Kernel Selection in Nonparametric Regression

Hélène Halconruy, Nicolas Marie

KERNEL SELECTION IN NONPARAMETRIC REGRESSION

HÉLÈNE HALCONRUY* AND NICOLAS MARIE†

Abstract. In the regression model $Y = b(X) + \sigma(X)\varepsilon$, where X has a density f, this paper deals with an oracle inequality for an estimator of bf, involving a kernel in the sense of Lerasle et al. (2016), selected via the PCO method. In addition to the bandwidth selection for kernel-based estimators already studied in Lacour, Massart and Rivoirard (2017) and Comte and Marie (2020), the dimension selection for anisotropic projection estimators of f and bf is covered.

Contents

MSC2010: 62G05 ; 62G08.

1. Introduction

Consider $n \in \mathbb{N}^*$ independent $\mathbb{R}^d \times \mathbb{R}$-valued ($d \in \mathbb{N}^*$) random variables $(X_1, Y_1), \ldots, (X_n, Y_n)$, having the same probability distribution assumed to be absolutely continuous with respect to Lebesgue's measure, and

$$\widehat{s}_{K,\ell}(n; x) := \frac{1}{n} \sum_{i=1}^{n} K(X_i, x)\ell(Y_i) \ ; \ x \in \mathbb{R}^d,$$

where $\ell : \mathbb{R} \to \mathbb{R}$ is a Borel function and K is a symmetric continuous map from $\mathbb{R}^d \times \mathbb{R}^d$ into $\mathbb{R}$. This is an estimator of the function $s : \mathbb{R}^d \to \mathbb{R}$ defined by

$$s(x) := \mathbb{E}(\ell(Y_1)|X_1 = x)f(x) \ ; \forall x \in \mathbb{R}^d,$$

where f is a density of X_1. For $\ell = 1$, $\widehat{s}_{K,\ell}(n; .)$ coincides with the estimator of f studied in Lerasle et al. [13], covering Parzen-Rosenblatt's and projection estimators already deeply studied in the literature (see Parzen [16], Rosenblatt [17], Tsybakov [18], etc.), but for $\ell \neq 1$, it covers estimators involved in nonparametric regression. Assume that for every $i \in \{1, \ldots, n\}$,

$$Y_i = b(X_i) + \sigma(X_i)\varepsilon_i \tag{1}$$

Key words and phrases. Nonparametric estimators ; Projection estimators ; Model selection ; Regression model.

where ε_i is a centered random variable of variance 1, independent of X_i, and $b, \sigma : \mathbb{R}^d \to \mathbb{R}$ are Borel functions.

- If $\ell = \mathrm{Id}_{\mathbb{R}}$, k is a symmetric kernel and

$$K(x', x) = \prod_{q=1}^{d} \frac{1}{h_q} k\left(\frac{x'_q - x_q}{h_q}\right) \text{ with } h_1, \ldots, h_d > 0 \tag{2}$$

  for every $x, x' \in \mathbb{R}^d$, then $\widehat{s}_{K,\ell}(n;.)$ is the numerator of the well-known Nadaraya-Watson estimator of the regression function b (see Nadaraya [15] and Watson [20]). Precisely, $\widehat{s}_{K,\ell}(n;.)$ is an estimator of $s = bf$ because ε_1 is independent to X_1 and $\mathbb{E}(\varepsilon_1) = 0$. If $\ell \neq \mathrm{Id}_{\mathbb{R}}$, then $\widehat{s}_{K,\ell}(n;.)$ is the numerator of the estimator studied in Einmahl and Mason [7, 8].
- If $\ell = \mathrm{Id}_{\mathbb{R}}$, $\mathcal{B}_{m_q} = \{\varphi_1^{m_q}, \ldots, \varphi_{m_q}^{m_q}\}$ ($m_q \in \mathbb{N}^*$ and $q \in \{1, \ldots, d\}$) is an orthonormal family of $\mathbb{L}^2(\mathbb{R})$ and

$$K(x', x) = \prod_{q=1}^{d} \sum_{j=1}^{m_q} \varphi_j^{m_q}(x_q) \varphi_j^{m_q}(x'_q) \tag{3}$$

  for every $x, x' \in \mathbb{R}^d$, then $\widehat{s}_{K,\ell}(n;.)$ is the projection estimator on $\mathcal{S} = \mathrm{span}(\mathcal{B}_{m_1} \otimes \cdots \otimes \mathcal{B}_{m_d})$ of $s = bf$.

Now, assume that $b = 0$ in Model (1): for every $i \in \{1, \ldots, n\}$,

$$Y_i = \sigma(X_i)\varepsilon_i. \tag{4}$$

If $\ell(x) = x^2$ for every $x \in \mathbb{R}$, then $\widehat{s}_{K,\ell}(n;.)$ is an estimator of $s = \sigma^2 f$.

These ten last years, several data-driven procedures have been proposed in order to select the bandwidth of Parzen-Rosenblatt's estimator ($\ell = 1$ and K defined by (2)). First, Goldenshluger-Lepski's method, introduced in [10], which reaches the adequate bias-variance compromise, but is not completely satisfactory on the numerical side (see Comte and Rebafka [5]). More recently, in [12], Lacour, Massart and Rivoirard proposed the PCO (Penalized Comparison to Overfitting) method and proved an oracle inequality for the associated adaptive Parzen-Rosenblatt's estimator by using a concentration inequality for the U-statistics due to Houdré and Reynaud-Bouret [11]. Together with Varet, they established the numerical efficiency of the PCO method in Varet et al. [19]. Still in the density estimation framework, the PCO method has been extended to bandwidths selection for the recursive Wolverton-Wagner estimator in Comte and Marie [3].
Comte and Marie [4] deal with an oracle inequality and numerical experiments for an adaptive Nadaraya-Watson's estimator with a numerator and a denominator having distinct bandwidths, both selected via the PCO method. Since the output variable in a regression model has no reason to be bounded, there were significant additional difficulties, bypassed in [4], to establish an oracle inequality for the numerator's adaptive estimator. Via similar arguments, the present article deals with an oracle inequality for $\widehat{s}_{\widehat{K},\ell}(n;.)$, where $\widehat{K}$ is selected via the PCO method in the spirit of Lerasle et al. [13]. As in Comte and Marie [4], one can deduce an oracle inequality for the adaptive quotient estimator $\widehat{s}_{\widehat{K},\ell}(n;.)/\widehat{s}_{\widehat{L},1}(n;.)$ of $\mathbb{E}(\ell(Y_1)|X_1 = \cdot)$, where $\widehat{K}$ and $\widehat{L}$ are both selected via the PCO method.
In addition to the bandwidth selection for kernel-based estimators already studied in [12, 4], the present paper covers the dimension selection for projection estimators of f, bf when $Y_1, \ldots, Y_n$ are defined by Model (1) with $\ell = \mathrm{Id}_{\mathbb{R}}$, and $\sigma^2 f$ when $Y_1, \ldots, Y_n$ are defined by Model (4) with $\ell(x) = x^2$ for every $x \in \mathbb{R}$. For projection estimators, when $d = 1$, the usual model selection method (see Comte [2], Chapter 2, Section 5) seems hard to beat. However, when $d > 1$ and K is defined by (3), $m_1, \ldots, m_d$ are selected via a Goldenshluger-Lepski type method (see Chagny [1]), which has the same numerical weakness than the Goldenshluger-Lepski method for bandwidth selection when K is defined by (2). So, for the dimension selection for anisotropic projection estimators, the PCO method is interesting.

In Section 2, some examples of kernels sets are provided and a risk bound on $\widehat{s}_{K,\ell}(n;.)$ is established. Section 3 deals with an oracle inequality for $\widehat{s}_{\widehat{K},\ell}(n;.)$, where $\widehat{K}$ is selected via the PCO method.

2. Risk bound

Throughout the paper, $s \in \mathbb{L}^2(\mathbb{R}^d)$. Let $\mathcal{K}_n$ be a set of symmetric continuous maps from $\mathbb{R}^d \times \mathbb{R}^d$ into $\mathbb{R}$, of cardinality less or equal than n, fulfilling the following assumption.

Assumption 2.1. *There exists a deterministic constant $\mathfrak{m}_{\mathcal{K},\ell} > 0$, not depending on n, such that*

(1) *For every $K \in \mathcal{K}_n$,*
$$\sup_{x' \in \mathbb{R}^d} \|K(x', .)\|_2^2 \leqslant \mathfrak{m}_{\mathcal{K},\ell} n.$$

(2) *For every $K \in \mathcal{K}_n$,*
$$\|s_{K,\ell}\|_2^2 \leqslant \mathfrak{m}_{\mathcal{K},\ell}$$
with
$$s_{K,\ell} := \mathbb{E}(\widehat{s}_{K,\ell}(n; .)) = \mathbb{E}(K(X_1, .)\ell(Y_1)).$$

(3) *For every $K, K' \in \mathcal{K}_n$,*
$$\mathbb{E}(\langle K(X_1, .), K'(X_2, .)\ell(Y_2)\rangle_2^2) \leqslant \mathfrak{m}_{\mathcal{K},\ell} \overline{s}_{K',\ell}$$
with
$$\overline{s}_{K',\ell} := \mathbb{E}(\|K'(X_1, .)\ell(Y_1)\|_2^2).$$

(4) *For every $K \in \mathcal{K}_n$ and $\psi \in \mathbb{L}^2(\mathbb{R}^d)$,*
$$\mathbb{E}(\langle K(X_1, .), \psi\rangle_2^2) \leqslant \mathfrak{m}_{\mathcal{K},\ell} \|\psi\|_2^2.$$

The elements of $\mathcal{K}_n$ are called kernels. Let us provide two natural examples of kernels sets.

Proposition 2.2. *Consider*
$$\mathcal{K}_k(h_{\min}) := \left\{ (x', x) \mapsto \prod_{q=1}^{d} \frac{1}{h_q} k\left(\frac{x'_q - x_q}{h_q}\right) \ ; \ h_1, \ldots, h_d \in \mathcal{H}(h_{\min}) \right\},$$
where k is a symmetric kernel (in the usual sense), $h_{\min} \in [n^{-1/d}, 1]$ and $\mathcal{H}(h_{\min})$ is a finite subset of $[h_{\min}, 1]$. The kernels set $\mathcal{K}_k(h_{\min})$ fulfills Assumption 2.1 and, for any $K \in \mathcal{K}_k(h_{\min})$ (i.e. defined by (2) with $h_1, \ldots, h_d \in \mathcal{H}(h_{\min})$),
$$\overline{s}_{K,\ell} = \|k\|_2^{2d} \mathbb{E}(\ell(Y_1)^2) \prod_{q=1}^{d} \frac{1}{h_q}.$$

Proposition 2.3. *Consider*
$$\mathcal{K}_{\mathcal{B}_1,\ldots,\mathcal{B}_n}(m_{\max}) := \left\{ (x', x) \mapsto \prod_{q=1}^{d} \sum_{j=1}^{m_q} \varphi_j^{m_q}(x_q) \varphi_j^{m_q}(x'_q) \ ; \ m_1, \ldots, m_d \in \{1, \ldots, m_{\max}\} \right\},$$
where $m_{\max}^d \in \{1, \ldots, n\}$ and, for every $m \in \{1, \ldots, n\}$, $\mathcal{B}_m = \{\varphi_1^m, \ldots, \varphi_m^m\}$ is an orthonormal family of $\mathbb{L}^2(\mathbb{R})$ such that
$$\sup_{x' \in \mathbb{R}} \sum_{j=1}^{m} \varphi_j^m(x')^2 \leqslant \mathfrak{m}_{\mathcal{B}} m$$
with $\mathfrak{m}_{\mathcal{B}} > 0$ not depending on m and n, and such that one of the two following conditions is satisfied:

(5) $$\mathcal{B}_m \subset \mathcal{B}_{m+1} \ ; \forall m \in \{1, \ldots, n-1\}$$

or

(6) $\overline{\mathfrak{m}}_{\mathcal{B}} := \sup\{|\mathbb{E}(K(X_1, x))| \ ; \ K \in \mathcal{K}_{\mathcal{B}_1,\ldots,\mathcal{B}_n}(m_{\max}) \text{ and } x \in \mathbb{R}^d\}$ *is finite and doesn't depend on n.*

The kernels set $\mathcal{K}_{\mathcal{B}_1,\ldots,\mathcal{B}_n}(m_{\max})$ fulfills Assumption 2.1 and, for any $K \in \mathcal{K}_{\mathcal{B}_1,\ldots,\mathcal{B}_n}(m_{\max})$ (i.e. defined by (3) with $m_1, \ldots, m_n \in \{1, \ldots, m_{\max}\}$),
$$\overline{s}_{K,\ell} \leqslant \mathfrak{m}_{\mathcal{B}}^d \mathbb{E}(\ell(Y_1)^2) \prod_{q=1}^{d} m_q.$$

Remark 2.4. *For the sake of simplicity, the present paper focuses on $\mathcal{K}_{\mathcal{B}_1,\dots,\mathcal{B}_n}(m_{\max})$, but Proposition 2.3 is still true for the weighted projection kernels set*

$$\mathcal{K}_{\mathcal{B}_1,\dots,\mathcal{B}_n}(w_1,\dots,w_n;m_{\max}) := \left\{ (x',x) \mapsto \prod_{q=1}^{d}\sum_{j=1}^{m_q} w_j \varphi_j^{m_q}(x_q)\varphi_j^{m_q}(x'_q) \ ; \ m_1,\dots,m_d \in \{1,\dots,m_{\max}\} \right\},$$

where $w_1,\dots,w_n \in [0,1]$.

Remark 2.5. *Note that Condition (5) is close, but more restrictive than Condition (19) of Lerasle et al.* [13]*, Proposition 3.2, which is that the spaces* $\mathrm{span}(\mathcal{B}_m)$*, $m \in \mathbb{N}$ are nested. See Massart* [14]*, Subsection 7.5.2 for examples of nested spaces. Our Condition (5) is fulfilled by the trigonometric basis, Hermite's basis or Laguerre's basis.*
Note also that in the same proposition of Lerasle et al. [13]*, Condition (20) coincides with our Condition (6). The regular histograms basis satisfies Condition (6). Indeed, by taking $\varphi_j^m = \psi_j^m := \sqrt{m}\mathbf{1}_{[(j-1)/m,j/m[}$ for every $m \in \{1,\dots,n\}$ and $j \in \{1,\dots,m\}$,*

$$\begin{aligned}\left|\mathbb{E}\left[\prod_{q=1}^{d}\sum_{j=1}^{m_q}\psi_j^{m_q}(X_{1,q})\psi_j^{m_q}(x_q)\right]\right| &= \sum_{j_1=1}^{m_1}\cdots\sum_{j_d=1}^{m_d}\left(\prod_{q=1}^{d} m_q \mathbf{1}_{[(j_q-1)/m_q,j_q/m_q[}(x_q)\right) \\ &\quad\times \int_{(j_1-1)/m_1}^{j_1/m_1}\cdots\int_{(j_d-1)/m_d}^{j_d/m_d} f(x'_1,\dots,x'_d)dx'_1\cdots dx'_d \\ &\leqslant \|f\|_\infty \prod_{q=1}^{d}\sum_{j=1}^{m_q}\mathbf{1}_{[(j-1)/m_q,j/m_q[}(x_q) \leqslant \|f\|_\infty\end{aligned}$$

for every $m_1,\dots,m_d \in \{1,\dots,n\}$ and $x \in \mathbb{R}^d$.

The following proposition shows that Legendre's basis also fulfills Condition (6).

Proposition 2.6. *For every $m \in \{1,\dots,n\}$ and $j \in \{1,\dots,m\}$, let ξ_j^m be the function defined on $[-1,1]$ by*

$$\xi_j^m(x) := \sqrt{\frac{2j+1}{2}}Q_j(x) \ ; \ \forall x \in [-1,1],$$

where

$$Q_j : x \in [-1,1] \longmapsto \frac{1}{2^j j!}\cdot\frac{d^j}{dx^j}(x^2-1)^j$$

is the j-th Legendre's polynomial. If $f \in C^{2d}([0,1]^d)$ and $\mathcal{B}_m = \{\xi_1^m,\dots,\xi_m^m\}$ for every $m \in \{1,\cdots,n\}$, then $\mathcal{K}_{\mathcal{B}_1,\dots,\mathcal{B}_m}(m_{\max})$ fulfills Condition (6).

The following proposition provides a suitable control of the variance of $\widehat{s}_{K,\ell}(n;.)$.

Proposition 2.7. *Under Assumption 2.1.(1,2,3), if $s \in \mathbb{L}^2(\mathbb{R}^d)$ and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|\ell(Y_1)|)) < \infty$, then there exists a deterministic constant $\mathfrak{c}_{2.7} > 0$, not depending on n, such that for every $\theta \in]0,1[$,*

$$\mathbb{E}\left(\sup_{K\in\mathcal{K}_n}\left\{\left|\|\widehat{s}_{K,\ell}(n;.)-s_{K,\ell}\|_2^2 - \frac{\overline{s}_{K,\ell}}{n}\right| - \frac{\theta}{n}\overline{s}_{K,\ell}\right\}\right) \leqslant \mathfrak{c}_{2.7}\frac{\log(n)^5}{\theta n}.$$

Finally, let us state the main result of this section.

Theorem 2.8. *Under Assumption 2.1, if $s \in \mathbb{L}^2(\mathbb{R}^d)$ and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|\ell(Y_1)|)) < \infty$, then there exist deterministic constants $\mathfrak{c}_{2.8}, \overline{\mathfrak{c}}_{2.8} > 0$, not depending on n, such that for every $\theta \in]0,1[$,*

$$\mathbb{E}\left(\sup_{K\in\mathcal{K}_n}\left\{\|\widehat{s}_{K,\ell}(n;.)-s\|_2^2 - (1+\theta)\left(\|s_{K,\ell}-s\|_2^2 + \frac{\overline{s}_{K,\ell}}{n}\right)\right\}\right) \leqslant \mathfrak{c}_{2.8}\frac{\log(n)^5}{\theta n}$$

and

$$\mathbb{E}\left(\sup_{K\in\mathcal{K}_n}\left\{\|s_{K,\ell}-s\|_2^2 + \frac{\overline{s}_{K,\ell}}{n} - \frac{1}{1-\theta}\|\widehat{s}_{K,\ell}(n;.)-s\|_2^2\right\}\right) \leqslant \overline{\mathfrak{c}}_{2.8}\frac{\log(n)^5}{\theta(1-\theta)n}.$$

Remark 2.9. *Note that the first inequality in Theorem 2.8 gives a risk bound on the estimator* $\widehat{s}_{K,\ell}(n;.)$*:*

$$\mathbb{E}(\|\widehat{s}_{K,\ell}(n;.)-s\|_2^2) \leqslant (1+\theta)\left(\|s_{K,\ell}-s\|_2^2+\frac{\overline{s}_{K,\ell}}{n}\right)+\mathfrak{c}_{2.8}\frac{\log(n)^5}{\theta n}$$

for every $\theta \in]0,1[$*. The second inequality is useful in order to establish a risk bound on the adaptive estimator defined in the next section (see Theorem 3.2).*

Remark 2.10. *In Proposition 2.7 and Theorem 2.8, the exponential moment condition may appear too strong. Nevertheless, this is de facto satisfied when*

$$\ell(Y_1),\dots,\ell(Y_n) \text{ have a compactly supported distribution.} \tag{7}$$

This last condition is satisfied in the density estimation framework because $\ell = 1$*, but even in the non-parametric regression framework, where* ℓ *is not bounded, when* $Y_1,\dots,Y_n$ *have a compactly supported distribution. Moreover, note that under Condition (7), the risk bounds of Theorem 2.8 can be stated in deviation, without additional steps in the proof. Precisely, under Assumption 2.1 and Condition (7), if* $s \in \mathbb{L}^2(\mathbb{R}^d)$*, then there exists a deterministic constant* $\mathfrak{c}_L > 0$*, depending on* $L = \sup_{z\in\mathrm{supp}(\mathbb{P}_{Y_1})}|\ell(z)|$ *but not on* n*, such that for every* $\vartheta \in]0,1[$ *and* $\lambda > 0$*,*

$$\sup_{K\in\mathcal{K}_n}\left|\|s_{K,\ell}-s\|_2^2+\frac{\overline{s}_{K,\ell}}{n}-\frac{1}{1-\vartheta}\|\widehat{s}_{K,\ell}(n;.)-s\|_2^2\right| \leqslant \frac{\mathfrak{c}_L}{\vartheta n}(1+\lambda)^3$$

with probability larger than $1-9.4|\mathcal{K}_n|e^{-\lambda}$*.*
When Condition (7) doesn't hold true, one can replace the exponential moment condition of Proposition 2.7 and Theorem 2.8 by a q*-th order moment condition on* $\ell(Y_1)$ *(*$q \in \mathbb{N}^*$*), but with a damaging effect on the rate of convergence of* $\widehat{s}_{K,\ell}(n;.)$*. For instance, at Remark B.5, it is established that under a* $(12-4\varepsilon)/\beta$*-th moment condition (*$\varepsilon \in]0,1[$ *and* $0<\beta<\varepsilon/2$*), the rate of convergence is of order* $O(1/n^{1-\varepsilon})$ *(instead of* $1/n$*) in Lemma B.2. This holds true for the three technical lemmas of Subsection B.1, and then for Proposition 2.7 and Theorem 2.8.*

3. Kernel selection

This section deals with a risk bound on the adaptive estimator $\widehat{s}_{\widehat{K},\ell}(n;.)$, where

$$\widehat{K} \in \arg\min_{K\in\mathcal{K}_n}\{\|\widehat{s}_{K,\ell}(n;\cdot)-\widehat{s}_{K_0,\ell}(n;\cdot)\|_2^2+\mathrm{pen}_\ell(K)\},$$

K_0 is an overfitting proposal for K in the sense that

$$K_0 \in \arg\max_{K\in\mathcal{K}_n}\left\{\sup_{x\in\mathbb{R}^d}|K(x,x)|\right\},$$

and

$$\mathrm{pen}_\ell(K) := \frac{2}{n^2}\sum_{i=1}^n\langle K(.,X_i),K_0(.,X_i)\rangle_2\ell(Y_i)^2 \ ; \forall K\in\mathcal{K}_n. \tag{8}$$

Example. On the one hand, for any $K \in \mathcal{K}_k(h_{\min})$ (i.e. defined by (2) with $h_1,\dots,h_d \in \mathcal{H}(h_{\min})$),

$$\sup_{x\in\mathbb{R}^d}|K(x,x)| = |k(0)|^d\prod_{q=1}^d\frac{1}{h_q}.$$

Then, for $\mathcal{K}_n = \mathcal{K}_k(h_{\min})$,

$$K_0(x',x) = \frac{1}{h_{\min}^d}\prod_{q=1}^d k\left(\frac{x'_q-x_q}{h_{\min}}\right) \ ; \forall x,x'\in\mathbb{R}^d.$$

On the other hand, for any $K \in \mathcal{K}_{\mathcal{B}_1,\dots,\mathcal{B}_n}(m_{\max})$ (i.e. defined by (3) with $m_1,\dots,m_n \in \{1,\dots,m_{\max}\}$),

$$\sup_{x\in\mathbb{R}^d}|K(x,x)| = \sup_{x\in\mathbb{R}^d}\prod_{q=1}^d\sum_{j=1}^{m_q}\varphi_j^{m_q}(x_q)^2.$$

Then, for $\mathcal{K}_n = \mathcal{K}_{\mathcal{B}_1,\ldots,\mathcal{B}_n}(m_{\max})$, at least for the usual bases mentioned at Remark 2.5,

$$K_0(x',x) = \prod_{q=1}^{d} \sum_{j=1}^{m_{\max}} \varphi_j^{m_{\max}}(x_q)\varphi_j^{m_{\max}}(x'_q) \ ; \ \forall x, x' \in \mathbb{R}^d.$$

In the sequel, in addition to Assumption 2.1, the kernels set $\mathcal{K}_n$ fulfills the following assumption.

Assumption 3.1. *There exists a deterministic constant $\overline{\mathfrak{m}}_{\mathcal{K},\ell} > 0$, not depending on n, such that*

$$\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n} \langle K(X_1,.), s_{K',\ell}\rangle_2^2\right) \leqslant \overline{\mathfrak{m}}_{\mathcal{K},\ell}.$$

The following theorem provides an oracle inequality for the adaptive estimator $\widehat{s}_{\widehat{K},\ell}(n;.)$.

Theorem 3.2. *Under Assumptions 2.1 and 3.1, if $s \in \mathbb{L}^2(\mathbb{R}^d)$ and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|\ell(Y_1)|)) < \infty$, then there exists a deterministic constant $\mathfrak{c}_{3.2} > 0$, not depending on n, such that for every $\vartheta \in]0,1[$,*

$$\mathbb{E}(\|\widehat{s}_{\widehat{K},\ell}(n;.) - s\|_2^2) \leqslant (1+\vartheta) \min_{K\in\mathcal{K}_n} \mathbb{E}(\|\widehat{s}_{K,\ell}(n;.) - s\|_2^2) + \frac{\mathfrak{c}_{3.2}}{\vartheta}\left(\|s_{K_0,\ell} - s\|_2^2 + \frac{\log(n)^5}{n}\right).$$

Remark 3.3. *As mentioned in Comte and Marie [4], p. 6, when $\mathcal{K}_n = \mathcal{K}_k(h_{\min})$, if s belongs to a Nikol'skii ball and $h_{\min} = 1/n$, then Theorem 3.2 says that the PCO estimator has a performance of same order than $O_n := \min_{K\in\mathcal{K}_n} \mathbb{E}(\|\widehat{s}_{K,\ell}(n;.) - s\|_2^2)$ up to a factor $1+\vartheta$. When $\mathcal{K}_n = \mathcal{K}_{\mathcal{B}_1,\ldots,\mathcal{B}_n}(m_{\max})$, it depends on the bases $\mathcal{B}_1,\ldots,\mathcal{B}_n$. For instance, with the same ideas than in Comte and Marie [4], thanks to DeVore and Lorentz [6], Theorem 2.3 p. 205, if s belongs to a Sobolev space and $m_{\max} = n$, then our Theorem 3.2 also says that the PCO estimator has a performance of same order than O_n.*

Notation. For any $B \in \mathcal{B}(\mathbb{R}^d)$, $\|.\|_{2,f,B}$ is the norm on $\mathbb{L}^2(B, f(x)\lambda_d(dx))$ defined by

$$\|\varphi\|_{2,f,B} := \left(\int_B \varphi(x)^2 f(x)\lambda_d(dx)\right)^{1/2} \ ; \ \forall \varphi \in \mathbb{L}^2(B, f(x)\lambda_d(dx)).$$

The following corollary provides an oracle inequality for $\widehat{s}_{\widehat{K},\ell}(n;.)/\widehat{s}_{\widehat{L},1}(n;.)$, where $\widehat{K}$ and $\widehat{L}$ are both selected via the PCO method.

Corollary 3.4. *Let $(\beta_j)_{j\in\mathbb{N}}$ be a decreasing sequence of elements of $]0,\infty[$ such that $\lim_\infty \beta_j = 0$ and, for every $j \in \mathbb{N}$, consider*

$$B_j := \{x \in \mathbb{R}^d : f(x) \geqslant \beta_j\}.$$

Under Assumptions 2.1 and 3.1 for ℓ and 1, if $s, f \in \mathbb{L}^2(\mathbb{R}^d)$ and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|\ell(Y_1)|)) < \infty$, then there exists a deterministic constant $\mathfrak{c}_{3.2} > 0$, not depending on n, such that for every $\vartheta \in]0,1[$,

$$\mathbb{E}\left[\left\|\frac{\widehat{s}_{\widehat{K},\ell}(n;.)}{\widehat{s}_{\widehat{L},1}(n;.)} - \frac{s}{f}\right\|_{2,f,B_n}^2\right] \leqslant \frac{\mathfrak{c}_{3.2}}{\beta_n^2}\left[(1+\vartheta)\min_{(K,L)\in\mathcal{K}_n^2}\{\mathbb{E}(\|\widehat{s}_{K,\ell}(n;.) - s\|_2^2) + \mathbb{E}(\|\widehat{s}_{L,1}(n;.) - f\|_2^2)\}\right.$$
$$\left. + \frac{1}{\vartheta}\left(\|s_{K_0,\ell} - s\|_2^2 + \|s_{K_0,1} - f\|_2^2 + \frac{\log(n)^5}{n}\right)\right]$$

where

$$\widehat{K} \in \arg\min_{K\in\mathcal{K}_n}\{\|\widehat{s}_{K,\ell}(n;\cdot) - \widehat{s}_{K_0,\ell}(n;\cdot)\|_2^2 + \mathrm{pen}_\ell(K)\}$$

and

$$\widehat{L} \in \arg\min_{L\in\mathcal{K}_n}\{\|\widehat{s}_{L,1}(n;\cdot) - \widehat{s}_{K_0,1}(n;\cdot)\|_2^2 + \mathrm{pen}_1(L)\}.$$

The proof of Corollary 3.4 is the same than the proof of Comte and Marie [4], Corollary 4.3.

Finally, let us discuss about Assumption 3.1. This assumption is difficult to check in practice, then let us provide a sufficient condition.

Assumption 3.5. *The function s is bounded and*

$$\mathfrak{m}_{\mathcal{K}} := \sup\{\|K(x',.)\|_1^2 \ ; \ K \in \mathcal{K}_n \text{ and } x' \in \mathbb{R}^d\}$$

doesn't depend on n.

Under Assumption 3.5, $\mathcal{K}_n$ fulfills Assumption 3.1. Indeed,

$$\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}\langle K(X_1,.),s_{K',\ell}\rangle_2^2\right) \leqslant \left(\sup_{K'\in\mathcal{K}_n}\|s_{K',\ell}\|_\infty^2\right)\mathbb{E}\left(\sup_{K\in\mathcal{K}_n}\|K(X_1,.)\|_1^2\right)$$
$$\leqslant \mathfrak{m}_{\mathcal{K}}\sup\left\{\left(\int_{-\infty}^{\infty}|K'(x',x)s(x)|dx\right)^2 \ ; \ K'\in\mathcal{K}_n \text{ and } x'\in\mathbb{R}\right\} \leqslant \mathfrak{m}_{\mathcal{K}}^2\|s\|_\infty^2.$$

Note that in the nonparametric regression framework (see Model (1)), to assume s bounded means that bf is bounded. For instance, this condition is fulfilled by the linear regression models with Gaussian inputs.

Let us provide two examples of kernels sets fulfilling Assumption 3.5, the sufficient condition for Assumption 3.1:

- Consider $K \in \mathcal{K}_k(h_{\min})$. Then, there exist $h_1,\dots,h_d \in \mathcal{H}(h_{\min})$ such that
$$K(x',x) = \prod_{q=1}^d \frac{1}{h_q}k\left(\frac{x'_q - x_q}{h_q}\right) \ ; \forall x,x' \in \mathbb{R}^d.$$
Clearly, $\|K(x',.)\|_1 = \|k\|_1^d$ for every $x' \in \mathbb{R}^d$. So, for $\mathcal{K}_n = \mathcal{K}_k(h_{\min})$, $\mathfrak{m}_{\mathcal{K}} \leqslant \|k\|_1^{2d}$.
- For $\mathcal{K}_n = \mathcal{K}_{\mathcal{B}_1,\dots,\mathcal{B}_n}(m_{\max})$, the condition on $\mathfrak{m}_{\mathcal{K}}$ seems harder to check in general. Let us show that it is satisfied for the regular histograms basis defined in Section 2. For every $m_1,\dots,m_d \in \{1,\dots,n\}$,
$$\left\|\prod_{q=1}^d\sum_{j=1}^{m_q}\psi_j^{m_q}(x'_q)\psi_j^{m_q}(.)\right\|_1 \leqslant \prod_{q=1}^d\left(m_q\sum_{j=1}^{m_q}\mathbf{1}_{[(j-1)/m_q,j/m_q[}(x'_q)\int_{(j-1)/m_q}^{j/m_q}dx\right) \leqslant 1.$$

Now, let us show that even if it doesn't fulfill Assumption 3.5, the trigonometric basis fulfills Assumption 3.1.

Proposition 3.6. *Consider $\chi_1 := \mathbf{1}_{[0,1]}$ and, for every $j \in \mathbb{N}^*$, the functions χ_{2j} and χ_{2j+1} defined on $\mathbb{R}$ by*

$$\chi_{2j}(x) := \sqrt{2}\cos(2\pi jx)\mathbf{1}_{[0,1]}(x) \text{ and } \chi_{2j+1}(x) := \sqrt{2}\sin(2\pi jx)\mathbf{1}_{[0,1]}(x) \ ; \forall x \in \mathbb{R}.$$

If $s \in C^2(\mathbb{R}^d)$ and $\mathcal{B}_m = \{\chi_1,\dots,\chi_m\}$ for every $m \in \{1,\dots,n\}$, then $\mathcal{K}_{\mathcal{B}_1,\dots,\mathcal{B}_n}(m_{\max})$ fulfills Assumption 3.1.

Appendix A. Details on kernels sets: proofs of Propositions 2.2, 2.3, 2.6 and 3.6

A.1. **Proof of Proposition 2.2.** Consider $K,K' \in \mathcal{K}_k(h_{\min})$. Then, there exist $h,h' \in \mathcal{H}(h_{\min})^d$ such that

$$K(x',x) = k_h(x'-x) \text{ and } K'(x',x) = k_{h'}(x'-x)$$

for every $x,x' \in \mathbb{R}^d$, where

$$k_h(x) := \prod_{q=1}^d \frac{1}{h_q}k\left(\frac{x_q}{h_q}\right) \ ; \forall x \in \mathbb{R}^d.$$

(1) For every $x' \in \mathbb{R}^d$, since $nh_{\min}^d \geqslant 1$,

$$\|K(x',.)\|_2^2 = \left(\prod_{q=1}^d\frac{1}{h_q^2}\right)\left[\int_{\mathbb{R}^d}\prod_{q=1}^d k\left(\frac{x'_q-x_q}{h_q}\right)^2\lambda_d(dx)\right] = \|k\|_2^{2d}\prod_{q=1}^d\frac{1}{h_q} \tag{9}$$
$$\leqslant \|k\|_2^{2d}\frac{1}{h_{\min}^d} \leqslant \|k\|_2^{2d}n.$$

(2) Since $s_{K,\ell} = K * s$ and by Young's inequality, $\|s_{K,\ell}\|_2^2 \leqslant \|k\|_1^{2d}\|s\|_2^2$.
(3) On the one hand, thanks to Equality (9),

$$\overline{s}_{K',\ell} = \mathbb{E}(\|K'(X_1,.)\ell(Y_1)\|_2^2) = \|k\|_2^{2d}\mathbb{E}(\ell(Y_1)^2)\prod_{q=1}^{d}\frac{1}{h'_q}.$$

On the other hand, for every $x, x' \in \mathbb{R}^d$,

$$\langle K(x,.), K'(x',.)\rangle_2 = \int_{\mathbb{R}^d} k_h(x-x'')k_{h'}(x'-x'')\lambda_d(dx'') = (k_h * k_{h'})(x-x').$$

Then,

$$\begin{aligned}\mathbb{E}(\langle K(X_1,.), K'(X_2,.)\ell(Y_2)\rangle_2^2) &= \mathbb{E}((k_h * k_{h'})(X_1-X_2)^2\ell(Y_2)^2)\\ &= \int_{\mathbb{R}^{d+1}}\left[\ell(y)^2\int_{\mathbb{R}^d}(k_h * k_{h'})(x'-x)^2 f(x')\lambda_d(dx')\right]\mathbb{P}_{(X_2,Y_2)}(dx,dy)\\ &\leqslant \|f\|_\infty\|k_h * k_{h'}\|_2^2\mathbb{E}(\ell(Y_2)^2) \leqslant \|f\|_\infty\|k\|_1^{2d}\overline{s}_{K',\ell}.\end{aligned}$$

(4) For every $\psi \in \mathbb{L}^2(\mathbb{R}^d)$,

$$\begin{aligned}\mathbb{E}(\langle K(X_1,.),\psi\rangle_2^2) &= \mathbb{E}((k_h * \psi)(X_1)^2)\\ &\leqslant \|f\|_\infty\|k_h * \psi\|_2^2 \leqslant \|f\|_\infty\|k\|_1^{2d}\|\psi\|_2^2.\end{aligned}$$

A.2. **Proof of Proposition 2.3.** Consider $K, K' \in \mathcal{K}_{\mathcal{B}_1,\ldots,\mathcal{B}_n}(m_{\max})$. Then, there exist $m, m' \in \{1,\ldots,m_{\max}\}^d$ such that

$$K(x',x) = \prod_{q=1}^{d}\sum_{j=1}^{m_q}\varphi_j^{m_q}(x_q)\varphi_j^{m_q}(x'_q) \text{ and } K'(x',x) = \prod_{q=1}^{d}\sum_{j=1}^{m'_q}\varphi_j^{m'_q}(x_q)\varphi_j^{m'_q}(x'_q)$$

for every $x, x' \in \mathbb{R}^d$.

(1) For every $x' \in \mathbb{R}^d$, since $m_{\max}^d \leqslant n$,

$$\begin{aligned}\|K(x',.)\|_2^2 &= \prod_{q=1}^{d}\sum_{j,j'=1}^{m_q}\varphi_{j'}^{m_q}(x'_q)\varphi_j^{m_q}(x'_q)\int_{-\infty}^{\infty}\varphi_{j'}^{m_q}(x)\varphi_j^{m_q}(x)dx = \prod_{q=1}^{d}\sum_{j=1}^{m_q}\varphi_j^{m_q}(x'_q)^2\\ &\leqslant \mathfrak{m}_{\mathcal{B}}^d\prod_{q=1}^{d}m_q \leqslant \mathfrak{m}_{\mathcal{B}}^d n.\end{aligned} \tag{10}$$

(2) Since

$$s_{K,\ell}(.) = \sum_{j_1=1}^{m_1}\cdots\sum_{j_d=1}^{m_d}\langle s, \varphi_{j_1}^{m_1}\otimes\cdots\otimes\varphi_{j_d}^{m_d}\rangle_2(\varphi_{j_1}^{m_1}\otimes\cdots\otimes\varphi_{j_d}^{m_d})(.),$$

by Pythagoras theorem, $\|s_{K,\ell}\|_2^2 \leqslant \|s\|_2^2$.
(3) First of all, thanks to Equality (10),

$$\overline{s}_{K',\ell} = \mathbb{E}\left[\ell(Y_1)^2\prod_{q=1}^{d}\sum_{j=1}^{m'_q}\varphi_j^{m'_q}(X_{1,q})^2\right] \leqslant \mathfrak{m}_{\mathcal{B}}^d\mathbb{E}(\ell(Y_1)^2)\prod_{q=1}^{d}m'_q.$$

On the one hand, under Condition (5) on $\mathcal{B}_1,\dots,\mathcal{B}_n$, for any $j\in\{1,\dots,m\}$, φ_j^m doesn't depend on m, so it can be denoted by φ_j, and then

$$\begin{aligned}\mathbb{E}(\langle K(X_1,.),K'(X_2,.)\ell(Y_2)\rangle_2^2) &= \int_{\mathbb{R}^d}\mathbb{E}\left[\left(\prod_{q=1}^d\sum_{j=1}^{m_q\wedge m_q'}\varphi_j(x_q')\varphi_j(X_{2,q})\right)^2\ell(Y_2)^2\right]f(x')\lambda_d(dx')\\ &\leqslant \|f\|_\infty\mathbb{E}\left[\ell(Y_2)^2\prod_{q=1}^d\sum_{j,j'=1}^{m_q\wedge m_q'}\varphi_{j'}(X_{2,q})\varphi_j(X_{2,q})\int_{-\infty}^{\infty}\varphi_{j'}(x')\varphi_j(x')dx'\right]\\ &\leqslant \|f\|_\infty\overline{s}_{K',\ell}.\end{aligned}$$

On the other hand, under Condition (6) on $\mathcal{B}_1,\dots,\mathcal{B}_n$, since X_1 and (X_2,Y_2) are independent, and since $K(x,x)\geqslant 0$ for every $x\in\mathbb{R}^d$,

$$\begin{aligned}\mathbb{E}(\langle K(X_1,.),K'(X_2,.)\ell(Y_2)\rangle_2^2) &\leqslant \mathbb{E}(\|K(X_1,.)\|_2^2\|K'(X_2,.)\|_2^2\ell(Y_2)^2)\\ &= \mathbb{E}(K(X_1,X_1))\mathbb{E}(\|K'(X_2,.)\|_2^2\ell(Y_2)^2)\leqslant\overline{\mathfrak{m}}_{\mathcal{B}}\overline{s}_{K',\ell}.\end{aligned}$$

(4) For every $\psi\in\mathbb{L}^2(\mathbb{R}^d)$,

$$\begin{aligned}\mathbb{E}(\langle K(X_1,.),\psi\rangle_2^2) &= \mathbb{E}\left[\left|\sum_{j_1=1}^{m_1}\cdots\sum_{j_d=1}^{m_d}\langle\psi,\varphi_{j_1}^{m_1}\otimes\cdots\otimes\varphi_{j_d}^{m_d}\rangle_2(\varphi_{j_1}^{m_1}\otimes\cdots\otimes\varphi_{j_d}^{m_d})(X_1)\right|^2\right]\\ &\leqslant \|f\|_\infty\left\|\sum_{j_1=1}^{m_1}\cdots\sum_{j_d=1}^{m_d}\langle\psi,\varphi_{j_1}^{m_1}\otimes\cdots\otimes\varphi_{j_d}^{m_d}\rangle_2(\varphi_{j_1}^{m_1}\otimes\cdots\otimes\varphi_{j_d}^{m_d})(.)\right\|_2^2\leqslant\|f\|_\infty\|\psi\|_2^2.\end{aligned}$$

A.3. **Proof of Proposition 2.6.** For the sake of readability, assume that $d=1$. Consider $m\in\{1,\dots,m_{\max}\}$. Since each Legendre's polynomial is uniformly bounded by 1,

$$\left|\mathbb{E}\left[\sum_{j=1}^m\xi_j^m(X_1)\xi_j^m(x')\right]\right|\leqslant\sum_{j=1}^m\frac{2j+1}{2}\left|\int_{-1}^1Q_j(x)f(x)dx\right|.$$

Moreover, since Q_j is a solution to Legendre's differential equation for any $j\in\{1,\dots,m\}$, thanks to the integration by parts formula,

$$\begin{aligned}\int_{-1}^1Q_j(x)f(x)dx &= -\frac{1}{j(j+1)}\int_{-1}^1\frac{d}{dx}[(1-x^2)Q_j'(x)]f(x)dx\\ &= -\frac{1}{j(j+1)}[(1-x^2)Q_j'(x)f(x)]_{-1}^1+\frac{1}{j(j+1)}\int_{-1}^1(1-x^2)Q_j'(x)f'(x)dx\\ &= -\frac{1}{j(j+1)}\int_{-1}^1Q_j(x)\frac{d}{dx}[(1-x^2)f'(x)]dx.\end{aligned}$$

Then,

$$\left|\int_{-1}^1Q_j(x)f(x)dx\right|\leqslant\frac{2\mathfrak{c}_1}{j(j+1)}\|Q_j\|_2=\frac{2\sqrt{2}\mathfrak{c}_1}{j(j+1)(2j+1)^{1/2}}$$

with $\mathfrak{c}_1=\max\{2\|f'\|_\infty,\|f''\|_\infty\}$. So,

$$\left|\mathbb{E}\left[\sum_{j=1}^m\xi_j^m(X_1)\xi_j^m(x')\right]\right|\leqslant 2\mathfrak{c}_1\sum_{j=1}^m\frac{1}{j^{3/2}}\leqslant 2\mathfrak{c}_1\zeta\left(\frac{3}{2}\right)$$

where ζ is Riemann's zeta function. Thus, Legendre's basis satisfies Condition (6).

A.4. **Proof of Proposition 3.6.** The proof of Proposition 3.6 relies on the following technical lemma.

Lemma A.1. *For every $x \in [0, 2\pi]$ and $p, q \in \mathbb{N}^*$ such that $q > p$,*

$$\left| \sum_{j=p+1}^{q} \frac{\sin(jx)}{j} \right| \leqslant \frac{2}{(1+p)\sin(x/2)}.$$

See Subsubsection A.4.1 for a proof.

For the sake of readability, assume that $d = 1$. Consider $K, K' \in \mathcal{K}_{\mathcal{B}_1,\dots,\mathcal{B}_n}(m_{\max})$. Then, there exist $m, m' \in \{1, \dots, m_{\max}\}$ such that

$$K(x', x) = \sum_{j=1}^{m} \chi_j(x)\chi_j(x') \text{ and } K'(x', x) = \sum_{j=1}^{m'} \chi_j(x)\chi_j(x') \ ; \forall x, x' \in \mathbb{R}.$$

First, there exist $\mathfrak{m}_1(m, m') \in \{0, \dots, n\}$ and $\mathfrak{c}_1 > 0$, not depending on n, K and K', such that for any $x' \in [0, 1]$,

$$\begin{aligned}
|\langle K(x', .), s_{K',\ell}\rangle_2| &= \left| \sum_{j=1}^{m\wedge m'} \mathbb{E}(\ell(Y_1)\chi_j(X_1))\chi_j(x') \right| \\
&\leqslant \mathfrak{c}_1 + 2 \left| \sum_{j=1}^{\mathfrak{m}_1(m,m')} \mathbb{E}(\ell(Y_1)(\cos(2\pi jX_1)\cos(2\pi jx') + \sin(2\pi jX_1)\sin(2\pi jx'))\mathbf{1}_{[0,1]}(X_1)) \right| \\
&= \mathfrak{c}_1 + 2 \left| \sum_{j=1}^{\mathfrak{m}_1(m,m')} \mathbb{E}(\ell(Y_1)\cos(2\pi j(X_1 - x'))\mathbf{1}_{[0,1]}(X_1)) \right|.
\end{aligned}$$

Moreover, for any $j \in \{2, \dots, \mathfrak{m}_1(m, m')\}$,

$$\begin{aligned}
\mathbb{E}(\ell(Y_1)\cos(2\pi j(X_1 - x'))\mathbf{1}_{[0,1]}(X_1)) &= \int_0^1 \cos(2\pi j(x - x'))s(x)dx \\
&= \frac{1}{j}\left[\frac{\sin(2\pi j(x - x'))}{2\pi}s(x)\right]_0^1 \\
&\quad + \frac{1}{j^2}\left[\frac{\cos(2\pi j(x - x'))}{4\pi^2}s'(x)\right]_0^1 - \frac{1}{j^2}\int_0^1 \frac{\cos(2\pi j(x - x'))}{4\pi^2}s''(x)dx \\
&= \frac{s(0) - s(1)}{2\pi} \cdot \frac{\alpha_j(x')}{j} + \frac{\beta_j(x')}{j^2}
\end{aligned}$$

where $\alpha_j(x') := \sin(2\pi jx')$ and

$$\beta_j(x') := \frac{1}{4\pi^2}\left((s'(1) - s'(0))\cos(2\pi jx') - \int_0^1 \cos(2\pi j(x - x'))s''(x)dx\right).$$

Then, there exists a deterministic constant $\mathfrak{c}_2 > 0$, not depending on n, K, K' and x', such that

$$\langle K(x', .), s_{K',\ell}\rangle_2^2 \leqslant \mathfrak{c}_2 \left[1 + \left(\sum_{j=1}^{\mathfrak{m}_1(m,m')} \frac{\alpha_j(x')}{j}\right)^2 + \left(\sum_{j=1}^{\mathfrak{m}_1(m,m')} \frac{\beta_j(x')}{j^2}\right)^2\right]. \tag{11}$$

Let us show that each term of the right-hand side of Inequality (11) is uniformly bounded in x', m and m'. On the one hand,

$$\left| \sum_{j=1}^{\mathfrak{m}_1(m,m')} \frac{\beta_j(x')}{j^2} \right| \leqslant \max_{j\in\{1,\dots,n\}} \|\beta_j\|_\infty \sum_{j=1}^{n} \frac{1}{j^2} \leqslant \frac{1}{24}(2\|s'\|_\infty + \|s''\|_\infty).$$

On the other hand, for every $x \in]0, \pi[$ such that $[\pi/x] + 1 \leqslant \mathfrak{m}_1(m, m')$ (without loss of generality), by Lemma A.1,

$$\begin{aligned}\left|\sum_{j=1}^{\mathfrak{m}_1(m,m')} \frac{\sin(jx)}{j}\right| &\leqslant \left|\sum_{j=1}^{[\pi/x]} \frac{\sin(jx)}{j}\right| + \left|\sum_{j=[\pi/x]+1}^{\mathfrak{m}_1(m,m')} \frac{\sin(jx)}{j}\right| \\ &\leqslant x\left[\frac{\pi}{x}\right] + \frac{2}{(1+[\pi/x])\sin(x/2)} \leqslant \pi + 2. \qquad (12)\end{aligned}$$

Since $x \mapsto \sin(x)$ is continuous, odd and 2π-periodic, Inequality (12) holds true for every $x \in \mathbb{R}$. So,

$$\left|\sum_{j=1}^{\mathfrak{m}_1(m,m')} \frac{\alpha_j(x')}{j}\right| \leqslant \pi + 2.$$

Therefore,

$$\mathbb{E}\left[\sup_{K,K' \in \mathcal{K}_{\mathcal{B}_1,\ldots,\mathcal{B}_n}(m_{\max})} \langle K(X_1,.), s_{K',\ell}\rangle_2^2\right] \leqslant \mathfrak{c}_2\left(1 + (\pi+2)^2 + \frac{1}{24^2}(2\|s'\|_\infty + \|s''\|_\infty)^2\right).$$

A.4.1. *Proof of Lemma A.1.* For any $x \in [0, 2\pi]$ and $q \in \mathbb{N}^*$, consider

$$f_q(x) := \sum_{j=1}^{q} \frac{\sin(jx)}{j},\ g_q(x) := \sum_{j=1}^{q}\left(\frac{1}{j} - \frac{1}{j+1}\right) h_j(x) \text{ and } h_q(x) := \sum_{j=1}^{q} \sin(jx).$$

On the one hand,

$$g_q(x) = h_1(x) - \frac{1}{q+1}h_q(x) + \sum_{j=2}^{q} \frac{1}{j}(h_j(x) - h_{j-1}(x)).$$

Then,

$$f_q(x) = g_q(x) + \frac{1}{q+1}h_q(x).$$

On the other hand,

$$\begin{aligned}h_q(x) &= \mathrm{Im}\left(\sum_{j=1}^{q} e^{\mathbf{i}jx}\right) = \mathrm{Im}\left[e^{\mathbf{i}(q+1)x/2}\frac{\sin(qx/2)}{\sin(x/2)}\right] \\ &= \frac{\sin((q+1)x/2)\sin(qx/2)}{\sin(x/2)} = \frac{\cos(x/2) - \cos((q+1/2)x)}{2\sin(x/2)}.\end{aligned}$$

Then,

$$\sin\left(\frac{x}{2}\right)|h_q(x)| \leqslant 1$$

and, for any $p \in \mathbb{N}^*$ such that $q > p$,

$$\sin\left(\frac{x}{2}\right)|g_q(x) - g_p(x)| \leqslant \frac{1}{p+1} - \frac{1}{q+1}.$$

Therefore,

$$\begin{aligned}\sin\left(\frac{x}{2}\right)|f_q(x) - f_p(x)| &\leqslant \sin\left(\frac{x}{2}\right)|g_q(x) - g_p(x)| + \sin\left(\frac{x}{2}\right)\frac{|h_q(x)|}{q+1} + \sin\left(\frac{x}{2}\right)\frac{|h_p(x)|}{p+1} \\ &\leqslant \frac{2}{p+1}.\end{aligned}$$

In conclusion,

$$\left|\sum_{j=p+1}^{q} \frac{\sin(jx)}{k}\right| \leqslant \frac{2}{(1+p)\sin(x/2)}.$$

Appendix B. Proofs of risk bounds

B.1. **Preliminary results.** This subsection provides three lemmas used several times in the sequel.

Lemma B.1. *Consider*

$$U_{K,K',\ell}(n) := \sum_{i\neq j} \langle K(X_i,.)\ell(Y_i) - s_{K,\ell}, K'(X_j,.)\ell(Y_j) - s_{K',\ell}\rangle_2 \ ; \ \forall K, K' \in \mathcal{K}_n. \tag{13}$$

Under Assumption 2.1.(1,2,3), if $s \in \mathbb{L}^2(\mathbb{R}^d)$ and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|\ell(Y_1)|)) < \infty$, then there exists a deterministic constant $\mathfrak{c}_{B.1} > 0$, not depending on n, such that for every $\theta \in]0,1[$,

$$\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}\left\{\frac{|U_{K,K',\ell}(n)|}{n^2} - \frac{\theta}{n}\overline{s}_{K',\ell}\right\}\right) \leqslant \mathfrak{c}_{B.1}\frac{\log(n)^5}{\theta n}.$$

Lemma B.2. *Consider*

$$V_{K,\ell}(n) := \frac{1}{n}\sum_{i=1}^{n} \|K(X_i,.)\ell(Y_i) - s_{K,\ell}\|_2^2 \ ; \ \forall K \in \mathcal{K}_n.$$

Under Assumption 2.1.(1,2), if $s \in \mathbb{L}^2(\mathbb{R}^d)$ and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|\ell(Y_1)|)) < \infty$, then there exists a deterministic constant $\mathfrak{c}_{B.2} > 0$, not depending on n, such that for every $\theta \in]0,1[$,

$$\mathbb{E}\left(\sup_{K\in\mathcal{K}_n}\left\{\frac{1}{n}|V_{K,\ell}(n) - \overline{s}_{K,\ell}| - \frac{\theta}{n}\overline{s}_{K,\ell}\right\}\right) \leqslant \mathfrak{c}_{B.2}\frac{\log(n)^3}{\theta n}.$$

Lemma B.3. *Consider*

$$W_{K,K',\ell}(n) := \langle \widehat{s}_{K,\ell}(n;.) - s_{K,\ell}, s_{K',\ell} - s\rangle_2 \ ; \ \forall K, K' \in \mathcal{K}_n. \tag{14}$$

Under Assumption 2.1.(1,2,4), if $s \in \mathbb{L}^2(\mathbb{R}^d)$ and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|\ell(Y_1)|)) < \infty$, then there exists a deterministic constant $\mathfrak{c}_{B.3} > 0$, not depending on n, such that for every $\theta \in]0,1[$,

$$\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}\{|W_{K,K',\ell}(n)| - \theta\|s_{K',\ell} - s\|_2^2\}\right) \leqslant \mathfrak{c}_{B.3}\frac{\log(n)^4}{\theta n}.$$

B.1.1. *Proof of Lemma B.1.* The proof of Lemma B.1 relies on the following concentration inequality for U-statistics, proved in dimension 1 in Houdré and Reynaud-Bouret [11] first, and then extended to the infinite-dimensional framework by Giné and Nickl in [9].

Lemma B.4. *Let $\xi_1, \ldots, \xi_n$ be i.i.d. random variables on a Polish space Ξ equipped with its Borel σ-algebra. Let $f_{i,j}$, $1 \leqslant i \neq j \leqslant n$, be some bounded and symmetric measurable maps from Ξ^2 into $\mathbb{R}$ such that, for every $i \neq j$,*

$$f_{i,j} = f_{j,i} \text{ and } \mathbb{E}(f_{i,j}(z,\xi_1)) = 0 \ dz\text{-a.e.}$$

Consider the totally degenerate second order U-statistic

$$U_n := \sum_{i\neq j} f_{i,j}(\xi_i,\xi_j).$$

There exists a universal constant $\mathfrak{m} > 0$ such that for every $\lambda > 0$,

$$\mathbb{P}(U_n \leqslant \mathfrak{m}(\mathfrak{c}_n\lambda^{1/2} + \mathfrak{d}_n\lambda + \mathfrak{b}_n\lambda^{3/2} + \mathfrak{a}_n\lambda^2)) \geqslant 1 - 2.7e^{-\lambda}$$

where

$$\begin{aligned}
\mathfrak{a}_n &= \sup_{i,j=1,\dots,n}\left\{\sup_{z,z'\in\Xi}|f_{i,j}(z,z')|\right\},\\
\mathfrak{b}_n^2 &= \max\left\{\sup_{i,z}\sum_{j=1}^{i-1}\mathbb{E}(f_{i,j}(z,\xi_j)^2)\ ;\ \sup_{j,z'}\sum_{i=j+1}^{n}\mathbb{E}(f_{i,j}(\xi_i,z')^2)\right\},\\
\mathfrak{c}_n^2 &= \sum_{i\neq j}\mathbb{E}(f_{i,j}(\xi_i,\xi_j)^2)\ \textit{and}\\
\mathfrak{d}_n &= \sup_{(a,b)\in\mathcal{A}}\mathbb{E}\left[\sum_{i<j}f_{i,j}(\xi_i,\xi_j)a_i(\xi_i)b_j(\xi_j)\right]
\end{aligned}$$

with

$$\mathcal{A}=\left\{(a,b):\mathbb{E}\left(\sum_{i=1}^{n-1}a_i(\xi_i)^2\right)\leqslant 1\ \textit{and}\ \mathbb{E}\left(\sum_{j=2}^{n}b_j(\xi_j)^2\right)\leqslant 1\right\}.$$

See Giné and Nickl [9], Theorem 3.4.8 for a proof.

Consider $\mathfrak{m}(n):=8\log(n)/\alpha$. For any $K,K'\in\mathcal{K}_n$,

$$U_{K,K',\ell}(n)=U^1_{K,K',\ell}(n)+U^2_{K,K',\ell}(n)+U^3_{K,K',\ell}(n)+U^4_{K,K',\ell}(n)$$

where

$$U^l_{K,K',\ell}(n):=\sum_{i\neq j}g^l_{K,K',\ell}(n;X_i,Y_i,X_j,Y_j)\ ;\ l=1,2,3,4$$

with, for every $(x',y),(x'',y')\in E=\mathbb{R}^d\times\mathbb{R}$,

$$\begin{aligned}
g^1_{K,K',\ell}(n;x',y,x'',y') &:= \langle K(x',.)\ell(y)\mathbf{1}_{|\ell(y)|\leqslant\mathfrak{m}(n)}-s^+_{K,\ell}(n;.),K'(x'',.)\ell(y')\mathbf{1}_{|\ell(y)|\leqslant\mathfrak{m}(n)}-s^+_{K',\ell}(n;.)\rangle_2,\\
g^2_{K,K',\ell}(n;x',y,x'',y') &:= \langle K(x',.)\ell(y)\mathbf{1}_{|\ell(y)|>\mathfrak{m}(n)}-s^-_{K,\ell}(n;.),K'(x'',.)\ell(y')\mathbf{1}_{|\ell(y)|\leqslant\mathfrak{m}(n)}-s^+_{K',\ell}(n;.)\rangle_2,\\
g^3_{K,K',\ell}(n;x',y,x'',y') &:= \langle K(x',.)\ell(y)\mathbf{1}_{|\ell(y)|\leqslant\mathfrak{m}(n)}-s^+_{K,\ell}(n;.),K'(x'',.)\ell(y')\mathbf{1}_{|\ell(y)|>\mathfrak{m}(n)}-s^-_{K',\ell}(n;.)\rangle_2,\\
g^4_{K,K',\ell}(n;x',y,x'',y') &:= \langle K(x',.)\ell(y)\mathbf{1}_{|\ell(y)|>\mathfrak{m}(n)}-s^-_{K,\ell}(n;.),K'(x'',.)\ell(y')\mathbf{1}_{|\ell(y)|>\mathfrak{m}(n)}-s^-_{K',\ell}(n;.)\rangle_2
\end{aligned}$$

and, for every $k\in\mathcal{K}_n$,

$$s^+_{k,\ell}(n;.):=\mathbb{E}(k(X_1,.)\ell(Y_1)\mathbf{1}_{|\ell(Y_1)|\leqslant\mathfrak{m}(n)})\text{ and }s^-_{k,\ell}(n;.):=\mathbb{E}(k(X_1,.)\ell(Y_1)\mathbf{1}_{|\ell(Y_1)|>\mathfrak{m}(n)}).$$

On the one hand, since $\mathbb{E}(g^1_{K,K',\ell}(n;x',y,X_1,Y_1))=0$ for every $(x',y)\in E$, by Lemma B.4, there exists a universal constant $\mathfrak{m}\geqslant 1$ such that for any $\lambda>0$, with probability larger than $1-5.4e^{-\lambda}$,

$$\frac{|U^1_{K,K',\ell}(n)|}{n^2}\leqslant\frac{\mathfrak{m}}{n^2}(\mathfrak{c}_{K,K',\ell}(n)\lambda^{1/2}+\mathfrak{d}_{K,K',\ell}(n)\lambda+\mathfrak{b}_{K,K',\ell}(n)\lambda^{3/2}+\mathfrak{a}_{K,K',\ell}(n)\lambda^2)$$

where the constants $\mathfrak{a}_{K,K',\ell}(n)$, $\mathfrak{b}_{K,K',\ell}(n)$, $\mathfrak{c}_{K,K',\ell}(n)$ and $\mathfrak{d}_{K,K',\ell}(n)$ are defined and controlled later. First, note that

$$\begin{aligned}
U^1_{K,K',\ell}(n) = \sum_{i\neq j}(&\varphi_{K,K',\ell}(n;X_i,Y_i,X_j,Y_j)\\
&-\psi_{K,K',\ell}(n;X_i,Y_i)-\psi_{K',K,\ell}(n;X_j,Y_j)+\mathbb{E}(\varphi_{K,K',\ell}(n;X_i,Y_i,X_j,Y_j))),
\end{aligned}\tag{15}$$

where

$$\varphi_{K,K',\ell}(n;x',y,x'',y'):=\langle K(x',.)\ell(y)\mathbf{1}_{|\ell(y)|\leqslant\mathfrak{m}(n)},K'(x'',.)\ell(y')\mathbf{1}_{|\ell(y')|\leqslant\mathfrak{m}(n)}\rangle_2$$

and

$$\psi_{k,k',\ell}(n;x',y):=\langle k(x',.)\ell(y)\mathbf{1}_{|\ell(y)|\leqslant\mathfrak{m}(n)},s^+_{k',\ell}(n;.)\rangle_2=\mathbb{E}(\varphi_{k,k',\ell}(n;x',y,X_1,Y_1))$$

for every $k,k'\in\mathcal{K}_n$ and $(x',y),(x'',y')\in E$. Let us now control $\mathfrak{a}_{K,K',\ell}(n)$, $\mathfrak{b}_{K,K',\ell}(n)$, $\mathfrak{c}_{K,K',\ell}(n)$ and $\mathfrak{d}_{K,K',\ell}(n)$:

- **The constant $\mathfrak{a}_{K,K',\ell}(n)$.** Consider

$$\mathfrak{a}_{K,K',\ell}(n) := \sup_{(x',y),(x'',y')\in E} |g^1_{K,K',\ell}(n;x',y,x'',y')|.$$

By (15), Cauchy-Schwarz's inequality and Assumption 2.1.(1),

$$\begin{aligned}\mathfrak{a}_{K,K',\ell}(n) &\leqslant 4 \sup_{(x',y),(x'',y')\in E} |\langle K(x',.)\ell(y)\mathbf{1}_{|\ell(y)|\leqslant \mathfrak{m}(n)}, K'(x'',.)\ell(y')\mathbf{1}_{|\ell(y')|\leqslant \mathfrak{m}(n)}\rangle_2| \\ &\leqslant 4\mathfrak{m}(n)^2 \left(\sup_{x'\in\mathbb{R}^d} \|K(x',.)\|_2\right)\left(\sup_{x''\in\mathbb{R}^d} \|K'(x'',.)\|_2\right) \leqslant 4\mathfrak{m}_{\mathcal{K},\ell}\mathfrak{m}(n)^2 n.\end{aligned}$$

So,

$$\frac{1}{n^2}\mathfrak{a}_{K,K',\ell}(n)\lambda^2 \leqslant \frac{4}{n}\mathfrak{m}_{\mathcal{K},\ell}\mathfrak{m}(n)^2\lambda^2.$$

- **The constant $\mathfrak{b}_{K,K',\ell}(n)$.** Consider

$$\mathfrak{b}_{K,K',\ell}(n)^2 := n \sup_{(x',y)\in E} \mathbb{E}(g^1_{K,K',\ell}(n;x',y,X_1,Y_1)^2).$$

By (15), Jensen's inequality, Cauchy-Schwarz's inequality and Assumption 2.1.(1),

$$\begin{aligned}\mathfrak{b}_{K,K',\ell}(n)^2 &\leqslant 16n \sup_{(x',y)\in E} \mathbb{E}(\langle K(x',.)\ell(y)\mathbf{1}_{|\ell(y)|\leqslant \mathfrak{m}(n)}, K'(X_1,.)\ell(Y_1)\mathbf{1}_{|\ell(Y_1)|\leqslant \mathfrak{m}(n)}\rangle_2^2) \\ &\leqslant 16n\mathfrak{m}(n)^2 \sup_{x'\in\mathbb{R}^d} \|K(x',.)\|_2^2 \mathbb{E}(\|K'(X_1,.)\ell(Y_1)\mathbf{1}_{|\ell(Y_1)|\leqslant \mathfrak{m}(n)}\|_2^2) \leqslant 16\mathfrak{m}_{\mathcal{K},\ell}n^2\mathfrak{m}(n)^2\overline{s}_{K',\ell}.\end{aligned}$$

So, for any $\theta\in]0,1[$,

$$\begin{aligned}\frac{1}{n^2}\mathfrak{b}_{K,K',\ell}(n)\lambda^{3/2} &\leqslant 2\left(\frac{3\mathfrak{m}}{\theta}\right)^{1/2}\frac{2}{n^{1/2}}\mathfrak{m}_{\mathcal{K},\ell}^{1/2}\mathfrak{m}(n)\lambda^{3/2} \times \left(\frac{\theta}{3\mathfrak{m}}\right)^{1/2}\frac{1}{n^{1/2}}\overline{s}_{K',\ell}^{1/2} \\ &\leqslant \frac{\theta}{3\mathfrak{m}n}\overline{s}_{K',\ell} + \frac{12\mathfrak{m}\lambda^3}{\theta n}\mathfrak{m}_{\mathcal{K},\ell}\mathfrak{m}(n)^2.\end{aligned}$$

- **The constant $\mathfrak{c}_{K,K',\ell}(n)$.** Consider

$$\mathfrak{c}_{K,K',\ell}(n)^2 := n^2\mathbb{E}(g^1_{K,K',\ell}(n;X_1,Y_1,X_2,Y_2)^2).$$

By (15), Jensen's inequality and Assumption 2.1.(3),

$$\begin{aligned}\mathfrak{c}_{K,K',\ell}(n)^2 &\leqslant 16n^2\mathbb{E}(\langle K(X_1,.)\ell(Y_1)\mathbf{1}_{|\ell(Y_1)|\leqslant \mathfrak{m}(n)}, K'(X_2,.)\ell(Y_2)\mathbf{1}_{|\ell(Y_2)|\leqslant \mathfrak{m}(n)}\rangle_2^2) \\ &\leqslant 16n^2\mathfrak{m}(n)^2\mathbb{E}(\langle K(X_1,.), K'(X_2,.)\ell(Y_2)\rangle_2^2) \leqslant 16\mathfrak{m}_{\mathcal{K},\ell}n^2\mathfrak{m}(n)^2\overline{s}_{K',\ell}.\end{aligned}$$

So,

$$\frac{1}{n^2}\mathfrak{c}_{K,K',\ell}(n)\lambda^{1/2} \leqslant \frac{\theta}{3\mathfrak{m}n}\overline{s}_{K',\ell} + \frac{12\mathfrak{m}\lambda}{\theta n}\mathfrak{m}_{\mathcal{K},\ell}\mathfrak{m}(n)^2.$$

- **The constant $\mathfrak{d}_{K,K',\ell}(n)$.** Consider

$$\mathfrak{d}_{K,K',\ell}(n) := \sup_{(a,b)\in\mathcal{A}} \mathbb{E}\left[\sum_{i<j} a_i(X_i,Y_i)b_j(X_j,Y_j)g^1_{K,K',\ell}(n;X_i,Y_i,X_j,Y_j)\right],$$

where

$$\mathcal{A} := \left\{(a,b) : \sum_{i=1}^{n-1}\mathbb{E}(a_i(X_i,Y_i)^2) \leqslant 1 \text{ and } \sum_{j=2}^{n}\mathbb{E}(b_j(X_j,Y_j)^2) \leqslant 1\right\}.$$

By (15), Jensen's inequality, Cauchy-Schwarz's inequality and Assumption 2.1.(3),

$$\begin{aligned}\mathfrak{d}_{K,K',\ell}(n) &\leqslant 4 \sup_{(a,b)\in\mathcal{A}} \mathbb{E}\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} |a_i(X_i,Y_i)b_j(X_j,Y_j)\varphi_{K,K',\ell}(n;X_i,Y_i,X_j,Y_j)|\right] \\ &\leqslant 4n\mathfrak{m}(n)\mathbb{E}(\langle K(X_1,.), K'(X_2,.)\ell(Y_2)\rangle_2^2)^{1/2} \leqslant 4\mathfrak{m}_{\mathcal{K},\ell}^{1/2}n\mathfrak{m}(n)\overline{s}_{K',\ell}^{1/2}.\end{aligned}$$

So,

$$\frac{1}{n^2}\mathfrak{d}_{K,K',\ell}(n)\lambda \leqslant \frac{\theta}{3\mathfrak{m}n}\overline{s}_{K',\ell} + \frac{12\mathfrak{m}\lambda^2}{\theta n}\mathfrak{m}_{\mathcal{K},\ell}\mathfrak{m}(n)^2.$$

Then, since $\mathfrak{m} \geqslant 1$ and $\lambda > 0$, with probability larger than $1 - 5.4e^{-\lambda}$,

$$\frac{|U^1_{K,K',\ell}(n)|}{n^2} \leqslant \frac{\theta}{n}\overline{s}_{K',\ell} + \frac{40\mathfrak{m}^2}{\theta n}\mathfrak{m}_{\mathcal{K},\ell}\mathfrak{m}(n)^2(1+\lambda)^3.$$

So, with probability larger than $1 - 5.4|\mathcal{K}_n|^2e^{-\lambda}$,

$$S_{\mathcal{K},\ell}(n,\theta) := \sup_{K,K'\in\mathcal{K}_n}\left\{\frac{|U^1_{K,K',\ell}(n)|}{n^2} - \frac{\theta}{n}\overline{s}_{K',\ell}\right\} \leqslant \frac{40\mathfrak{m}^2}{\theta n}\mathfrak{m}_{\mathcal{K},\ell}\mathfrak{m}(n)^2(1+\lambda)^3.$$

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

$$\lambda_{\mathcal{K},\ell}(n,\theta,t) := -1 + \left(\frac{t}{\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)}\right)^{1/3} \text{ with } \mathfrak{m}_{\mathcal{K},\ell}(n,\theta) = \frac{40\mathfrak{m}^2}{\theta n}\mathfrak{m}_{\mathcal{K},\ell}\mathfrak{m}(n)^2.$$

Then, for any $T > 0$,

$$\begin{aligned}
\mathbb{E}(S_{\mathcal{K},\ell}(n,\theta)) &\leqslant T + \int_T^\infty \mathbb{P}(S_{\mathcal{K},\ell}(n,\theta) \geqslant (1+\lambda_{\mathcal{K},\ell}(n,\theta,t))^3\mathfrak{m}_{\mathcal{K},\ell}(n,\theta))dt \\
&\leqslant T + 5.4|\mathcal{K}_n|^2\int_T^\infty \exp(-\lambda_{\mathcal{K},\ell}(n,\theta,t))dt \\
&= T + 5.4|\mathcal{K}_n|^2\int_T^\infty \exp\left(-\frac{t^{1/3}}{2\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)^{1/3}}\right)\exp\left(1 - \frac{t^{1/3}}{2\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)^{1/3}}\right)dt \\
&\leqslant T + 5.4\mathfrak{c}_1|\mathcal{K}_n|^2\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)\exp\left(-\frac{T^{1/3}}{2\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)^{1/3}}\right) \text{ with } \mathfrak{c}_1 = \int_0^\infty e^{1-r^{1/3}/2}dr.
\end{aligned}$$

Moreover,

$$\mathfrak{m}_{\mathcal{K},\ell}(n,\theta) \leqslant \mathfrak{c}_2\frac{\log(n)^2}{\theta n} \text{ with } \mathfrak{c}_2 = \frac{40\cdot 8^2\mathfrak{m}^2}{\alpha^2}\mathfrak{m}_{\mathcal{K},\ell}.$$

So, by taking

$$T = 2^4\mathfrak{c}_2\frac{\log(n)^5}{\theta n},$$

and since $|\mathcal{K}_n| \leqslant n$,

$$\mathbb{E}(S_{\mathcal{K},\ell}(n,\theta)) \leqslant 2^4\mathfrak{c}_2\frac{\log(n)^5}{\theta n} + 5.4\mathfrak{c}_1\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)\frac{|\mathcal{K}_n|^2}{n^2} \leqslant (2^4 + 5.4\mathfrak{c}_1)\mathfrak{c}_2\frac{\log(n)^5}{\theta n}.$$

On the other hand, by Assumption 2.1.(1), Cauchy-Schwarz's inequality and Markov's inequality,

$$\begin{aligned}
\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}|g^2_{K,K',\ell}(n;X_1,Y_1,X_2,Y_2)|\right) &\leqslant 4\mathfrak{m}(n)\sum_{K,K'\in\mathcal{K}_n}\mathbb{E}(|\ell(Y_1)|\mathbf{1}_{|\ell(Y_1)|>\mathfrak{m}(n)}|\langle K(X_1,.),K'(X_2,.)\rangle_2|) \\
&\leqslant 4\mathfrak{m}(n)\mathfrak{m}_{\mathcal{K},\ell}n|\mathcal{K}_n|^2\mathbb{E}(\ell(Y_1)^2)^{1/2}\mathbb{P}(|\ell(Y_1)| > \mathfrak{m}(n))^{1/2} \leqslant \mathfrak{c}_3\frac{\log(n)}{n}
\end{aligned}$$

with

$$\mathfrak{c}_3 = \frac{32}{\alpha}\mathfrak{m}_{\mathcal{K},\ell}\mathbb{E}(\ell(Y_1)^2)^{1/2}\mathbb{E}(\exp(\alpha|\ell(Y_1)|))^{1/2}.$$

So,

$$\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}\frac{|U^2_{K,K',\ell}(n)|}{n^2}\right) \leqslant \mathfrak{c}_3\frac{\log(n)}{n}$$

and, symmetrically,

$$\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}\frac{|U^3_{K,K',\ell}(n)|}{n^2}\right) \leqslant \mathfrak{c}_3\frac{\log(n)}{n}.$$

By Assumption 2.1.(1), Cauchy-Schwarz's inequality and Markov's inequality,

$$\begin{aligned}\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}|g^4_{K,K',\ell}(n;X_1,Y_1,X_2,Y_2)|\right) &\leqslant 4\sum_{K,K'\in\mathcal{K}_n}\mathbb{E}(|\ell(Y_1)\ell(Y_2)|\mathbf{1}_{|\ell(Y_1)|,|\ell(Y_2)|>\mathfrak{m}(n)}|\langle K(X_1,.),K'(X_2,.)\rangle_2|)\\ &\leqslant 4\mathfrak{m}_{\mathcal{K},\ell}n|\mathcal{K}_n|^2\mathbb{E}(\ell(Y_1)^2)\mathbb{P}(|\ell(Y_1)|>\mathfrak{m}(n))\leqslant\frac{\mathfrak{c}_4}{n^5}\end{aligned}$$

with

$$\mathfrak{c}_4=4\mathfrak{m}_{\mathcal{K},\ell}\mathbb{E}(\ell(Y_1)^2)\mathbb{E}(\exp(\alpha|\ell(Y_1)|)).$$

So,

$$\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}\frac{|U^4_{K,K',\ell}(n)|}{n^2}\right)\leqslant\frac{\mathfrak{c}_4}{n^5}.$$

Therefore,

$$\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}\left\{\frac{|U_{K,K',\ell}(n)|}{n^2}-\frac{\theta}{n}\overline{s}_{K',\ell}\right\}\right)\leqslant(2^4+5.4\mathfrak{c}_1)\mathfrak{c}_2\frac{\log(n)^5}{\theta n}+2\mathfrak{c}_3\frac{\log(n)}{n}+\frac{\mathfrak{c}_4}{n^5}.$$

B.1.2. *Proof of Lemma B.2.* First, the two following results are used several times in the sequel:

$$\begin{aligned}\|s_{K,\ell}\|_2^2 &\leqslant \mathbb{E}(\ell(Y_1)^2)\int_{\mathbb{R}^d}f(x')\int_{\mathbb{R}^d}K(x',x)^2\lambda_d(dx)\lambda_d(dx')\\ &\leqslant \mathbb{E}(\ell(Y_1)^2)\mathfrak{m}_{\mathcal{K},\ell}n \qquad (16)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}(V_{K,\ell}(n)) &= \mathbb{E}(\|K(X_1,.)\ell(Y_1)-s_{K,\ell}\|_2^2)\\ &= \mathbb{E}(\|K(X_1,.)\ell(Y_1)\|_2^2)+\|s_{K,\ell}\|_2^2-2\int_{\mathbb{R}^d}s_{K,\ell}(x)\mathbb{E}(K(X_1,x)\ell(Y_1))\lambda_d(dx)=\overline{s}_{K,\ell}-\|s_{K,\ell}\|_2^2. \qquad (17)\end{aligned}$$

Consider $\mathfrak{m}(n):=2\log(n)/\alpha$ and

$$v_{K,\ell}(n):=V_{K,\ell}(n)-\mathbb{E}(V_{K,\ell}(n))=v^1_{K,\ell}(n)+v^2_{K,\ell}(n),$$

where

$$v^j_{K,\ell}(n)=\frac{1}{n}\sum_{i=1}^n(g^j_{K,\ell}(n;X_i,Y_i)-\mathbb{E}(g^j_{K,\ell}(n;X_i,Y_i)))\ ;\ j=1,2$$

with, for every $(x',y)\in E$,

$$g^1_{K,\ell}(n;x',y):=\|K(x',.)\ell(y)-s_{K,\ell}\|_2^2\mathbf{1}_{|\ell(y)|\leqslant\mathfrak{m}(n)}$$

and

$$g^2_{K,\ell}(n;x',y):=\|K(x',.)\ell(y)-s_{K,\ell}\|_2^2\mathbf{1}_{|\ell(y)|>\mathfrak{m}(n)}.$$

On the one hand, by Bernstein's inequality, for any $\lambda>0$, with probability larger than $1-2e^{-\lambda}$,

$$|v^1_{K,\ell}(n)|\leqslant\sqrt{\frac{2\lambda}{n}\mathfrak{v}_{K,\ell}(n)}+\frac{\lambda}{n}\mathfrak{c}_{K,\ell}(n)$$

where

$$\mathfrak{c}_{K,\ell}(n)=\frac{\|g^1_{K,\ell}(n;.)\|_\infty}{3}\text{ and }\mathfrak{v}_{K,\ell}(n)=\mathbb{E}(g^1_{K,\ell}(n;X_1,Y_1)^2).$$

Moreover,

$$\begin{aligned}\mathfrak{c}_{K,\ell}(n) &= \frac{1}{3}\sup_{(x',y)\in E}\|K(x',.)\ell(y)-s_{K,\ell}\|_2^2\mathbf{1}_{|\ell(y)|\leqslant\mathfrak{m}(n)}\\ &\leqslant \frac{2}{3}\left(\mathfrak{m}(n)^2\sup_{x'\in\mathbb{R}^d}\|K(x',.)\|_2^2+\|s_{K,\ell}\|_2^2\right)\leqslant\frac{2}{3}(\mathfrak{m}(n)^2+\mathbb{E}(\ell(Y_1)^2))\mathfrak{m}_{\mathcal{K},\ell}n\end{aligned}$$

by Inequality (16), and

$$\begin{aligned}\mathfrak{v}_{K,\ell}(n) &\leqslant \|g^1_{K,\ell}(n;.)\|_\infty \mathbb{E}(V_{K,\ell}(n)) \\ &\leqslant 2(\mathfrak{m}(n)^2 + \mathbb{E}(\ell(Y_1)^2))\mathfrak{m}_{\mathcal{K},\ell} n(\overline{s}_{K,\ell} - \|s_{K,\ell}\|_2^2)\end{aligned}$$

by Inequality (16) and Equality (17). Then, for any $\theta \in]0,1[$,

$$\begin{aligned}|v^1_{K,\ell}(n)| &\leqslant 2\sqrt{\lambda(\mathfrak{m}(n)^2 + \mathbb{E}(\ell(Y_1)^2))\mathfrak{m}_{\mathcal{K},\ell}(\overline{s}_{K,\ell} - \|s_{K,\ell}\|_2^2)} + \frac{2\lambda}{3}(\mathfrak{m}(n)^2 + \mathbb{E}(\ell(Y_1)^2))\mathfrak{m}_{\mathcal{K},\ell} \\ &\leqslant \theta \overline{s}_{K,\ell} + \frac{5\lambda}{3\theta}(1 + \mathbb{E}(\ell(Y_1)^2))\mathfrak{m}_{\mathcal{K},\ell}\mathfrak{m}(n)^2\end{aligned}$$

with probability larger than $1 - 2e^{-\lambda}$. So, with probability larger than $1 - 2|\mathcal{K}_n|e^{-\lambda}$,

$$S_{\mathcal{K},\ell}(n,\theta) := \sup_{K\in\mathcal{K}_n}\left\{\frac{|v^1_{K,\ell}(n)|}{n} - \frac{\theta}{n}\overline{s}_{K,\ell}\right\} \leqslant \frac{5\lambda}{3\theta n}(1 + \mathbb{E}(\ell(Y_1)^2))\mathfrak{m}_{\mathcal{K},\ell}\mathfrak{m}(n)^2.$$

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

$$\lambda_{\mathcal{K},\ell}(n,\theta,t) := \frac{t}{\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)} \text{ with } \mathfrak{m}_{\mathcal{K},\ell}(n,\theta) = \frac{5}{3\theta n}(1 + \mathbb{E}(\ell(Y_1)^2))\mathfrak{m}_{\mathcal{K},\ell}\mathfrak{m}(n)^2.$$

Then, for any $T > 0$,

$$\begin{aligned}\mathbb{E}(S_{\mathcal{K},\ell}(n,\theta)) &\leqslant T + \int_T^\infty \mathbb{P}(S_{\mathcal{K},\ell}(n,\theta) \geqslant \lambda_{\mathcal{K},\ell}(n,\theta,t)\mathfrak{m}_{\mathcal{K},\ell}(n,\theta))dt \\ &\leqslant T + 2|\mathcal{K}_n|\int_T^\infty \exp(-\lambda_{\mathcal{K},\ell}(n,\theta,t))dt \\ &= T + 2|\mathcal{K}_n|\int_T^\infty \exp\left(-\frac{t}{2\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)}\right)\exp\left(-\frac{t}{2\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)}\right)dt \\ &\leqslant T + 2\mathfrak{c}_1|\mathcal{K}_n|\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)\exp\left(-\frac{T}{2\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)}\right) \text{ with } \mathfrak{c}_1 = \int_0^\infty e^{-r/2}dr = 2. \qquad (18)\end{aligned}$$

Moreover,

$$\mathfrak{m}_{\mathcal{K},\ell}(n,\theta) \leqslant \mathfrak{c}_2\frac{\log(n)^2}{\theta n} \text{ with } \mathfrak{c}_2 = \frac{20}{3\alpha^2}(1 + \mathbb{E}(\ell(Y_1)^2))\mathfrak{m}_{\mathcal{K},\ell}.$$

So, by taking

$$T = 2\mathfrak{c}_2\frac{\log(n)^3}{\theta n},$$

and since $|\mathcal{K}_n| \leqslant n$,

$$\mathbb{E}(S_{\mathcal{K},\ell}(n,\theta)) \leqslant 2\mathfrak{c}_2\frac{\log(n)^3}{\theta n} + 4\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)\frac{|\mathcal{K}_n|}{n} \leqslant 6\mathfrak{c}_2\frac{\log(n)^3}{\theta n}.$$

On the other hand, by Inequality (16) and Markov's inequality,

$$\begin{aligned}\mathbb{E}\left[\sup_{K\in\mathcal{K}_n}\frac{|v^2_{K,\ell}(n)|}{n}\right] &\leqslant \frac{2}{n}\mathbb{E}\left(\sup_{K\in\mathcal{K}_n}\|K(X_1,.)\ell(Y_1) - s_{K,\ell}\|_2^2\mathbf{1}_{|\ell(Y_1)|>\mathfrak{m}(n)}\right) \\ &\leqslant \frac{4}{n}\mathbb{E}\left[\left|\ell(Y_1)^2\sup_{K\in\mathcal{K}_n}\|K(X_1,.)\|_2^2 + \sup_{K\in\mathcal{K}_n}\|s_{K,\ell}\|_2^2\right|^2\right]^{1/2}\mathbb{P}(|\ell(Y_1)| > \mathfrak{m}(n))^{1/2} \leqslant \frac{\mathfrak{c}_3}{n}\end{aligned}$$

with

$$\mathfrak{c}_3 = 8\mathfrak{m}_{\mathcal{K},\ell}\mathbb{E}(\ell(Y_1)^4)^{1/2}\mathbb{E}(\exp(\alpha|\ell(Y_1)|))^{1/2}.$$

Therefore,

$$\mathbb{E}\left(\sup_{K\in\mathcal{K}_n}\left\{\frac{|v_{K,\ell}(n)|}{n} - \frac{\theta}{n}\overline{s}_{K,\ell})\right\}\right) \leqslant 6\mathfrak{c}_2\frac{\log(n)^3}{\theta n} + \frac{\mathfrak{c}_3}{n}$$

and, by Equality (17), the definition of $v_{K,\ell}(n)$ and Assumption 2.1.(2),

$$\mathbb{E}\left(\sup_{K\in\mathcal{K}_n}\left\{\frac{1}{n}|V_{K,\ell}(n)-\overline{s}_{K,\ell}|-\frac{\theta}{n}\overline{s}_{K,\ell}\right\}\right)\leqslant 6\mathfrak{c}_2\frac{\log(n)^3}{\theta n}+\frac{\mathfrak{c}_3+\mathfrak{m}_{\mathcal{K},\ell}}{n}.$$

Remark B.5. *As mentioned in Remark 2.10, replacing the exponential moment condition by the weaker q-th moment condition with $q=(12-4\varepsilon)/\beta$, $\varepsilon\in]0,1[$ and $0<\beta<\varepsilon/2$, allows to get a rate of convergence of order $1/n^{1-\varepsilon}$. Indeed, by Inequality (18), with $\mathfrak{m}(n)=n^\beta$ and*

$$T=\frac{2\mathfrak{c}_1}{\theta n^{1-\varepsilon}}\ \textit{with}\ \mathfrak{c}_1=\frac{5}{3}(1+\mathbb{E}(\ell(Y_1)^2))\mathfrak{m}_{\mathcal{K},\ell},$$

and by letting $\alpha=1+2\beta-\varepsilon$, there exist $n_{\varepsilon,\alpha}\in\mathbb{N}^*$ and $\mathfrak{c}_{\varepsilon,\alpha}>0$ not depending on n, such that for any $n\geqslant n_{\varepsilon,\alpha}$,

$$\begin{aligned}\mathbb{E}(S_{\mathcal{K},\ell}(n,\theta))&\leqslant\frac{2\mathfrak{c}_1}{\theta n^{1-\varepsilon}}+4\mathfrak{c}_1|\mathcal{K}_n|\frac{n^{2\beta-1}}{\theta}\exp(-n^{\varepsilon-2\beta})\\&\leqslant\frac{2\mathfrak{c}_1}{\theta n^{1-\varepsilon}}+4\mathfrak{c}_1\mathfrak{c}_{\varepsilon,\alpha}\frac{n^{2\beta}}{\theta n^\alpha}=\frac{2\mathfrak{c}_1(1+2\mathfrak{c}_{\varepsilon,\alpha})}{\theta n^{1-\varepsilon}}.\end{aligned}$$

Furthermore, by Markov's inequality,

$$\mathbb{P}(|\ell(Y_1)|>n^\beta)\leqslant\frac{\mathbb{E}(|\ell(Y_1)|^{(12-4\varepsilon)/\beta})}{n^{12-4\varepsilon}}.$$

So, as previously, there exists a deterministic constant $\mathfrak{c}_2>0$ such that

$$\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}|W^2_{K,K',\ell}(n)|\right)\leqslant\mathfrak{c}_2|\mathcal{K}_n|^2\mathbb{P}(|\ell(Y_1)|>\mathfrak{m}(n))^{1/4}\leqslant\frac{\mathfrak{c}_3\mathbb{E}(|\ell(Y_1)|^{(12-4\varepsilon)/\beta})^{1/4}}{n^{1-\varepsilon}},$$

and then

$$\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}\{|W_{K,K',\ell}(n)|-\theta\|s_{K',\ell}-s\|_2^2\}\right)\leqslant\frac{\mathfrak{c}_3}{\theta n^{1-\varepsilon}}\ \textit{with}\ \mathfrak{c}_3=2\mathfrak{c}_1(1+2\mathfrak{c}_{\varepsilon,\alpha})+\mathfrak{c}_2\mathbb{E}(|\ell(Y_1)|^{(12-4\varepsilon)/\beta})^{1/4}.$$

B.1.3. *Proof of Lemma B.3.* Consider $\mathfrak{m}(n)=12\log(n)/\alpha$. For any $K,K'\in\mathcal{K}_n$,

$$W_{K,K',\ell}(n)=W^1_{K,K',\ell}(n)+W^2_{K,K',\ell}(n)$$

where

$$W^j_{K,K',\ell}(n):=\frac{1}{n}\sum_{i=1}^n(g^j_{K,K',\ell}(n;X_i,Y_i)-\mathbb{E}(g^j_{K,K',\ell}(n;X_i,Y_i)))\ ;\ j=1,2$$

with, for every $(x',y)\in E$,

$$g^1_{K,K',\ell}(n;x',y):=\langle K(x',.)\ell(y),s_{K',\ell}-s\rangle_2\mathbf{1}_{|\ell(y)|\leqslant\mathfrak{m}(n)}$$

and

$$g^2_{K,K',\ell}(n;x',y):=\langle K(x',.)\ell(y),s_{K',\ell}-s\rangle_2\mathbf{1}_{|\ell(y)|>\mathfrak{m}(n)}.$$

On the one hand, by Bernstein's inequality, for any $\lambda>0$, with probability larger than $1-2e^{-\lambda}$,

$$|W^1_{K,K',\ell}(n)|\leqslant\sqrt{\frac{2\lambda}{n}\mathfrak{v}_{K,K',\ell}(n)}+\frac{\lambda}{n}\mathfrak{c}_{K,K',\ell}(n)$$

where

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

Moreover,

$$\begin{aligned}\mathfrak{c}_{K,K',\ell}(n)&=\frac{1}{3}\sup_{(x',y)\in E}|\langle K(x',.)\ell(y),s_{K',\ell}-s\rangle_2|\mathbf{1}_{|\ell(y)|\leqslant\mathfrak{m}(n)}\\&\leqslant\frac{1}{3}\mathfrak{m}(n)\|s_{K',\ell}-s\|_2\sup_{x'\in\mathbb{R}^d}\|K(x',.)\|_2\leqslant\frac{1}{3}\mathfrak{m}_{\mathcal{K},\ell}^{1/2}n^{1/2}\mathfrak{m}(n)\|s_{K',\ell}-s\|_2\end{aligned}$$

by Assumption 2.1.(1), and

$$\mathfrak{v}_{K,\ell}(n) \leqslant \mathbb{E}(\langle K(X_1,.)\ell(Y_1), s_{K',\ell} - s\rangle_2^2 \mathbf{1}_{|\ell(Y_1)|\leqslant \mathfrak{m}(n)}) \leqslant \mathfrak{m}(n)^2 \mathfrak{m}_{\mathcal{K},\ell} \|s_{K',\ell} - s\|_2^2$$

by Assumption 2.1.(4). Then, since $\lambda > 0$, for any $\theta \in]0,1[$,

$$\begin{aligned} |W^1_{K,K',\ell}(n)| &\leqslant \sqrt{\frac{2\lambda}{n}\mathfrak{m}(n)^2\mathfrak{m}_{\mathcal{K},\ell}\|s_{K',\ell} - s\|_2^2} + \frac{\lambda}{3n^{1/2}}\mathfrak{m}_{\mathcal{K},\ell}^{1/2}\mathfrak{m}(n)\|s_{K',\ell} - s\|_2 \\ &\leqslant \theta\|s_{K',\ell} - s\|_2^2 + \frac{\mathfrak{m}_{\mathcal{K},\ell}}{2\theta n}\mathfrak{m}(n)^2(1+\lambda)^2 \end{aligned}$$

with probability larger than $1 - 2e^{-\lambda}$. So, with probability larger than $1 - 2|\mathcal{K}_n|^2 e^{-\lambda}$,

$$S_{\mathcal{K},\ell}(n,\theta) := \sup_{K,K'\in\mathcal{K}_n} \{|W^1_{K,K',\ell}(n)| - \theta\|s_{K',\ell} - s\|_2^2\} \leqslant \frac{\mathfrak{m}_{\mathcal{K},\ell}}{2\theta n}\mathfrak{m}(n)^2(1+\lambda)^2.$$

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

$$\lambda_{\mathcal{K},\ell}(n,\theta,t) := -1 + \left(\frac{t}{\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)}\right)^{1/2} \text{ with } \mathfrak{m}_{\mathcal{K},\ell}(n,\theta) = \frac{\mathfrak{m}_{\mathcal{K},\ell}}{2\theta n}\mathfrak{m}(n)^2.$$

Then, for any $T > 0$,

$$\begin{aligned} \mathbb{E}(S_{\mathcal{K},\ell}(n,\theta)) &\leqslant T + \int_T^\infty \mathbb{P}(S_{\mathcal{K},\ell}(n,\theta) \geqslant (1+\lambda_{\mathcal{K},\ell}(n,\theta,t))^2 \mathfrak{m}_{\mathcal{K},\ell}(n,\theta))dt \\ &\leqslant T + 2|\mathcal{K}_n|^2 \int_T^\infty \exp(-\lambda_{\mathcal{K},\ell}(n,\theta,t))dt \\ &= T + 2|\mathcal{K}_n|^2 \int_T^\infty \exp\left(-\frac{t^{1/2}}{2\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)^{1/2}}\right)\exp\left(1 - \frac{t^{1/2}}{2\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)^{1/2}}\right)dt \\ &\leqslant T + 2\mathfrak{c}_1|\mathcal{K}_n|^2\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)\exp\left(-\frac{T^{1/2}}{2\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)^{1/2}}\right) \text{ with } \mathfrak{c}_1 = \int_0^\infty e^{1-r^{1/2}/2}dr. \end{aligned}$$

Moreover,

$$\mathfrak{m}_{\mathcal{K},\ell}(n,\theta) \leqslant \mathfrak{c}_2\frac{\log(n)^2}{\theta n} \text{ with } \mathfrak{c}_2 = \frac{12^2}{2\alpha^2}\mathfrak{m}_{\mathcal{K},\ell}.$$

So, by taking

$$T = 2^3\mathfrak{c}_2\frac{\log(n)^4}{\theta n},$$

and since $|\mathcal{K}_n| \leqslant n$,

$$\mathbb{E}(S_{\mathcal{K},\ell}(n,\theta)) \leqslant 2^3\mathfrak{c}_2\frac{\log(n)^4}{\theta n} + 2\mathfrak{c}_1\mathfrak{m}_{\mathcal{K},\ell}(n,\theta)\frac{|\mathcal{K}_n|^2}{n^2} \leqslant (2^3 + 2\mathfrak{c}_1)\mathfrak{c}_2\frac{\log(n)^4}{\theta n}.$$

On the other hand, by Assumption 2.1.(2,4), Cauchy-Schwarz's inequality and Markov's inequality,

$$\begin{aligned} \mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}|W^2_{K,K',\ell}(n)|\right) &\leqslant 2\mathbb{E}(\ell(Y_1)^2\mathbf{1}_{|\ell(Y_1)|>\mathfrak{m}(n)})^{1/2}\sum_{K,K'\in\mathcal{K}_n}\mathbb{E}(\langle K(X_1,.), s_{K',\ell} - s\rangle_2^2)^{1/2} \\ &\leqslant 2\mathfrak{m}_{\mathcal{K},\ell}^{1/2}\|s_{K',\ell} - s\|_2\mathbb{E}(\ell(Y_1)^4)^{1/4}|\mathcal{K}_n|^2\mathbb{P}(|\ell(Y_1)| > \mathfrak{m}(n))^{1/4} \leqslant \frac{\mathfrak{c}_3}{n} \end{aligned}$$

with

$$\mathfrak{c}_3 = 2\mathfrak{m}_{\mathcal{K},\ell}^{1/2}(\mathfrak{m}_{\mathcal{K},\ell}^{1/2} + \|s\|_2)\mathbb{E}(\ell(Y_1)^4)^{1/4}\mathbb{E}(\exp(\alpha|\ell(Y_1)|))^{1/4}.$$

Therefore,

$$\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}\{|W_{K,K',\ell}(n)| - \theta\|s_{K',\ell} - s\|_2^2\}\right) \leqslant (2^3 + 2\mathfrak{c}_1)\mathfrak{c}_2\frac{\log(n)^4}{\theta n} + \frac{\mathfrak{c}_3}{n} \leqslant \mathfrak{c}_4\frac{\log(n)^4}{\theta n}$$

with $\mathfrak{c}_4 = (2^3 + 2\mathfrak{c}_1)\mathfrak{c}_2 + \mathfrak{c}_3$.

B.2. **Proof of Proposition 2.7.** For any $K \in \mathcal{K}_n$,

$$\|\widehat{s}_{K,\ell}(n;.) - s_{K,\ell}\|_2^2 = \frac{U_{K,\ell}(n)}{n^2} + \frac{V_{K,\ell}(n)}{n} \tag{19}$$

with $U_{K,\ell}(n) = U_{K,K,\ell}(n)$ and $V_{K,\ell}(n) = V_{K,K,\ell}(n)$. Then, by Lemmas B.1 and B.2,

$$\mathbb{E}\left(\sup_{K\in\mathcal{K}_n}\left\{\left|\|\widehat{s}_{K,\ell}(n;.) - s_{K,\ell}\|_2^2 - \frac{\overline{s}_{K,\ell}}{n}\right| - \frac{\theta}{n}\overline{s}_{K,\ell}\right\}\right) \leqslant \mathfrak{c}_{2.7}\frac{\log(n)^5}{\theta n}$$

with $\mathfrak{c}_{2.7} = \mathfrak{c}_{B.1} + \mathfrak{c}_{B.2}$.

B.3. **Proof of Theorem 2.8.** On the one hand, for every $K \in \mathcal{K}_n$,

$$\|\widehat{s}_{K,\ell}(n;.) - s\|_2^2 - (1+\theta)\left(\|s_{K,\ell} - s\|_2^2 + \frac{\overline{s}_{K,\ell}}{n}\right)$$

can be written

$$\|\widehat{s}_{K,\ell}(n;.) - s_{K,\ell}\|_2^2 - (1+\theta)\frac{\overline{s}_{K,\ell}}{n} + 2W_{K,\ell}(n) - \theta\|s_{K,\ell} - s\|_2^2,$$

where $W_{K,\ell}(n) := W_{K,K,\ell}(n)$ (see (14)). Then, by Proposition 2.7 and Lemma B.3,

$$\mathbb{E}\left(\sup_{K\in\mathcal{K}_n}\left\{\|\widehat{s}_{K,\ell}(n;.) - s\|_2^2 - (1+\theta)\left(\|s_{K,\ell} - s\|_2^2 + \frac{\overline{s}_{K,\ell}}{n}\right)\right\}\right) \leqslant \mathfrak{c}_{2.8}\frac{\log(n)^5}{\theta n}$$

with $\mathfrak{c}_{2.8} = \mathfrak{c}_{2.7} + \mathfrak{c}_{B.3}$. On the other hand, for any $K \in \mathcal{K}_n$,

$$\|s_{K,\ell} - s\|_2^2 = \|\widehat{s}_{K,\ell}(n;.) - s\|_2^2 - \|\widehat{s}_{K,\ell}(n;.) - s_{K,\ell}\|_2^2 - W_{K,\ell}(n).$$

Then,

$$(1-\theta)\left(\|s_{K,\ell} - s\|_2^2 + \frac{\overline{s}_{K,\ell}}{n}\right) - \|\widehat{s}_{K,\ell}(n;.) - s\|_2^2 \leqslant |W_{K,\ell}(n)| - \theta\|s_{K,\ell} - s\|_2^2 + \Lambda_{K,\ell}(n) - \theta\frac{\overline{s}_{K,\ell}}{n}$$

where

$$\Lambda_{K,\ell}(n) := \left|\|\widehat{s}_{K,\ell} - s_{K,\ell}\|_2^2 - \frac{\overline{s}_{K,\ell}}{n}\right|.$$

By Equalities (19) and (17),

$$\Lambda_{K,\ell}(n) = \left|\frac{U_{K,\ell}(n)}{n^2} + \frac{v_{K,\ell}(n)}{n} - \frac{\|s_{K,\ell}\|_2^2}{n}\right|$$

with $U_{K,\ell}(n) = U_{K,K,\ell}(n)$ (see (13)). By Lemmas B.2 and B.1, there exists a deterministic constant $\mathfrak{c}_1 > 0$, not depending n and θ, such that

$$\mathbb{E}\left(\sup_{K\in\mathcal{K}_n}\left\{\Lambda_{K,\ell}(n) - \theta\frac{\overline{s}_{K,\ell}}{n}\right\}\right) \leqslant \mathfrak{c}_1\frac{\log(n)^5}{\theta n}.$$

By Lemma B.3,

$$\mathbb{E}\left(\sup_{K\in\mathcal{K}_n}\{|W_{K,\ell}(n)| - \theta\|s_{K,\ell} - s\|_2^2\}\right) \leqslant \mathfrak{c}_{B.3}\frac{\log(n)^4}{\theta n}.$$

Therefore,

$$\mathbb{E}\left(\sup_{K\in\mathcal{K}_n}\left\{\|s_{K,\ell} - s\|_2^2 + \frac{\overline{s}_{K,\ell}}{n} - \frac{1}{1-\theta}\|\widehat{s}_{K,\ell}(n;.) - s\|_2^2\right\}\right) \leqslant \overline{\mathfrak{c}}_{2.8}\frac{\log(n)^5}{\theta(1-\theta)n}$$

with $\overline{\mathfrak{c}}_{2.8} = \mathfrak{c}_{B.3} + \mathfrak{c}_1$.

B.4. **Proof of Theorem 3.2.** The proof of Theorem 3.2 is dissected in three steps.

Step 1. This first step is devoted to provide a suitable decomposition of

$$\|\widehat{s}_{\widehat{K},\ell}(n;\cdot)-s\|_2^2.$$

First,

$$\|\widehat{s}_{\widehat{K},\ell}(n;\cdot)-s\|_2^2=\|\widehat{s}_{\widehat{K},\ell}(n;\cdot)-\widehat{s}_{K_0,\ell}(n;\cdot)\|_2^2+\|\widehat{s}_{K_0,\ell}(n;\cdot)-s\|_2^2-2\langle\widehat{s}_{K_0,\ell}(n;\cdot)-\widehat{s}_{\widehat{K},\ell}(n;\cdot),\widehat{s}_{K_0,\ell}(n;\cdot)-s\rangle_2$$

From (8), it follows that for any $K\in\mathcal{K}_n$,

$$\begin{aligned}\|\widehat{s}_{\widehat{K},\ell}(n;\cdot)-s\|_2^2 &\leqslant \|\widehat{s}_{K,\ell}(n;\cdot)-s\|_2^2+\mathrm{pen}_\ell(K)-\mathrm{pen}_\ell(\widehat{K})+\|\widehat{s}_{K_0,\ell}(n;\cdot)-s\|_2^2\\ &\quad -2\langle\widehat{s}_{K,\ell}(n;\cdot)-\widehat{s}_{\widehat{K},\ell}(n\,\cdot),\widehat{s}_{K_0,\ell}(n;\cdot)-s\rangle_2\\ &= \|\widehat{s}_{K,\ell}(n;\cdot)-s\|_2^2+\psi_n(K)-\psi_n(\widehat{K}) \qquad (20)\end{aligned}$$

where

$$\psi_n(K):=2\langle\widehat{s}_{K,\ell}(n;\cdot)-s,\widehat{s}_{K_0,\ell}(n;\cdot)-s\rangle_2-\mathrm{pen}_\ell(K).$$

Let's complete the decomposition of $\|\widehat{s}_{\widehat{K},\ell}(n;\cdot)-s\|_2^2$ by writing

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

where

$$\begin{aligned}\psi_{1,n}(K) &:= \frac{U_{K,K_0,\ell}(n)}{n^2},\\ \psi_{2,n}(K) &:= -\frac{1}{n^2}\left(\sum_{i=1}^n\ell(Y_i)\langle K_0(X_i,.),s_{K,\ell}\rangle_2+\sum_{i=1}^n\ell(Y_i)\langle K(X_i,.),s_{K_0,\ell}\rangle_2\right)+\frac{1}{n}\langle s_{K_0,\ell},s_{K,\ell}\rangle_2 \text{ and}\\ \psi_{3,n}(K) &:= W_{K,K_0,\ell}(n)+W_{K_0,K,\ell}(n)+\langle s_{K,\ell}-s,s_{K_0,\ell}-s\rangle_2.\end{aligned}$$

Step 2. In this step, we give controls of the quantities

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

- By Lemma B.1, for any $\theta\in]0,1[$,

$$\mathbb{E}(|\psi_{1,n}(K)|)\leqslant\frac{\theta}{n}\overline{s}_{K,\ell}+\mathfrak{c}_{B.1}\frac{\log(n)^5}{\theta n}$$

and

$$\mathbb{E}(|\psi_{1,n}(\widehat{K})|)\leqslant\frac{\theta}{n}\mathbb{E}(\overline{s}_{\widehat{K},\ell})+\mathfrak{c}_{B.1}\frac{\log(n)^5}{\theta n}.$$

- On the one hand, for any $K,K'\in\mathcal{K}_n$, consider

$$\Psi_{2,n}(K,K'):=\frac{1}{n}\sum_{i=1}^n\ell(Y_i)\langle K(X_i,.),s_{K',\ell}\rangle_2.$$

Then, by Assumption 3.1,

$$\begin{aligned}\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}|\Psi_{2,n}(K,K')|\right) &\leqslant \mathbb{E}(\ell(Y_1)^2)^{1/2}\mathbb{E}\left(\sup_{K,K'\in\mathcal{K}_n}\langle K(X_1,.),s_{K',\ell}\rangle_2^2\right)^{1/2}\\ &\leqslant \overline{\mathfrak{m}}_{\mathcal{K},\ell}^{1/2}\mathbb{E}(\ell(Y_1)^2)^{1/2}.\end{aligned}$$

On the other hand, by Assumption 2.1.(2),

$$|\langle s_{K,\ell},s_{K_0,\ell}\rangle_2|\leqslant\mathfrak{m}_{\mathcal{K},\ell}.$$

Then, there exists a deterministic constant $\mathfrak{c}_1>0$, not depending on n and K, such that

$$\mathbb{E}(|\psi_{2,n}(K)|)\leqslant\frac{\mathfrak{c}_1}{n} \text{ and } \mathbb{E}(|\psi_{2,n}(\widehat{K})|)\leqslant\frac{\mathfrak{c}_1}{n}.$$

- By Lemma B.3,

$$\begin{aligned}\mathbb{E}(|\psi_{3,n}(K)|) &\leqslant \frac{\theta}{4}(\|s_{K,\ell}-s\|_2^2+\|s_{K_0,\ell}-s\|_2^2)+8\mathfrak{c}_{B.3}\frac{\log(n)^4}{\theta n}\\ &\quad+\left(\frac{\theta}{2}\right)^{1/2}\|s_{K,\ell}-s\|_2\times\left(\frac{2}{\theta}\right)^{1/2}\|s_{K_0,\ell}-s\|_2\\ &\leqslant \frac{\theta}{2}\|s_{K,\ell}-s\|_2^2+\left(\frac{\theta}{4}+\frac{1}{\theta}\right)\|s_{K_0,\ell}-s\|_2^2+8\mathfrak{c}_{B.3}\frac{\log(n)^4}{\theta n}\end{aligned}$$

and

$$\mathbb{E}(|\psi_{3,n}(\widehat{K})|)\leqslant\frac{\theta}{2}\mathbb{E}(\|s_{\widehat{K},\ell}-s\|_2^2)+\left(\frac{\theta}{4}+\frac{1}{\theta}\right)\|s_{K_0,\ell}-s\|_2^2+8\mathfrak{c}_{B.3}\frac{\log(n)^4}{\theta n}.$$

Step 3. By the previous step, there exists a deterministic constant $\mathfrak{c}_2>0$, not depending on n, θ, K and K_0, such that

$$\mathbb{E}(|\psi_n(K)|)\leqslant\theta\left(\|s_{K,\ell}-s\|_2^2+\frac{\overline{s}_{K,\ell}}{n}\right)+\left(\frac{\theta}{2}+\frac{2}{\theta}\right)\|s_{K_0,\ell}-s\|_2^2+\mathfrak{c}_2\frac{\log(n)^5}{\theta n}$$

and

$$\mathbb{E}(|\psi_n(\widehat{K})|)\leqslant\theta\mathbb{E}\left(\|s_{\widehat{K},\ell}-s\|_2^2+\frac{\overline{s}_{\widehat{K},\ell}}{n}\right)+\left(\frac{\theta}{2}+\frac{2}{\theta}\right)\|s_{K_0,\ell}-s\|_2^2+\mathfrak{c}_2\frac{\log(n)^5}{\theta n}.$$

Then, by Theorem 2.8,

$$\mathbb{E}(|\psi_n(K)|)\leqslant\frac{\theta}{1-\theta}\mathbb{E}(\|\widehat{s}_{K,\ell}(n;.)-s\|_2^2)+\left(\frac{\theta}{2}+\frac{2}{\theta}\right)\|s_{K_0,\ell}-s\|_2^2+\left(\frac{\mathfrak{c}_2}{\theta}+\frac{\mathfrak{c}_{2.8}}{1-\theta}\right)\frac{\log(n)^5}{n}$$

and

$$\mathbb{E}(|\psi_n(\widehat{K})|)\leqslant\frac{\theta}{1-\theta}\mathbb{E}(\|\widehat{s}_{\widehat{K},\ell}(n;.)-s\|_2^2)+\left(\frac{\theta}{2}+\frac{2}{\theta}\right)\|s_{K_0,\ell}-s\|_2^2+\left(\frac{\mathfrak{c}_2}{\theta}+\frac{\mathfrak{c}_{2.8}}{1-\theta}\right)\frac{\log(n)^5}{n}.$$

By decomposition (20), there exist two deterministic constants $\mathfrak{c}_3,\mathfrak{c}_4>0$, not depending on n, θ, K and K_0, such that

$$\begin{aligned}\mathbb{E}(\|\widehat{s}_{\widehat{K},\ell}(n;\cdot)-s\|_2^2)&\leqslant\mathbb{E}(\|\widehat{s}_{K,\ell}(n;\cdot)-s\|_2^2)+\mathbb{E}(|\psi_n(K)|)+\mathbb{E}(|\psi_n(\widehat{K})|)\\ &\leqslant\left(1+\frac{\theta}{1-\theta}\right)\mathbb{E}(\|\widehat{s}_{K,\ell}(n;\cdot)-s\|_2^2)+\frac{\theta}{1-\theta}\mathbb{E}(\|\widehat{s}_{\widehat{K},\ell}(n;.)-s\|_2^2)\\ &\quad+\frac{\mathfrak{c}_3}{\theta}\|s_{K_0,\ell}-s\|_2^2+\frac{\mathfrak{c}_4}{\theta(1-\theta)}\cdot\frac{\log(n)^5}{n}.\end{aligned}$$

This concludes the proof.

Acknowledgments. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°811017. The authors want also to thank Fabienne Comte for her careful reading and advices.

References

[1] G. Chagny. *Warped Bases for Conditional Density Estimation.* Math. Methods Statist. 22, 253-282, 2013.
[2] F. Comte. *Estimation non-paramétrique.* Spartacus IDH, 2014.
[3] F. Comte and N. Marie. *Bandwidth Selection for the Wolverton-Wagner Estimator.* Journal of Statistical Planning and Inference 207, 198-214, 2020.
[4] F. Comte and N. Marie. *On a Nadaraya-Watson Estimator with Two Bandwidths.* Submitted, 2020.
[5] F. Comte and T. Rebafka. *Nonparametric Weighted Estimators for Biased Data.* Journal of Statistical Planning and Inference 174, 104-128, 2016.
[6] R.A. DeVore and G.G. Lorentz. *Constructive Approximation.* Springer-Verlag, 1993.
[7] U. Einmahl and D.M. Mason. *An Empirical Process Approach to the Uniform Consistency of Kernel-Type Function Estimators.* Journal of Theoretical Probability 13, 1-37, 2000.
[8] U. Einmahl and D.M. Mason. *Uniform in Bandwidth Consistency of Kernel-Type Function Estimators.* Annals of Statistics 33, 1380-1403, 2005.

[9] E. Giné and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models.* Cambridge university press, 2015.
[10] 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.
[11] C. Houdré and P. Reynaud-Bouret. *Exponential Inequalities, with Constants, for U-statistics of Order Two.* Stochastic Inequalities and Applications, vol. 56 of Progr. Proba., 55-69, Birkhauser, 2003.
[12] C. Lacour, P. Massart and V. Rivoirard. *Estimator Selection: a New Method with Applications to Kernel Density Estimation.* Sankhya A 79, 2, 298-335, 2017.
[13] M. Lerasle, N.M. Magalhaes and P. Reynaud-Bouret. *Optimal Kernel Selection for Density Estimation.* High dimensional probabilities VII: The Cargese Volume, vol. 71 of Prog. Proba., 435–460, Birkhauser, 2016.
[14] P. Massart. *Concentration Inequalities and Model Selection.* Lecture Notes in Mathematics 1896, Springer, 2007.
[15] E.A. Nadaraya. *On a Regression Estimate.* (Russian) Verojatnost. i Primenen. 9, 157-159, 1964.
[16] E. Parzen. *On the Estimation of a Probability Density Function and the Mode.* The Annals of Mathematical Statistics 33, 1065-1076, 1962.
[17] M. Rosenblatt. *Remarks on some Nonparametric Estimates of a Density Function.* The Annals of Mathematical Statistics 27, 832-837, 1956.
[18] A. Tsybakov. *Introduction to Nonparametric Estimation.* Springer, 2009.
[19] S. Varet, C. Lacour, P. Massart and V. Rivoirard. *Numerical Performance of Penalized Comparison to Overfitting for Multivariate Density Estimation.* Preprint, 2020.
[20] G.S. Watson. *Smooth Regression Analysis.* Sankhya A 26, 359-372, 1964.

*LTCI, Télécom Paris, Palaiseau, France
Email address: helene.halconruy@telecom-paris.fr

†Laboratoire Modal'X, Université Paris Nanterre, Nanterre, France
Email address: nmarie@parisnanterre.fr

*,†ESME Sudria, Paris, France
Email address: helene.halconruy@esme.fr
Email address: nicolas.marie@esme.fr