_1^4)^{1/2}+\mathfrak{c}_{K,Y})\mathbb{P}(|Y_1|>\mathfrak{m}(n))^{1/2}\\ &\leqslant 4(\|K\|_2^2\mathbb{E}(Y_1^4)^{1/2}+\mathfrak{c}_{K,Y})\mathbb{E}(\exp(\alpha|Y_1|))^{1/2}\frac{1}{\theta n^2h_{\min}}.\end{aligned}$$

Then, there exists a deterministic constant $\mathfrak{c}_5>0$, not depending on n and $h_{\min}$, such that

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\left\{\frac{|W_n(h)|}{n}-\theta\frac{\mathfrak{c}_{K,Y}}{nh}\right\}\right)\leqslant\frac{\mathfrak{c}_5}{\theta}\cdot\frac{\log(n)^3}{n}.$$

Therefore, by Lemma 6.2, there exists a deterministic constant $\mathfrak{c}_6>0$, not depending on n and $h_{\min}$, such that

$$\begin{aligned}\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\left\{\Lambda_n(h)-2\theta\frac{\mathfrak{c}_{K,Y}}{nh}\right\}\right) &\leqslant \frac{\mathfrak{c}_U}{\theta}\cdot\frac{\log(n)^5}{n}+\frac{\mathfrak{c}_5}{\theta}\cdot\frac{\log(n)^3}{n}\\ &\quad+\frac{1}{n}\|K\|_1^2\|f\|_{\infty}\mathbb{E}(b(X_1)^2)\\ &\leqslant \frac{\mathfrak{c}_6}{\theta}\cdot\frac{\log(n)^5}{n}.\end{aligned}$$

Moreover, by Lemma 6.3,

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\{|V_n(h,h)|-\theta\|(bf)_h-bf\|_2^2\}\right)\leqslant\frac{\mathfrak{c}_V}{\theta}\cdot\frac{\log(n)^3}{n}.$$

In conclusion, by Inequality (12),

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\left\{\|(bf)_h - bf\|_2^2 + \frac{\mathfrak{c}_{K,Y}}{nh} - \frac{1}{1-2\theta}\|\widehat{bf}_{n,h} - bf\|_2^2\right\}\right) \leqslant \frac{\mathfrak{c}_L}{\theta(1-2\theta)}\cdot\frac{\log(n)^5}{n}.$$

References

[1] X. Chang, X., S.-B. Lin and Y. Wang. Divide and conquer local average regression. *Electron. J. Stat.* **11**, no. 1, 1326-1350, 2017.

[2] F. Comte and N. Marie. Bandwidth selection for the Wolverton-Wagner estimator. To appear in *Journal of Statistical Planning and Inference*, 2019.

[3] F. Comte. *Estimation non-paramétrique.* Spartacus IDH, 2nd edition, 2017.

[4] F. Comte and T. Rebafka. Nonparametric weighted estimators for biased data. *J. Statist. Plann. Inference* **174**, 104-128, 2016.

[5] A. Goldenshluger and O. Lepski. Bandwidth Selection in Kernel Density Estimation: Oracle Inequalities and Adaptive Minimax Optimality. *The Annals of Statistics* **39**, 1608-1632, 2011.

[6] L. Györfi, M. Kohler, A. Krzyzak and H. Walk. *A distribution-free theory of nonparametric regression.* Springer Series in Statistics. Springer-Verlag, New York, 2002.

[7] Härdle, W.; Tsybakov, A.; Yang, L. Nonparametric vector autoregression. *J. Statist. Plann. Inference* **68**, no. 2, 221-245, 1998.

[8] C. Houdré and P. Reynaud-Bouret. Exponential Inequalities, with Constants, for U-Statistics of Order Two. *Stochastic Inequalities and Applications, Progr. Probab.*, **56** Birkhäuser, Basel, 55-69, 2003.

[9] T. Klein and E. Rio. Concentration Around the Mean for Maxima of Empirical Processes. *The Annals of Probability*, **33**, 1060-1077, 2005.

[10] C. Lacour, P. Massart and V. Rivoirard. Estimator Selection: a New Method with Applications to Kernel Density Estimation. *Sankhya* **79**, 298-335, 2017.

[11] E.A. Nadaraya. On a regression estimate. (Russian) *Verojatnost. i Primenen.* **9**, 157-159, 1964.

[12] E. Parzen. On the Estimation of a Probability Density Function and the Mode. *The Annals of Mathematical Statistics* **33**, 1065-1076, 1962.

[13] M. Rosenblatt. Remarks on some Nonparametric Estimates of a Density Function. *Ann. Math. Statist.* **27**, 832-837, 1956.

[14] C. J. Stone. Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics* **10**, no. 4, 1040-1053, 1982.

[15] A. Tsybakov. *Introduction to Nonparametric Estimation.* Springer, 2009.

[16] S. Varet, C. Lacour, P. Massart and V. Rivoirard. Numerical performance of Penalized Comparison to Overfitting for multivariate kernel density estimation. *Preprint* Hal, https://hal.archives-ouvertes.fr/hal-02002275

[17] M. P. Wand and M. C. Jones. *Kernel smoothing.* Monographs on Statistics and Applied Probability, 60. Chapman and Hall, Ltd., London, 1995.

[18] G. S. Watson. Smooth regression analysis. *Sankhya* Ser. A **26**, 359-372, 1964.

*Laboratoire MAP5, Université Paris Descartes, Paris, France
E-mail address: fabienne.comte@parisdescartes.fr

**Laboratoire Modal'X, Université Paris Nanterre, Nanterre, France
E-mail address: nmarie@parisnanterre.fr

**ESME Sudria, Paris, France
E-mail address: nicolas.marie@esme.fr