Should we estimate a product of density functions by a product of estimators ?

F. Comte [*], C. Duval[†]

Abstract

In this paper, we consider the inverse problem of estimating the product fg of two densities, given a n-sample of i.i.d. observations drawn from each. We propose both projection estimators with model selection device and kernel estimators with bandwidth selection strategies. The procedures do not consist in making the product of each density estimator, but in plugging an overfitted estimator of one of the two densities, in an estimator based on the second sample. Our findings are a first step toward a better understanding of the good performances of overfitting in regression Nadaraya-Watson estimator. **March 9, 2022**

Keywords. Bandwidth selection. Density estimation. Kernel estimation. Model selection. Nonparametric estimation. Penalized Comparison to Overfitting. Projection estimator.
AMS Classification. 62G05-62G07

1 Introduction

In this work, we consider that we have n observations $X_i, i = 1, \dots, n$ independent and identically distributed (i.i.d.) with density f and independent from n additional observations $Y_i, i = 1, \dots, n$ i.i.d. with density g. We study the question of estimating the product function fg from these observations. Note that the resulting function is not a density, and none of the observations are directly related to this product. In that sense, we face an inverse problem. Our framework contains the case where $f = g$ and the goal is to estimate f^2 by splitting a $2n$-sample. These quantities may be of interest in some testing problems or as a first step for estimating the $\mathbb{L}^2$-norm of f, see Laurent and Massart (2000); other product problems are considered in Butucea *et al.* (2018).

However, we must explain that we considered this problem as a simplified setting (a toy-problem, in some sense) for a more complicated question. Let us explain it. Consider a regression model with $Y_i = b(X_i) + \varepsilon_i$ with i.i.d. and independent sequences $(X_i)_{1\le i\le n}$ and $(\varepsilon_i)_{1\le i\le n}$. The question is to estimate the regression function $b(\cdot)$ from observations $(X_i, Y_i)_{1\le i\le n}$. A popular proposal is the Nadaraya-Watson estimator (see Györfi *et al.* (2002))

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

where K is a kernel and h a bandwidth parameter. This estimator can be seen as a weighted combination of the Y_i's (second equality) or as a ratio of an estimator of bf, where f is still the density of the X_i's, divided by an estimator of f (first equality). In this last case, it is not clear that the same bandwidth h must be chosen for the numerator and the denominator. Surprisingly, Comte and Marie (2021) proposed sophisticated strategies for these two terms, but noticed in the simulation experiments that, if the numerical results obtained for both functions separately were excellent, the performance of the ratio was almost systematically defeated by the single bandwidth method selected from a least squares criterion relying on the weighted view of the question. The unique bandwidth selected in this case is small, but the ratio of these two bad overfitted estimators

[*]Université de Paris, MAP5, UMR CNRS 8145.
[†]Université de Lille, CNRS, UMR 8524 - Laboratoire Paul Painlevé, F-59000 Lille, France

is undoubtedly very good, at least for not too high noise level. This is why we wondered if the product of two functional estimators was a good estimator of the product of two functions; we took these functions as densities for simplicity. Thus our motivation is mainly theoretical, but we believe that the question is of general interest.

Now, let us see why making a product of density estimators can be easily seen as an inadequate (sub-optimal) strategy. Assume that we set $\widehat{fg} := \widehat{f} \times \widehat{g}$, where $\widehat{f}$ and $\widehat{g}$ are minimax optimal estimators of f and g respectively. To get an upper bound result, there is no other way than to separate the role of each estimate : both individual risks of $\widehat{f}$ and $\widehat{g}$ would emerge. Then, the resulting rate is the slowest between the rates of estimation of f and of g: it is the rate induced by the less regular density between f and g, say g without loss or generality for the remaining of this discussion. Clearly, this is not optimal if the product fg is more regular than g. For instance, for f a $\beta(p,p)$ density with $p \geq 2$, p integer and g a uniform density, i.e. a $\beta(1,1)$. Then on $\mathbb{R}$, f has regularity $p-1$ and g regularity 0, but $fg = f$ has regularity $p-1$. Therefore, one can wonder if in these cases it is possible to build an estimator directly adapted to the regularity of the product fg.

A related disadvantage of an upper bound separating the roles of f and g is that it does not treat this problem as an inverse problem : both individual regularities of f and g intervene whereas one expects that the sole regularity of fg should matter. Especially since, depending on the regularity classes which are considered, there is often no universal rule relating the regularities of f and g to the one of the product.

To complete this discussion, notice that it is easy to derive a lower bound result, inspired by the former example on beta distributions. Denote by $\Sigma(s, L)$ where s and L are positive, a ball of radius L in space of functions with regularity s. Then it holds, for any measurable function T of $(X_i, Y_i)_{1 \leq i \leq n}$,

$$\sup_{fg \in \Sigma(s,L)} \|T - fg\|^2 \geq \sup_{\substack{fg \in \Sigma(s,L) \\ \mathrm{Supp} f \subset [0,1] \\ g = \mathbf{1}_{[0,1]}}} \|T - fg\|^2 = \sup_{\substack{f \in \Sigma(s,L) \\ \mathrm{Supp} f \subset [0,1]}} \|T - f\|^2.$$

It follows that

$$\inf_T \sup_{fg \in \Sigma(s,L)} \|T - fg\|^2 \geq \inf_T \sup_{\substack{f \in \Sigma(s,L) \\ \mathrm{Supp} f \subset [0,1]}} \|T - f\|^2, \tag{1}$$

we recover on the right side the lower bound of the direct density estimation problem. To summarize, if the regularity set $\Sigma(s, L)$ contains a $[0, 1]$ supported density f_0, a lower bound for the product is given by a lower bound for the direct estimation of f_0. This is enough to state that the upper bound results presented below are optimal. For instance, we recover rates in $n^{-\frac{2s}{2s+1}}$ if $(fg) \in \Sigma(s, L)$, a Sobolev class of regularity s, that are minimax.

The plan of the paper is the following. We propose in section 2 a projection strategy: we define a projection estimator of the product fg and prove a non-asymptotic risk bound showing that a rate related to the regularity of fg can be reached for a well-chosen dimension of the projection space (see section 2.2). As this choice depends on unknown parameters, we then propose a model selection strategy and prove that the resulting estimator automatically reaches the squared-bias/variance compromise. Then, we turn to kernel strategies, for which we propose in section 3 an estimator with similar properties. The bandwidth selection procedure is more complicated. We study a Goldenshluger and Lepski (2011) method which gives, following a way rather similar to the projection case, a theoretical result but is difficult to use in practice. Then we propose a method inspired from the recent proposal of Lacour *et al.* (2017), which has a very intricate proof, but is quite easy to implement. Comparisons of the different methods and associated strategies for product estimators are conducted in section 4, and concur to our theoretical findings. Several additional questions are presented in the concluding remarks of section 5. Lastly, proofs are gathered in section 6 concerning section 2 and in section 7 for section 3.

2 Projection method

2.1 Estimator and first risk bound

Let $(\varphi_j)_{j\geq 0}$ a $\mathbb{L}^2(I)$-orthonormal basis, where $I \subseteq \mathbb{R}$ is a subset of $\mathbb{R}$ and the domain on which fg is estimated. We set $S_m = \text{vect}(\varphi_0, \ldots, \varphi_{m-1})$ the m-dimensional functional space linearly generated by the m first φ_j's. For any square integrable function h on I, we denote by $h_m = \sum_{j=0}^{m-1} a_j(h)\varphi_j$ with $a_j(h) = \langle \varphi_j, h \rangle$, the orthogonal projection of h on S_m. We also define the quantity, assumed to be finite:

$$L(m) := \sup_{x\in I} \sum_{j=0}^{m-1} \varphi_j^2(x).$$

The order of $L(m)$ depends on the choice of the basis. For the trigonometric basis for m odd and $I = [0, 1]$, it holds $L(m) = m$. For the Hermite basis where $I = \mathbb{R}$, we have $L(m) \leq C_H\sqrt{m}$ (see Lemma 1 in Comte and Lacour (2021) and section 2.2). For the Legendre polynomial basis where $I = [-1, 1]$ it holds that $L(m) = m^2$, see Cohen *et al.* (2013, p.831). In any case, we consider that $L(m) \geq 1$, which holds at least for $m \geq m_0$.

For simplicity, we write $fg = fg\mathbf{1}_I$ and we recall the definition of the projection estimator of f on S_{m^*} where m^* is a positive integer:

$$\widehat{f}_{m^*} = \sum_{j=0}^{m^*-1} \widehat{a}_j\varphi_j, \quad \widehat{a}_j = \frac{1}{n}\sum_{i=1}^{n} \varphi_j(X_i). \tag{2}$$

Now, we propose the following estimator of fg:

$$\widehat{(fg)}_{m,m^*} = \sum_{j=0}^{m-1} \widehat{a}_j^{(m^*)}\varphi_j, \quad \widehat{a}_j^{(m^*)} = \frac{1}{n}\sum_{i=1}^{n} \varphi_j(Y_i)\widehat{f}_{m^*}(Y_i). \tag{3}$$

Clearly, $\mathbb{E}(\widehat{a}_j^{(m^*)}) = \langle \varphi_j, f_{m^*}g \rangle$, which shows that our estimator is indeed close to $f_{m^*}g$, which in turn should be near of fg for large m^*. Choosing m^* large is possible only if the variance of $\widehat{f}_{m^*}$ does not appear in the risk bound. This is established for the integrated risk bound for the estimator (3) in the following result.

Proposition 2.1. *Assume that f and g are bounded on I with bounds denoted by $\|f\|_\infty$ and $\|g\|_\infty$ respectively. Let $\widehat{(fg)}_{m,m^*}$ be the estimator defined by (3). Then for any m^* such that $L(m^*) \leq n$, we have*

$$\mathbb{E}\left(\|\widehat{(fg)}_{m,m^*} - fg\|^2\right) \leq \|(fg)_m - fg\|^2 + \|g\|_\infty^2\|f - f_{m^*}\|^2 + \mathfrak{C}(f, g)\frac{L(m)}{n}, \tag{4}$$

where

$$\mathfrak{C}(f, g) := \|g\|_\infty(1 + \|f\|^2) + \|f\|_\infty\|g\|^2 \leq \|g\|_\infty(1 + 2\|f\|_\infty)$$

and $(fg)_m$ is the orthogonal projection of fg on S_m, f_{m^*} the orthogonal projection of f on S_{m^*}.

The risk bound (4) contains two standard terms, the squared bias

$$\mathbb{B}_1 = \|(fg)_m - fg\|^2 = \sum_{j\geq m} [a_j(fg)]^2, \quad a_j(fg) = \langle fg, \varphi_j \rangle$$

and the variance $\mathfrak{C}(f, g)L(m)/n$, requiring a standard compromise (when m grows, the bias decreases while the variance increases). It also involves the bias term $\|f - f_{m^*}\|^2$ which has no counterpart: thus m^* can and should be chosen as large as possible in order to make it negligible.

Strategy suggested by (4). If in the initial problem, f and g have symmetric roles, this is no longer true in the definition (3) of the estimator, where one of the two densities is estimated first.

As a matter of fact, Proposition 2.1 suggests: 1. to plug in the product estimator (3) an over-fitted estimator, eliminating a selection issue for m^*, 2. to select for this over-fitted estimator the one corresponding to the smoother density. Indeed, this should make the additional bias term decrease faster. However, the information about which is smoother between f and g, is not available. From theoretical viewpoint, both $\|f - f_{m^*}\|^2$ and $\|g - g_{m^*}\|^2$ are negligible by assuming a minimal regularity for f and g and choosing m^* maximal with $L(m^*) \leq n$. From practical point of view, we propose to consider that the smoother density is the one for which a model selection method for the direct density estimation of f and g leads to the smallest selected dimension (see section 4).

2.2 Rates on Sobolev Hermite spaces

In this section, we give an example of rate induced by Proposition 2.1, in the case of the Hermite basis and associated Sobolev spaces. The Hermite functions $(\varphi_j)_{j\geq 0}$ are defined from Hermite polynomials $(H_j)_{j\geq 0}$ by:

$$\varphi_j(x) = c_j H_j(x) e^{-x^2/2}, \quad H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j}(e^{-x^2}), \quad c_j = (2^j j! \sqrt{\pi})^{-1/2}, \quad x \in \mathbb{R}. \tag{5}$$

The Hermite polynomials $(H_j)_{j\geq 0}$ are orthogonal with respect to the weight function e^{-x^2}, that is: $\int_{\mathbb{R}} H_j(x) H_k(x) e^{-x^2} dx = 2^j j! \sqrt{\pi} \delta_{j,k}$ (see Abramowitz and Stegun (1964), chap 22.2.14). Therefore, the Hermite basis $(\varphi_j)_{j\geq 0}$ is an orthonormal basis on $\mathbb{R}$. We note also that φ_j is bounded:

$$||\varphi_j||_\infty = \sup_{x\in\mathbb{R}} |\varphi_j(x)| \leq \Phi_0, \text{ with } \Phi_0 \simeq 1,086435/\pi^{1/4} \simeq 0,8160 \tag{6}$$

(see Abramowitz and Stegun (1964), chap.22.14.17). Moreover, it is proved in Lemma 1 of Comte and Lacour (2021) that $\sup_{x\in\mathbb{R}} \sum_{j=0}^{m-1} \varphi_j^2(x) \leq C_H \sqrt{m}$ for a finite constant $C_H > 0$.

For $s > 0$, the Sobolev-Hermite ball (see Bongioanni and Torrea (2006)) is defined by :

$$W_H^s(D) = \Big\{\theta \in \mathbb{L}^2(\mathbb{R}), \sum_{k\geq 0} k^s a_k^2(\theta) \leq D\Big\}, \quad D > 0, \tag{7}$$

where $a_k(\theta) = \langle \theta, \varphi_k \rangle$. It is proved in Belomestny *et al.* (2019) that, for s an integer, $s \geq 1$, $f \in W_H^s = \{\theta \in \mathbb{L}^2(\mathbb{R}), \sum_{k\geq 0} k^s a_k^2(\theta) < +\infty\}$ is equivalent to : f admits derivatives up to order s which satisfy: f, f', …, $f^{(s)}$, $x^{s-\ell} f^{(\ell)}$ for $\ell = 0, \dots, s-1$ belong to $\mathbb{L}^2(\mathbb{R})$. Moreover, for any function $f \in W_H^s(D)$, we have $\|f - f_m\|^2 \leq D m^{-s}$. It is also easy to see that if, in addition, $s > 1$, then f is bounded. Indeed

$$|\sum_{j\geq 0} a_j \varphi_j| \leq \Phi_0 \left(|a_0| + \sum_{j\geq 1} (|a_j| j^{s/2}) j^{-s/2}\right) \leq \Phi_0 \left(\|f\| + \sqrt{\sum_{j\geq 1} j^s a_j^2 \sum_{j\geq 1} j^{-s}}\right).$$

As the functions are assumed to be bounded, it holds $|\langle fg, \varphi_j\rangle| \leq \min(\|f\|_\infty |\langle g, \varphi_j\rangle|, \|g\|_\infty |\langle f, \varphi_j\rangle|)$. Thus if $fg \in W_H^s(D)$, $f \in W_H^{s'}(D')$ and $g \in W_H^{s''}(D'')$, then $s \geq \max(s', s'')$.

Then we obtain as a straightforward consequence of Proposition 2.1, the following result.

Proposition 2.2. *Let $s \geq s' \geq 1/2$ and assume that $fg \in W_H^s(D)$, $f \in W_H^{s'}(D')$ with f and g bounded and $g \in \mathbb{L}^2(\mathbb{R})$. Then choosing $m_{\rm opt} = [n^{1/(s+1/2)}]$ and $m_n^* = n^2/C_H^2$, we have*

$$\mathbb{E}\left(\|\widehat{(fg)}_{m_{\rm opt}, m_n^*} - fg\|^2\right) \leq C(D, D', \|f\|_\infty, \|g\|_\infty) n^{-\frac{2s}{2s+1}}.$$

We can conclude that the resulting rate is of order $n^{-2s/(2s+1)}$, and is optimal, see (1).

2.3 Model selection

As noticed in the comments of Proposition 4, we can and should choose m^* as large as possible. Then only the dimension m remains to be selected from the data. We define the collection of proposals for m as follows

$$\mathcal{M}_n = \{m \in \{1, \dots, n\}, \quad L(m) \leq n\},$$

and set

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^{n} \widehat{f}_{m^*}(Y_i)t(Y_i). \tag{8}$$

Then we select $\widehat{m}$ with the criterion

$$\widehat{m} := \arg\min_{m\in\mathcal{M}_n}\left\{\min_{t\in S_m}\gamma_n(t) + \text{pen}(m)\right\}, \quad \text{pen}(m) = \kappa\|f\|_\infty(\|f\|_\infty + \|g\|_\infty)\frac{L(m)}{n}$$

where κ is a numerical constant. Note that

$$\min_{t\in S_m}\gamma_n(t) = \gamma_n(\widehat{(fg)}_{m,m^*}) = -\|\widehat{(fg)}_{m,m^*}\|^2.$$

We consider the following set of assumptions:

[A1] f and g are bounded on I.

[A2] The model m^* is such that $L(m^*) \leq \frac{\|f\|_\infty}{16}\frac{n}{\log(n)}$.

[A3] There exist two real numbers $a > 1$ and $C_a > 0$, which need not to be known, such that $\sum_{j\geq 1} j^a a_j^2(f) \leq C_a < +\infty$.

[A4] The basis functions are bounded: $\forall j \in \mathbb{N}, \forall x \in I$, $|\varphi_j(x)| \leq C_\varphi$, and the collection of models is nested.

[A5] The collection of models is such that $\text{Card}(\mathcal{M}_n) \leq n$, and $\forall c > 0$, $\sum_{m\in\mathcal{M}_n} e^{-c\sqrt{L(m)}} \leq \Sigma < +\infty$ where $\Sigma = \Sigma(c)$ is a constant depending on c but not on n.

We can prove the following result.

Theorem 2.1. *If Assumptions [A1]-[A5] hold, then, there exists κ_0 such that, for any $\kappa \geq \kappa_0$, we have*

$$\begin{aligned}\mathbb{E}(\|\widehat{(fg)}_{\widehat{m},m^*} - fg\|^2) \quad &\leq \quad \inf_{m\in\mathcal{M}_n}\left\{3\|fg - (fg)_m\|^2 + 4\kappa\|f\|_\infty(\|f\|_\infty + \|g\|_\infty)\frac{L(m)}{n}\right\} \\ &\qquad + 16\|g\|_\infty^2\|f - f_{m^*}\|^2 + \frac{C}{n},\end{aligned} \tag{9}$$

where C is a constant depending on $\|f\|_\infty$, C_a.

The proof is relegated to section 6 and indicates that $\kappa_0 = 8 \times 12 = 96$ would suit. In practice, the estimate is replaced by its positive part, for which the same risk bound holds. Theorem 2.1 shows that our adaptive procedure automatically realizes the squared bias-variance tradeoff up to negligible terms. As previously noticed if $m^* = m_n^*$ is chosen large enough $\|f - f_{m^*}\|^2$ is negligible (less than $1/n$).

In Assumption [A2], the maximal value of m^* depends on $\|f\|_\infty$. This constraint can be replaced by $L(m^*) \leq n/\log^{3/2}(n)$ and the result follows for n large enough. Condition [A3] implies that the function f has a minimal regularity of $1/2$ on Sobolev-Fourier spaces for $I = [0, 1]$ and 1 on Hermite Sobolev spaces. Assumptions [A4] and [A5] are classical, for instance they are fulfilled for the trigonometric and Hermite bases.

The values $\|f\|_\infty$, $\|g\|_\infty$ in the penalty term are unknown and must be replaced by estimates. The bound $\|f\|_\infty$ can be estimated by the maximal value of a projection estimate of f on a middle-sized space, for instance $\sup_{x\in I}|\widehat{f}_{[\sqrt{n}]}(x)|$ and an analogous approach can be adopted for $\|g\|_\infty$. Let us denote these estimators by $\widehat{\|f\|}_\infty$ and $\widehat{\|g\|}_\infty$. This strategy is theoretically studied in Theorem 12 p.594 (Appendix A: Random penalty) in Lacour (2007).
We adopt in the numerical Section, the following strategy. The penalty is obtained from the theory as the sum of the bounds of two terms, a bound on $\frac{1}{n}\sum_{j=1}^{m-1}\mathbb{E}\left(\varphi_j^2(Y_1)[\widehat{f}_{m^*}(Y_1)]^2\right)$ and a bound on an additional term $\|f\|_\infty\|g\|_\infty L(m)/n$. Following ideas in Massart (2007) (see also Theorem 7.6 p.216, in the density case), we replace the first term by

$$\widehat{\text{pen}}_1(m) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=0}^{m-1}\left[\widehat{f}_{m^*}(Y_i)\varphi_j(Y_i)\right]^2$$

and the second term by $\widehat{\text{pen}}_2(m) = \widehat{\|f\|}_\infty\widehat{\|g\|}_\infty L(m)/n$. So, in the Hermite basis where $L(m) = C_H\sqrt{m}$ (with unknown C_H), our global penalty is

$$\widehat{\text{pen}}_1(m) + \kappa\widehat{\|f\|}_\infty\widehat{\|g\|}_\infty\frac{\sqrt{m}}{n}. \tag{10}$$

The constant κ is calibrated by preliminary simulations, see section 4.

3 Kernel estimators

3.1 Definition and risk bound

Let K be a symmetric kernel. We recall the definition of the classical kernel density estimator of f

$$\widetilde{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n}K_h(X_i - x), \quad K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right), \quad h > 0. \tag{11}$$

Let us denote by $\star$ the convolution product, $u\star v(x) = \int u(t)v(x-t)dt$ for functions u, v such that the integral is well defined. Then, for any function w, we set $w_h := w\star K_h$.

By analogy with the projection study, we define the kernel estimator of fg by

$$\widetilde{(fg)}_{h,h_o}(x) = \frac{1}{n}\sum_{i=1}^{n}\widetilde{f}_{h_o}(Y_i)K_h(Y_i - x). \tag{12}$$

We prove the following integrated risk bound.

Proposition 3.1. *Assume that f is square integrable on $\mathbb{R}$ and that g is bounded on $\mathbb{R}$ with bound denoted by $\|g\|_\infty$. Let $\widetilde{(fg)}_{h,h_o}$ be the estimator defined by (12). Then for any h_o such that $1/(nh_o) \le 1$, we have*

$$\mathbb{E}\left(\|\widetilde{(fg)}_{h,h_o} - fg\|^2\right) \le 2\|(fg)_h - fg\|^2 + 2\|g\|_\infty^2\|K\|_1^2\|f_{h_o} - f\|^2 + \frac{C(f,g,K)}{nh}, \tag{13}$$

where $C(K,f) = \|K\|^2\|g\|_\infty(\|K\|^2 + 2\|K\|_1^2\|f\|^2)$.

As observed in the projection context, Proposition 3.1 suggests to choose h_o the smallest as possible, in order to make this term negligible. For instance if f belongs to a Nikols'ki ball with regularity parameter α (see Tsybakov (2009), Definition 1.4 p.13), $\|f_{h_o} - f\|^2$ has order $(h_o)^{2\alpha}$ if the kernel K has order at least $\lfloor\alpha\rfloor$. It follows that for $\alpha > 1/2$ and $h_o = 1/n$, this term has order less than $1/n$ and is negligible. Then, there is only one bandwidth h that requires to be selected.

If in addition fg belongs to a Nikols'ki ball with regularity parameter β and the kernel K has order at least $\lfloor\beta\rfloor$, (see Tsybakov (2009), Definition 1.3, p.5) then the estimator will reach the minimax rate $n^{-2\beta/(2\beta+1)}$ (Tsybakov (2009), chapter 1, section 1.2.3 and Theorem 1.2) for h chosen of order $n^{-1/(2\beta+1)}$. As in the projection case, such a choice of h is not feasible since β is unknown, a data driven procedure for selecting h must be proposed.

3.2 Bandwidth selection with Goldenschluger and Lepski method

Our first proposal is a Goldenschluger and Lespki (2011) method. We define for $\mathcal{H}_n$ a discrete collection of bandwidths in $(1/n, 1)$ with cardinality less than n,

$$A(h) = \sup_{h' \in \mathcal{H}_n} \left[\| \left(K_{h'} \star \widetilde{(fg)}_h \right) - \widetilde{(fg)}_{h'} \|^2 - \kappa V(h') \right]_+, \quad V(h) = \frac{\|K\|_1^4 \|K\|^2 (\|f\|_\infty^2 + \|g\|_\infty^2)}{nh} \tag{14}$$

where we drop the index h_o for readibility in $\widetilde{(fg)}_h = \widetilde{(fg)}_{h,h_o}$, and the selection of h is done by the rule

$$\check{h} = \arg\min_{h \in \mathcal{H}_n} \{A(h) + \kappa' V(h)\}.$$

Note that by denoting $K_{h,h'}(x) = K_{h',h}(x) = K_h \star K_{h'}(x)$,

$$K_{h'} \star \widetilde{(fg)}_h(x) = \frac{1}{n} \sum_{i=1}^n \widetilde{f}_{h_o}(Y_i) K_{h,h'}(Y_i - x) = K_h \star \widetilde{(fg)}_{h'}(x).$$

We consider the following set of assumptions:

[B1] f and g are bounded on $\mathbb{R}$.

[B2] The kernel K is even, bounded and integrable.

[B3](p) The bandwidth h_o is such that $nh_o \geq 3p[\|K\|_\infty/(2\|f\|_\infty \|K\|_1)] \log(n)$.

[B4] The discrete collection of bandwidths in $(1/n, 1)$, $\mathcal{H}_n$, has cardinality less than n and such that for any $c_1 > 0$, $\sum_{h \in \mathcal{H}_n} \exp(-c_1/h) \leq \Sigma = \Sigma(c_1) < +\infty$.

Note that as $\int K = 1$, $\|K\|_1 \geq 1$. As for [A2], we can replace [B3](p) with $nh_o \geq \log(n)^{3/2}$, for large enough n, to get rid of the unknown constant $\|f\|_\infty$ in the bound defining h_o. Assumption [B4] is fulfilled for $\mathcal{H}_n = \{h_k = 1/k, k = 1, \dots, n\}$. Note that contrary to the projection, the kernel method does not require any regularity constraint of type [A5].

Theorem 3.1. *Under Assumptions* [B1]-[B2]-[B3]*(3) and* [B4]*, we have for $\kappa' \geq \kappa$, that*

$$\mathbb{E}\left(\|\widetilde{(fg)}_{\check{h}} - fg\|^2 \right) \leq C \inf_{h \in \mathcal{H}_n} \left\{ \|K\|_1^4 \|(fg)_h - fg\|^2 + \kappa' V(h) \right\} + 18 \|K\|_1^4 \|g\|_\infty^2 \|f_{h_o} - f\|^2 + \frac{C'}{n},$$

where C is numerical and C' depends on $\|f\|_\infty, \|g\|_\infty, K$.

Theorem 3.1 ensures that the adaptive estimator performs a squared bias-variance compromise. However, the Goldenschluger and Lespki (2011) method is often difficult to calibrate from an implementation viewpoint (see Comte and Rebafka (2012)) and suffers from important computational costs. Moreover, it involves the calibration of two constants, κ' and κ. Contrary to the model selection procedure this preliminary calibration step is difficult, probably because these constants act simultaneously on the bias and variance terms. Moreover, the "double" convolution $K_{h,h'}$ is numerically time consuming.

This is why we explore another PCO method, introduced for density estimation by Lacour *et al.* (2017). The PCO method is more complicated from theoretical point of view, because it involves the study of several U-statistics of order 2. But, it is much easier to calibrate and implement, from practical point of view. Still, the Goldenschluger and Lespki (2011) method has the advantage that its proof, though technical, is well delineated; it is enlightening to understand the order of the different terms involved in the decomposition of the key processes appearing in both methods.

3.3 Bandwidth selection with PCO

We keep omitting the index h_o and write $\widetilde{(fg)}_h = \widetilde{(fg)}_{h,h_o}$ as h_o is fixed equal to minimal value. We select

$$\widetilde{h} = \arg\min_{h\in\mathcal{H}_n} \left\{ \|\widetilde{(fg)}_h - \widetilde{(fg)}_{h_{\min}}\|^2 + 2\text{pen}(h) \right\}$$

with $h_{\min} = \min\{h, h \in \mathcal{H}_n\}$,

$$\text{pen}(h) = \text{pen}_1(h) + \text{pen}_2(h) \quad \text{where} \quad \text{pen}_1(h) = \frac{1}{n^2}\langle K_h, K_{h_{\min}}\rangle \sum_{i=1}^{n} \widetilde{f}^2_{h_0}(Y_i), \tag{15}$$

$$\text{pen}_2(h) = \kappa\frac{c_0(f,g,K)}{nh}, \quad c_0(f,g,K) = 4\|K\|_1^3\|K\|_\infty(\|g\|_\infty^2 + \|f\|_\infty^2). \tag{16}$$

Note that pen_1 and pen_2 are both of order $1/(nh)$. This is obvious for pen_2; for pen_1, observe that $|\langle K_h, K_{h_{\min}}\rangle| \leq \|K\|_\infty\|K\|_1/h$ and that, under [B3](p), $(1/n)\sum_{i=1}^n \widetilde{f}^2_{h_o}(Y_i)$ is bounded with large probability, see (35).

Theorem 3.2. *Assume that Assumptions* [B1]-[B2], [B3]*(4) and* [B4] *hold and that* $1/(nh_{\min}) \leq 1$. *Then, for any* $\theta \in (0, 1/4)$ *and* $\kappa \geq 1/4$, *we have*

$$\begin{aligned}\mathbb{E}\left(\|\widetilde{(fg)}_{\widetilde{h}} - fg\|^2\right) \quad &\leq \quad 2(1 + c_1(\theta)) \inf_{h\in\mathcal{H}_n} \left\{ \|(fg)_h - fg\|^2 + (1 + \frac{\kappa}{1+\theta})\frac{c_0(f,g,K)}{nh} \right\} \\ &\qquad + c_2(\theta)\|(fg)_{h_{\min}} - fg\|^2 + c_3\|f_{h_0} - f\|^2 + C\frac{\log(n)}{n},\end{aligned}$$

where

$$c_1(\theta) = 2\theta(1-\theta)/(1-3\theta) > 0, \quad c_2(\theta) = 2\frac{(1+\theta^2)(1-2\theta)}{\theta(1-2\theta)} > 0,$$

and c_3 *and* C *are positive constants depending on* θ, f, g, K.

The risk bound of Theorem 3.2 involves four terms. The first term in the first line is the minimal risk among the collection of estimators, up to multiplicative constants. The two following terms, $\|(fg)_{h_{\min}} - fg\|^2$ and $\|f_{h_0} - f\|^2$ are bias terms corresponding to small bandwidths, they are negligible if $h_{\min}$ is of order $1/n$ and h_o of order $\log(n)/n$ (as required by assumption [B3](4) and if the functions f and fg have regularity larger that 1/2. The last term $\log(n)/n$ has negligible order compared to the first one. Therefore, the adaptive estimator achieves the intended squared-bias variance compromise detailed at the end of section 3.1.

4 Examples and simulation experiments

4.1 Description of the procedures

In this section, we illustrate the performances of the projection estimator with Hermite basis (see section 2.2) and kernel estimator with kernel built as a Gaussian mixture defined by:

$$K(x) = 2n_1(x) - n_2(x), \tag{17}$$

where $n_j(x)$ is the density of a centered Gaussian with a variance equal to j. This kernel is of order 3 (*i.e.* $\int x^j K(x)dx = 0$, for $j = 1, \ldots, 3$). We consider four examples:

1. $X \sim f = \mathcal{B}(7,5)$ and $Y \sim g = \mathcal{U}(0,1)$,
2. $X \sim f = \Gamma(4, 1/4)$ and $Y \sim g = \mathcal{E}(1/4)$,
3. $X \sim f = \mathcal{N}(0,3)$ and $Y \sim g$ Laplace,
4. $X \sim f = \mathcal{N}(0,3)$ and $Y \sim g$ Cauchy.

We compute normalized $\mathbb{L}^2$-risks to allow the numerical comparaison of the different examples for which $\int (fg)^2$ varies a lot. Namely, we evaluate

$$\frac{\mathbb{E}[\|\widehat{(fg)} - fg\|^2]}{\|fg\|^2}$$

and the associated deviations, from $N = 100$ independent datasets with different values of sample size $n = 200$, 1000 and 2000. All adaptive methods require the calibration of constants κ's in penalties. This is done by preliminary simulation experiments. For calibration strategies (dimension jump and slope heuristics), the reader is referred to Baudry *et al.* (2012), and to Lerasle (2012) for theoretical justifications. Here, we test a grid of values of κ's, the tests are conducted on a set of densities which are different from the one considered hereafter, to avoid overfitting. The different estimators are computed on the same datasets and compared.

• **Product** : This estimator is obtained as the product of $\widehat{f}\widehat{g}$ where each estimator is an adaptive optimal estimator. In the projection case, the product is $\widehat{f}_{\widehat{m}_1}\widehat{g}_{\widehat{m}_2}$, where $\widehat{f}_m$ is defined by (2) with

$$\widehat{m}_1 = \arg\min_{m\in\{1,\ldots,D_{\max}\}} \left\{ -\|\widehat{f}_m\|^2 + \frac{4}{n^2}\sum_{i=1}^{n}\sum_{j=0}^{m-1}\varphi_j^2(X_i) \right\},$$

and $\widehat{g}_{\widehat{m}_2}$ is defined analogously. In the kernel case, $\widetilde{f}_{\widetilde{h}_1}\widetilde{g}_{\widetilde{h}_2}$, where $\widetilde{f}_{\widetilde{h}_1}$ is defined by (11) with

$$\widetilde{h}_1 = \arg\min_{h\in\{1/k,k=1,\ldots,n\}} \left\{ \|\widetilde{f}_h - \widetilde{f}_{\frac{1}{n}}\|^2 + \frac{4}{n}\langle K_h, K_{\frac{1}{n}}\rangle \right\},$$

and $\widetilde{g}_{\widetilde{h}_2}$ is defined analogously.

• **First X**: In all our examples f is smoother than g. The theoretical results suggest that one should consider for the preliminary estimate the dataset X which has density f. In the projection setting our estimate is $\widehat{(fg)}_{\widehat{m},m^*}$ of Theorem 2.1 and penalty given by (10), where $\widehat{\|f\|_\infty^2}$ is estimated by $\sup_{x\in I}|\widehat{f}_{10}^2(x)|$, $\widehat{\|g\|_\infty^2}$ is estimated similarly, and with $\kappa = 0.15$ after calibration. In the kernel case we consider the estimator $\widetilde{(fg)}_{\widetilde{h}}$ of Theorem 3.2 with penalty given by (15) where pen_2 is replaced by

$$\widehat{\mathrm{pen}}_2 = 0.32\frac{\widehat{\|f\|_\infty^2} + \widehat{\|g\|_\infty^2}}{nh},$$

where $\widehat{\|f\|_\infty^2}$ is estimated by $\sup_{x\in I}|\widetilde{f}^2_{\log n/\sqrt{n}}(x)|$, $\widehat{\|g\|_\infty^2}$ is estimated similarly. Note that $\|K\|_1 \simeq 1.133$ and $\|K\|_\infty \simeq 0.516$, $\|K\|_1^3\|K\|_\infty \simeq 0.75$.

• **Optimal first** : As the information about compared smoothness of f and g is unavailable in practice, we have proposed an adaptive method for choosing which estimate is plugged in: we perform a classical penalized (resp. PCO) procedure (see step **Product**) to the datasets X and Y and we take as preliminary projection (resp. kernel) estimate the one for which $\widehat{m}$ (resp. $\widetilde{h}$) is the smallest (resp. largest). Indeed, the optimal dimension (resp. bandwidth) is asymptotically a decreasing (resp. increasing) function of the regularity. For instance, if $\widehat{m}_1 < \widehat{m}_2$ we proceed as in **First X** step, otherwise the roles of X and Y are switched. We count the number of times where Y is selected first; thus, when this count is zero, **First X** and **Optimal first** are the same and give the same result.

• **Oracle (optimal first)** : Our benchmark is computed as follows. We consider for all dimensions or bandwidths the estimators of the step **Optimal first** and select the oracle that minimizes $m \mapsto \mathbb{E}[\|\widehat{(fg)}_{m,m^*} - fg\|^2]$ or $h \mapsto \mathbb{E}[\|\widetilde{(fg)}_h - fg\|^2]$. This quantity provides a numerical lower bound for the $\mathbb{L}^2$-risk of our procedure.

4.2 Numerical results

Let us comment the results of Tables 1-4. First, we compare separately projection and kernel procedures. Let us start with the two fully data driven methods **Product** and **Optimal first**. We

n	Product		First X		Optimal first		Oracle	
	Hermite	Kernel	Hermite	Kernel	Hermite	Kernel	Hermite	Kernel
200	$18.2_{(1.85)}$	$3.21_{(2.97)}$	$12.3_{(3.32)}$	$21.0_{(3.96)}$	$7.79^{(\mathbf{100})}_{(2.32)}$	$5.80^{(\mathbf{79})}_{(8.60)}$	$2.08_{(1.64)}$	$4.69_{(8.00)}$
1000	$4.41_{(0.57)}$	$1.22_{(1.05)}$	$3.79_{(1.09)}$	$1.05_{(0.68)}$	$2.19^{(\mathbf{100})}_{(0.66)}$	$0.93^{(\mathbf{26})}_{(0.6)}$	$0.54_{(0.40)}$	$0.74_{(0.54)}$
2000	$2.26_{(0.38)}$	$0.43_{(0.29)}$	$1.92_{(0.46)}$	$0.45_{(0.32)}$	$1.18^{(\mathbf{100})}_{(0.37)}$	$0.40^{(\mathbf{47})}_{(0.28)}$	$0.27_{(0.21)}$	$0.31_{(0.23)}$

Table 1: $\mathbb{L}^2$-risks with std in parenthesis (both multiplied by 10^2): $f \sim \mathcal{B}(7,5)$ and $g \sim \mathcal{U}(0,1)$, $D_{\max} = 130$. For the Optimal first the bold upper script is the number of times where Y is selected first.

n	Product		First X		Optimal first		Oracle	
	Hermite	Kernel	Hermite	Kernel	Hermite	Kernel	Hermite	Kernel
200	$27.4_{(8.13)}$	$16.6_{(9.11)}$	$10.5_{(5.07)}$	$5.21_{(3.36)}$	$10.5^{(\mathbf{0})}_{(5.07)}$	$5.21^{(\mathbf{0})}_{(3.36)}$	$4.27_{(2.87)}$	$3.79_{(2.18)}$
1000	$38.1_{(6.63)}$	$10.1_{(7.29)}$	$3.86_{(1.43)}$	$2.77_{(2.12)}$	$3.86^{(\mathbf{0})}_{(1.43)}$	$2.80^{(\mathbf{2})}_{(2.13)}$	$1.52_{(1.02)}$	$1.25_{(0.77)}$
2000	$36.2_{(4.98)}$	$11.0_{(10.8)}$	$2.62_{(0.80)}$	$1.75_{(1.57)}$	$2.62^{(\mathbf{0})}_{(0.80)}$	$1.75^{(\mathbf{0})}_{(1.57)}$	$0.90_{0.64)}$	$0.80_{(0.55)}$

Table 2: $\mathbb{L}^2$-risks with std in parenthesis (both multiplied by 10^2): $f \sim \Gamma(4,1/4)$ and $g \sim \mathcal{E}(1/4)$, $D_{\max} = 100$. For the Optimal first the bold upper script is the number of times where Y is selected first.

n	Product		First X		Optimal first		Oracle	
	Hermite	Kernel	Hermite	Kernel	Hermite	Kernel	Hermite	Kernel
200	$5.45_{(2.61)}$	$6.81_{(7.42)}$	$5.68_{(1.47)}$	$6.50_{(3.49)}$	$5.65^{(\mathbf{51})}_{(1.44)}$	$6.66^{(\mathbf{29})}_{(3.59)}$	$3.43_{(1.58)}$	$3.18_{(2.04)}$
1000	$4.62_{(0.93)}$	$4.36_{(6.96)}$	$2.25_{(0.93)}$	$2.94_{(2.39)}$	$2.53^{(\mathbf{100})}_{(0.95)}$	$2.80^{(\mathbf{42})}_{(2.32)}$	$1.16_{(0.60)}$	$0.98_{(0.57)}$
2000	$2.27_{(1.57)}$	$2.46_{(2.37)}$	$1.39_{(0.33)}$	$2.34_{(2.44)}$	$1.46^{(\mathbf{55})}_{(0.41)}$	$2.62^{(\mathbf{26})}_{(2.68)}$	$0.74_{(0.34)}$	$0.62_{(0.35)}$

Table 3: $\mathbb{L}^2$-risks with std in parenthesis (both multiplied by 10^2): $f \sim \mathcal{N}(0,3)$ and g Laplace, $D_{\max} = 100$. For the Optimal first the bold upper script is the number of times where Y is selected first.

n	Product		First X		Optimal first		Oracle	
	Hermite	Kernel	Hermite	Kernel	Hermite	Kernel	Hermite	Kernel
200	$4.60_{(2.59)}$	$49.8_{(71.8)}$	$3.24_{(2.14)}$	$5.19_{(2.48)}$	$3.24^{(\mathbf{48})}_{(2.14)}$	$5.18^{(\mathbf{23})}_{(2.48)}$	$2.43_{(1.67)}$	$2.65_{(1.86)}$
1000	$2.31_{(0.74)}$	$72.3_{(74.6)}$	$1.84_{(0.54)}$	$2.51_{(1.87)}$	$1.87^{(\mathbf{100})}_{(0.50)}$	$2.46^{(\mathbf{13})}_{(1.87)}$	$0.84_{(0.52)}$	$0.78_{(0.68)}$
2000	$0.92_{(0.74)}$	$64.1_{85.4)}$	$0.73_{(0.49)}$	$1.88_{(1.73)}$	$0.83^{(\mathbf{95})}_{(0.53)}$	$1.77^{(\mathbf{10})}_{(1.66)}$	$0.43_{(0.30)}$	$0.39_{(0.31)}$

Table 4: $\mathbb{L}^2$-risks with std in parenthesis (both multiplied by 10^2): $f \sim \mathcal{N}(0,3)$ and g Cauchy, $D_{\max} = 50$. For the Optimal first the bold upper script is the number of times where Y is selected first.

observe that the results of the corresponding columns nicely confirm the theory: the risks of our procedure is almost systematically and significantly smaller (see Table 2 in particular). Besides, the risk of **Optimal first** is always comparable and even sometimes better than the risk of the **First X** method which uses the unavailable knowledge of the smoothest density. The risk of **Optimal first** has the same order as the **Oracle** even if a multiplicative factor larger than 2 is observed. Lastly, as the risks are normalized we can compare the risks of the different Tables; we see that the first two examples (Tables 1 and 2) are slightly more difficult which was expected: these densities are less regular as functions on $\mathbb{R}$.

Second, we can compare projection and kernel methods. The kernel method is much more time consuming that the projection method (by a factor more than 10). We can see that for the operational **Optimal first** method the kernel strategy seems better for the first two examples while the projection method wins in the two other cases. Nevertheless, the gap between the risks is never very large.

5 Concluding remarks

In this paper, we have shown that an optimal strategy for estimating a product of densities was not to make a product of estimators but to plug an overfitted estimator of one of the densities in the estimator of the product. This can be done both with projection and kernel estimators and adequate model or bandwidth selection methods are proved to deliver adaptive estimators. We have implemented these methods and proved their good numerical performances.

We assumed that the two samples had the same sizes but the case where the X-sample has size n_X and the Y sample size n_Y is worth being studied, for instance if $n_X = \gamma n$, $n_Y = n$, $\gamma \in (0, \infty)$. Following the steps of the proof in the projection case suggests that the procedure can be adapted and leads to similar results with rate induced by the smallest sample size $(1 \wedge \gamma)n$.

We considered a product of two densities but generalizations to product of other functions or product of more than two densities may be worth studying. The real variables X and Y may also be replaced by vectors, leading to a multivariate and anisotropy problem. Lastly, it is likely that our methods would extend to dependent variables, provided that the two sequences remain independent, but this should be further investigated.

If we come back to the Nadaraya-Watson problem that initiated our question, we justified in our context that plugging an overfitted estimator is an optimal strategy. Other contexts where overfitting has been recognized as judicious exists (see Chinot and Lerasle (2020)). The next step, as the original problem is a ratio, is to address the question of estimating $1/f$ when f is a density.

6 Proofs of section 2

In the sequel C and C' denote generic constants whose value may change from line to line.

6.1 Proof of Proposition 2.1

First we write

$$\|\widehat{(fg)}_{m,m^*} - fg\|^2 = \|\widehat{(fg)}_{m,m^*} - (fg)_m\|^2 + \|(fg)_m - fg\|^2 := \mathbb{T} + \mathbb{B}_1. \tag{18}$$

The term $\mathbb{B}_1$ is the standard integrated squared bias and the first element of inequality (4). Next, we study $\mathbb{T}$. Using that the basis is orthonormal, we get

$$\mathbb{T} := \|\widehat{(fg)}_{m,m^*} - (fg)_m\|^2 = \sum_{j=0}^{m-1} \left(\widehat{a}_j^{(m^*)} - a_j(fg)\right)^2.$$

We know that $\mathbb{E}(\widehat{a}_j^{(m^*)}) = a_j(f_{m^*}g)$. Thus,

$$\mathbb{E}(\mathbb{T}) = \sum_{j=0}^{m-1} \mathbb{E}\left[\left(\widehat{a}_j^{(m^*)} - a_j(f_{m^*}g)\right)^2\right] + \sum_{j=0}^{m-1} (a_j(f_{m^*}g) - a_j(fg))^2 := \mathbb{V} + \mathbb{B}_2, \tag{19}$$

where $\mathbb{V}$ is a variance term and $\mathbb{B}_2$ a second bias term. We have

$$\mathbb{B}_2 = \sum_{j=0}^{m-1} \langle \varphi_j, g(f_{m^*} - f)\rangle^2 = \|[g(f_{m^*} - f)]_m\|^2 \leq \|g(f_{m^*} - f)\|^2 \leq \|g\|_\infty^2 \|f - f_{m^*}\|^2. \tag{20}$$

Then, we split $\mathbb{V}$ to involve the conditional variance given $(X_1, \ldots, X_n) := \mathbf{X}$. As

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \varphi_j(Y_i)\widehat{f}_{m^*}(Y_i) \,|\mathbf{X}\right] = \langle \varphi_j, g\widehat{f}_{m^*}\rangle,$$

this yields

$$\begin{aligned} \mathbb{V} &= \sum_{j=0}^{m-1}\left\{\mathbb{E}\left[\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\varphi_j(Y_i)\widehat{f}_{m^*}(Y_i)-\langle\varphi_j, g\widehat{f}_{m^*}\rangle\right)^2|\mathbf{X}\right]+\mathbb{E}\left[\langle\varphi_j, g(\widehat{f}_{m^*}-f_m^*)\rangle^2\right]\right\} \\ &:= \mathbb{V}_1+\mathbb{V}_2. \end{aligned}$$

We successively study $\mathbb{V}_1$ and $\mathbb{V}_2$. First, for $\mathbb{V}_1$, we have

$$\begin{aligned} \mathbb{V}_1 &= \sum_{j=1}^{m-1}\mathbb{E}\left[\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\varphi_j(Y_i)\widehat{f}_{m^*}(Y_i)\right)|\mathbf{X}\right] = \frac{1}{n}\sum_{j=0}^{m-1}\mathrm{Var}\left(\varphi_j(Y_1)\widehat{f}_{m^*}(Y_1)\right) \\ &\leq \frac{1}{n}\sum_{j=0}^{m-1}\mathbb{E}\left(\varphi_j^2(Y_1)[\widehat{f}_{m^*}(Y_1)]^2\right) \leq \frac{L(m)}{n}\mathbb{E}\left([\widehat{f}_{m^*}(Y_1)]^2\right). \end{aligned}$$

Now, as $\mathbb{E}\left([\widehat{f}_{m^*}(Y_1)]^2\right)=\mathbb{E}\left([\widehat{f}_{m^*}(Y_1)-f_{m^*}(Y_1)]^2\right)+\mathbb{E}\left([f_{m^*}(Y_1)]^2\right)$, it holds

$$\mathbb{E}\left([\widehat{f}_{m^*}(Y_1)]^2\right) \leq \|g\|_\infty\frac{L(m^*)}{n}+\|g\|_\infty\|f\|^2 \leq (1+\|f\|^2)\|g\|_\infty,$$

for m^* such that $L(m^*)\leq n$. Finally we get

$$\mathbb{V}_1 \leq \|g\|_\infty(1+\|f\|^2)\frac{L(m)}{n}.$$

Next, we turn to the study of $\mathbb{V}_2$,

$$\begin{aligned} \mathbb{V}_2 &= \sum_{j=0}^{m-1}\mathbb{E}\left(\langle\varphi_j, g(\widehat{f}_{m^*}-f_m^*)\rangle^2\right) = \sum_{j=0}^{m-1}\mathbb{E}\left(\langle\varphi_j, g\sum_{k=0}^{m^*-1}(\widehat{a}_k-a_k(f))\varphi_k\rangle^2\right) \\ &= \sum_{j=0}^{m-1}\sum_{k,\ell=0}^{m^*-1}\mathrm{cov}(\widehat{a}_k,\widehat{a}_\ell)\langle\varphi_j, g\varphi_k\rangle\langle\varphi_j, g\varphi_\ell\rangle \end{aligned}$$

As $\mathrm{cov}(\widehat{a}_k,\widehat{a}_\ell)=\frac{1}{n^2}\sum_{i,i'=1}^{n}\mathrm{cov}(\varphi_k(X_i),\varphi_\ell(X_{i'}))=\frac{1}{n}\mathrm{cov}(\varphi_k(X_1),\varphi_\ell(X_1))$, we get

$$\begin{aligned} \mathbb{V}_2 &= \frac{1}{n}\sum_{j=0}^{m-1}\sum_{k,\ell=0}^{m^*-1}\mathrm{cov}(\varphi_k(X_1),\varphi_\ell(X_1))\langle\varphi_j, g\varphi_k\rangle\langle\varphi_j, g\varphi_\ell\rangle = \frac{1}{n}\sum_{j=0}^{m-1}\mathrm{Var}\left(\sum_{k=0}^{m^*-1}\langle\varphi_j, g\varphi_k\rangle\varphi_k(X_1)\right) \\ &\leq \frac{1}{n}\sum_{j=0}^{m-1}\mathbb{E}\left[\left(\sum_{k=0}^{m^*-1}\langle\varphi_j, g\varphi_k\rangle\varphi_k(X_1)\right)^2\right] = \frac{1}{n}\sum_{j=0}^{m-1}\int\left(\sum_{k=0}^{m^*-1}\langle\varphi_j, g\varphi_k\rangle\varphi_k(u)\right)^2 f(u)du \\ &\leq \frac{\|f\|_\infty}{n}\sum_{j=0}^{m-1}\sum_{k=0}^{m^*-1}\langle\varphi_k,\varphi_j g\rangle^2 = \frac{\|f\|_\infty}{n}\sum_{j=0}^{m-1}\|(\varphi_j g)_{m^*}\|^2 \\ &\leq \frac{\|f\|_\infty}{n}\sum_{j=0}^{m-1}\|\varphi_j g\|^2 \leq \frac{\|f\|_\infty L(m)\|g\|^2}{n}. \end{aligned}$$

The bounds for $\mathbb{V}_1$ and $\mathbb{V}_2$ imply

$$\mathbb{V} \leq \left(\|g\|_\infty(1+\|f\|^2)+\|f\|_\infty\|g\|^2\right)\frac{L(m)}{n}. \tag{21}$$

Now, plugging (20) and (21) in (19) and the result in the expectation of (18) gives (4) and Proposition 2.1 is proved. □

6.2 Proof of Theorem 2.1

First we note that

$$\begin{aligned}\gamma_n(t)-\gamma_n(s) &= \|t-fg\|^2-\|s-fg\|^2+2\langle t-s,fg\rangle-\frac{2}{n}\sum_{i=1}^n \widehat{f}_{m^*}(Y_i)(t-s)(Y_i)\\ &= \|t-fg\|^2-\|s-fg\|^2+2\langle t-s,fg-f_{m^*}g\rangle-2\nu_n(t-s),\end{aligned}$$

where $\nu_n(t) := \frac{1}{n}\sum_{i=1}^n[\widehat{f}_{m^*}(Y_i)t(Y_i)-\langle f_{m^*}g,t\rangle]$. By definition of $\widehat{m}$, it holds, for any $m\in\mathcal{M}_n$,

$$\gamma_n(\widehat{(fg)}_{\widehat{m},m^*})+\text{pen}(\widehat{m})\le\gamma_n((fg)_m)+\text{pen}(m),$$

so that we get

$$\begin{aligned}\|\widehat{(fg)}_{\widehat{m},m^*}-fg\|^2 \le\ & \|(fg)_m-fg\|^2+2\langle\widehat{(fg)}_{\widehat{m},m^*}-(fg)_m,g(f_{m^*}-f)\rangle+\text{pen}(m)\\ &+2\nu_n(\widehat{(fg)}_{\widehat{m},m^*}-(fg)_m)-\text{pen}(\widehat{m}).\end{aligned}$$

Now we have

$$\begin{aligned}2|\langle\widehat{(fg)}_{\widehat{m},m^*}-(fg)_m,g(f_{m^*}-f)\rangle| &\le \frac{1}{8}\|\widehat{(fg)}_{\widehat{m},m^*}-(fg)_m\|^2+8\|g(f_{m^*}-f)\|^2\\ &\le \frac{1}{4}\|\widehat{(fg)}_{\widehat{m},m^*}-fg\|^2+\frac{1}{4}\|(fg)_m-fg\|^2+8\|g\|_\infty\|(f_{m^*}-f)\|^2,\end{aligned}$$

and as $t\mapsto\nu_n(t)$ is linear,

$$\begin{aligned}2|\nu_n(\widehat{(fg)}_{\widehat{m},m^*}-(fg)_m)| &\le \frac{1}{8}\|\widehat{(fg)}_{\widehat{m},m^*}-(fg)_m\|^2+8\sup_{t\in S_m+S_{\widehat{m}},\|t\|=1}\nu_n^2(t)\\ &\le \frac{1}{4}\|\widehat{(fg)}_{\widehat{m},m^*}-fg\|^2+\frac{1}{4}\|(fg)_m-fg\|^2+8\sup_{t\in S_m+S_{\widehat{m}},\|t\|=1}\nu_n^2(t).\end{aligned}$$

Thus, we find

$$\begin{aligned}\frac{1}{2}\|\widehat{(fg)}_{\widehat{m},m^*}-fg\|^2 \le\ & \frac{3}{2}\|(fg)_m-fg\|^2+8\|g\|_\infty\|(f_{m^*}-f)\|^2+\text{pen}(m)\\ &+8\sup_{t\in S_m+S_{\widehat{m}},\|t\|=1}\nu_n^2(t)-\text{pen}(\widehat{m}).\end{aligned}$$

Consequently, we get

$$\begin{aligned}\|\widehat{(fg)}_{\widehat{m},m^*}-fg\|^2 \le\ & 3\|(fg)_m-fg\|^2+16\|g\|_\infty\|(f_{m^*}-f)\|^2+2\text{pen}(m)\\ &+16\left(\sup_{t\in S_m+S_{\widehat{m}},\|t\|=1}\nu_n^2(t)-p(m\vee\widehat{m})\right)+16p(m\vee\widehat{m})-2\text{pen}(\widehat{m}).\end{aligned}$$

Now the following Lemma can be proved:

Lemma 6.1. *Under the Assumptions of Theorem 2.1, there exists κ_0 such that for any $\kappa\ge\kappa_0$ and $p(m)=\kappa\|f\|_\infty(1+\|g\|_\infty+\|f\|_\infty)L(m)/n$, then it holds*

$$\sum_m\mathbb{E}\left(\sup_{t\in S_m,\|t\|=1}\nu_n^2(t)-p(m)\right)_+\le\frac{C}{n}.$$

Then applying Lemma 6.1, for κ such that $8p(m)\le\text{pen}(m)$, we get $8p(m\vee\widehat{m})-\text{pen}(\widehat{m})\le 8p(m)+8p(\widehat{m})-\text{pen}(\widehat{m})\le\text{pen}(m)$ which gives the result of Theorem 2.1. □

Proof of Lemma 6.1. We split ν_n in four terms: $\nu_n = \nu_{n,1} + \nu_{n,2} + \nu_{n,3} + \nu_{n,4}$ where for some positive constant c_0 to be defined in the sequel, we set

$$A(x) = \{|\widehat{f}_{m^*}(x) - f_{m^*}(x)| < c_0\},$$

and

$$\nu_{n,1}(t) = \frac{1}{n}\sum_{i=1}^{n}[(\widehat{f}_{m^*}(Y_i) - f_{m^*}(Y_i))\mathbf{1}_{A(Y_i)}t(Y_i) - \langle(\widehat{f}_{m^*} - f_{m^*})\mathbf{1}_A, gt\rangle],$$

$$\nu_{n,2}(t) = \frac{1}{n}\sum_{i=1}^{n}[(\widehat{f}_{m^*}(Y_i) - f_{m^*}(Y_i))\mathbf{1}_{(A(Y_i))^c}t(Y_i) - \langle(\widehat{f}_{m^*} - f_{m^*})\mathbf{1}_{A^c}, gt\rangle],$$

$$\nu_{n,3}(t) = \langle(\widehat{f}_{m^*} - f_{m^*}), gt\rangle = \frac{1}{n}\sum_{i=1}^{n}\psi_t(X_i), \quad \psi_t(X) = \sum_{j=0}^{m^*-1}(\varphi_j(X) - \mathbb{E}(\varphi_j(X)))\int\varphi_j tg,$$

$$\nu_{n,4}(t) = \frac{1}{n}\sum_{i=1}^{n}[f_{m^*}(Y_i)t(Y_i) - \langle f_{m^*}, gt\rangle].$$

Study of $\nu_{n,2}$ Let $B_m := \{t \in S_m, \|t\| = 1\}$. We start by the study of $\nu_{n,2}$ as it leads to fix c_0, and we first establish that $\mathbb{E}(\sup_{t\in B_m}|\nu_{n,2}(t)|) \le n^{-p}$ for some positive p. It holds that $\mathbb{E}(\sup_{t\in B_m}[\nu_{n,2}^2(t)]) \le \sum_{j=0}^{m-1}\mathbb{E}[\nu_{n,2}^2(\varphi_j)]$. We note that

$$\mathbb{E}[\nu_{n,2}^2(\varphi_j)] = \mathrm{Var}(\nu_{n,2}(\varphi_j)) = \mathbb{E}(\mathrm{Var}(\nu_{n,2}(\varphi_j)|\mathbf{X})) + \mathrm{Var}(\mathbb{E}(\nu_{n,2}(\varphi_j)|\mathbf{X})),$$

since $\mathbb{E}(\nu_{n,2}(\varphi_j)|\mathbf{X}) = 0$, we get $\mathbb{E}[\nu_{n,2}^2(\varphi_j)] = \mathbb{E}[\mathrm{Var}(\nu_{n,2}(\varphi_j)|\mathbf{X})]$. Next, we derive that

$$\begin{aligned}\mathrm{Var}(\nu_{n,2}(\varphi_j)|\mathbf{X}) &= \frac{1}{n}\mathrm{Var}[(\widehat{f}_{m^*}(Y_1) - f_{m^*}(Y_1))\mathbf{1}_{(A(Y_1))^c}\varphi_j(Y_1)|\mathbf{X}] \\ &\le \frac{1}{n}\mathbb{E}\left\{\left[(\widehat{f}_{m^*}(Y_1) - f_{m^*}(Y_1))\mathbf{1}_{(A(Y_1))^c}\varphi_j(Y_1)\right]^2|\mathbf{X}\right\}\end{aligned}$$

and it follows

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in B_m}[\nu_{n,2}^2(t)]\right) &\le \frac{1}{n}\sum_{j=0}^{m-1}\mathbb{E}\left\{\left[(\widehat{f}_{m^*}(Y_1) - f_{m^*}(Y_1))\mathbf{1}_{(A(Y_1))^c}\varphi_j(Y_1)\right]^2\right\} \\ &\le \frac{L(m)}{n}\mathbb{E}\left\{\left[(\widehat{f}_{m^*}(Y_1) - f_{m^*}(Y_1))\mathbf{1}_{(A(Y_1))^c}\right]^2\right\}.\end{aligned}$$

The last term can be written as

$$\mathbb{E}\left\{\left[(\widehat{f}_{m^*}(Y_1) - f_{m^*}(Y_1))\mathbf{1}_{(A(Y_1))^c}\right]^2\right\} = \mathbb{E}\left[\int(\widehat{f}_{m^*}(u) - f_{m^*}(u))^2\mathbf{1}_{(A(u))^c}g(u)du\right].$$

Then, we find an upper bound for

$$\begin{aligned}(\widehat{f}_{m^*}(u) - f_{m^*}(u))^2 &= \left(\sum_{j=0}^{m^*-1}(\widehat{a}_j - \mathbb{E}(\widehat{a}_j))\varphi_j(u)\right)^2 \\ &\le L(m^*)\sum_{j=0}^{m^*-1}(\widehat{a}_j - \mathbb{E}(\widehat{a}_j))^2 = L(m^*)\sum_{j=0}^{m^*-1}\left(\frac{1}{n}\sum_{i=1}^{n}(\varphi_j(X_i) - \mathbb{E}(\varphi_j(X_i)))\right)^2 \\ &\le 4(L(m^*))^2 \le 4C^2n^2,\end{aligned}$$

since [A2] implies that $L(m^*) \le Cn$ for C a positive constant. Therefore,

$$\mathbb{E}\left(\sup_{t\in B_m}[\nu_{n,2}^2(t)]\right) \le 4n^2\int\mathbb{P}(|\widehat{f}_{m^*}(u) - f_{m^*}(u)| > c_0)g(u)du. \tag{22}$$

We complete by applying the Bernstein inequality to $Z_i = \sum_{j=0}^{m^*-1} \varphi_j(X_i)\varphi_j(u)$ yielding

$$\mathbb{P}\left(|\widehat{f}_{m^*}(u) - f_{m^*}(u)| > c_0\right) = \mathbb{P}\left(|\sum_{i=1}^{n} Z_i - \mathbb{E}(Z_i)| > nc_0\right) \leq 2\exp\left(-\frac{nc_0^2}{2(v_2^2 + b_2 c_0)}\right)$$

with v_2^2 a bound on $\mathrm{Var}(Z_i)$ and b_2 an a.s. bound on Z_i. We find that $b_2 = L(m^*)$ suits and

$$\mathrm{Var}(Z_i) \leq \mathbb{E}\left[\left(\sum_{j=0}^{m^*-1} \varphi_j(X_i)\varphi_j(u)\right)^2\right] \leq \|f\|_\infty \sum_{j=0}^{m^*-1} \varphi_j^2(u) \leq \|f\|_\infty L(m^*).$$

Therefore, choosing

$$c_0 = \|f\|_\infty, \tag{23}$$

and using that, by [A2] $L(m^*) \leq c_1 n/\log(n)$ where $c_1 = \|f\|_\infty/p$ (here $p = 16$), it follows that

$$\mathbb{P}\left(|\widehat{f}_{m^*}(u) - f_{m^*}(u)| > c_0\right) \leq 2n^{-\frac{p}{4}}. \tag{24}$$

Then, gathering (22) and (24) leads to $\mathbb{E}(\sup_{t\in B_m} \nu_{n,2}^2(t)) \leq 8n^{2-\frac{p}{4}} = \frac{8}{n^2}$, for $p = 16$. As a consequence under [A5], we get

$$\sum_{m\in\mathcal{M}_n} \mathbb{E}(\sup_{t\in B_m} [\nu_{n,2}^2(t)]) \leq \frac{C}{n}. \tag{25}$$

Study of $\nu_{n,1}$. We apply the Talagrand inequality (see Lemma 8.1) to $\nu_{n,1}$ conditionally to $\mathbf{X}$. Using that $t \mapsto \nu_{n,1}(t)$ is linear and the Cauchy-Schwarz inequality, we get

$$\begin{aligned}
\left(\mathbb{E}\left(\sup_{t\in B_m} |\nu_{n,1}(t)|\,|\mathbf{X}\right)\right)^2 &\leq \mathbb{E}\left(\sup_{t\in B_m} \nu_{n,1}^2(t)|\mathbf{X}\right) \leq \sum_{j=0}^{m-1} \mathbb{E}[\nu_{n,1}^2(\varphi_j)|\mathbf{X}] \\
&= \frac{1}{n}\sum_{j=0}^{m-1} \mathrm{Var}[(\widehat{f}_{m^*}(Y_1) - f_{m^*}(Y_1))\mathbf{1}_{A(Y_1)}\varphi_j(Y_1)|\mathbf{X}] \\
&\leq \frac{1}{n}\sum_{j=0}^{m-1} \mathbb{E}[(\widehat{f}_{m^*}(Y_1) - f_{m^*}(Y_1))^2\mathbf{1}_{A(Y_1)}\varphi_j^2(Y_1)|\mathbf{X}] \\
&\leq \frac{c_0^2 L(m)}{n} := H_1^2.
\end{aligned}$$

Next, note that $\sup_{x,t} |(\widehat{f}_{m^*}(x) - f_{m^*}(x))\mathbf{1}_{A(x)}t(x)| \leq c_0 \sup_{t,x}|t(x)| \leq c_0\sqrt{L(m)} := b_1$ and

$$\sup_t \mathrm{Var}((\widehat{f}_{m^*}(Y_1) - f_{m^*}(Y_1))^2\mathbf{1}_{A(Y_1)}t(Y_1)|\mathbf{X}) \leq c_0^2\|g\|_\infty := v_1^2.$$

Applying Lemma 8.1 with $\delta = \frac{1}{2}$, it follows that

$$\begin{aligned}
\mathbb{E}\left[\left(\sup_{t\in B_m} |\nu_{n,1}(t)|^2 - 4\frac{c_0^2 L(m)}{n}\right)_+ \Big|\mathbf{X}\right] &\leq \frac{4c_0^2}{nK_1}\left(\|g\|_\infty \exp\left(-K_1\frac{L(m)}{2\|g\|_\infty}\right)\right. \\
&\quad \left. + \frac{49}{K_1 C^2(1/2)}\exp\left(-\frac{K_1 C(1/2)}{7}\sqrt{n}\right)\right).
\end{aligned}$$

Since the latter bound does not depend on $\mathbf{X}$, the inequality holds unconditionally in expectation. Therefore under [A5] and as $L(m) \geq 1$ for $m \geq m_0$, we get, for C a positive constant, and using (23),

$$\sum_{m\in\mathcal{M}_n} \mathbb{E}\left[\left(\sup_{t\in B_m} |\nu_{n,1}(t)|^2 - 4\frac{\|f\|_\infty^2 L(m)}{n}\right)_+\right] \leq \frac{C}{n}. \tag{26}$$

Study of $\nu_{n,3}$ Similarly to $\nu_{n,1}$ we apply the Talagrand inequality

$$\left(\mathbb{E}\left(\sup_{t\in B_m}|\nu_{n,3}(t)|\right)\right)^2 \leq \sum_{j=0}^{m-1}\mathbb{E}[\nu_{n,3}^2(\varphi_j)] = \frac{1}{n}\sum_{j=0}^{m-1}\mathrm{Var}(\psi_{\varphi_j}(X_1))$$

$$\leq \frac{1}{n}\sum_{j=0}^{m-1}\mathbb{E}\left[\left(\sum_{k=0}^{m^*-1}\varphi_k(X_1)\langle\varphi_k,\varphi_j g\rangle\right)^2\right] = \frac{1}{n}\sum_{j=0}^{m-1}\int\left(\sum_{k=0}^{m^*-1}\varphi_k(u)\langle\varphi_k,\varphi_j g\rangle\right)^2 f(u)du$$

$$\leq \frac{\|f\|_\infty}{n}\sum_{j=0}^{m-1}\sum_{k=0}^{m^*-1}\langle\varphi_k,\varphi_j g\rangle^2 \leq \frac{\|f\|_\infty}{n}\sum_{j=0}^{m-1}\|\varphi_j g\|^2 \leq \frac{\|f\|_\infty\|g\|^2 L(m)}{n} \leq \frac{\|f\|_\infty\|g\|_\infty L(m)}{n} := H_3^2.$$

Next, note that $\sup_{x,t}\left|\sum_{k=0}^{m^*-1}\varphi_k(x)\langle\varphi_k,tg\rangle\right| \leq 2\sqrt{L(m^*)\|tg\|^2} \leq 2\sqrt{n}\|g\|_\infty := b_3$ and

$$\sup_t \mathrm{Var}\left(\sum_{k=0}^{m^*-1}\varphi_k(X_1)\langle\varphi_k,tg\rangle\right) \leq \|f\|_\infty\|g\|_\infty^2 := v_3^2.$$

Applying Lemma 8.1 with $\delta = \frac{1}{2}$, it follows that

$$\mathbb{E}\left[\left(\sup_{t\in B_m}|\nu_{n,3}(t)|^2 - 4\|f\|_\infty\|g\|_\infty\frac{L(m)}{n}\right)_+\right] \leq \frac{4\|g\|_\infty^2}{K_1\, n}\left(\|f\|_\infty\exp\left(-K_1\frac{L(m)}{2\|g\|_\infty}\right)\right.$$
$$\left. + \frac{49\times 4}{K_1 C^2(1/2)}\exp\left(-\frac{K_1 C(1/2)\sqrt{\|f\|_\infty}}{14\sqrt{\|g\|_\infty}}\sqrt{L(m)}\right)\right).$$

Therefore under [A5], we get, for C a positive constant,

$$\sum_{m\in\mathcal{M}_n}\mathbb{E}\left[\left(\sup_{t\in B_m}|\nu_{n,3}(t)|^2 - 4\frac{\|f\|_\infty\|g\|_\infty L(m)}{n}\right)_+\right] \leq \frac{C}{n}. \tag{27}$$

Study of $\nu_{n,4}$ Again we apply the Talagrand inequality, similar computations enable to derive $H_4^2 = \|f\|_\infty\|g\|_\infty L(m)/n$. To obtain v_4^2 we first write

$$\sup_{t\in B_m}\mathrm{Var}(f_{m^*}(Y_1)t(Y_1)) \leq \sup_{t\in B_m}\mathbb{E}(f_{m^*}^2(Y_1)t^2(Y_1)) = \sup_{t\in B_m}\int f_{m^*}(u)^2 t(u)^2 g(u)du$$
$$\leq \|g\|_\infty\|f_{m^*}\|_\infty^2.$$

It remains to bound $\|f_{m^*}\|_\infty$. Under [A4], we have

$$|f_{m^*}(x)| = \left|a_0(f)\varphi_0(x) + \sum_{j\geq m^*}a_j(f)\varphi_j(x)\right| \leq C_\varphi\left(|a_0| + \sum_{j\geq 1}|a_j(f)|\right).$$

Then using [A3], we have

$$|f_{m^*}(x)| \leq C_\varphi\left(C_\varphi + \sqrt{\sum_{j\geq 1}j^a a_j^2(f)\sum_{j\geq 1}j^{-a}}\right) \leq C_\varphi\left(C_\varphi + \sqrt{C_a\sum_{j\geq 1}j^{-a}}\right) := C(a,\varphi) < +\infty$$

since $a > 1$. Thus, we can set $v_4^2 := \|g\|_\infty C^2(a,\varphi)$. Similarly, we derive

$$\sup_{t\in B_m}\|f_{m^*}t - \langle f_{m^*}, tf\rangle\|_\infty \leq 2\sup_{t\in B_m}\|f_{m^*}t\|_\infty \leq 2\|f_{m^*}\|_\infty\sqrt{L(m)} \leq 2C(a,\varphi)\sqrt{L(m)} =: b_4.$$

It follows, by applying Lemma 8.1 with $\delta = \frac{1}{2}$, that

$$\mathbb{E}\left[\left(\sup_{t\in B_m} |\nu_{n,4}(t)|^2 - 4\|f\|_\infty\|g\|_\infty \frac{L(m)}{n}\right)_+\right] \leq \frac{4}{K_1}\left(\frac{\|f\|_\infty C^2(a,\varphi)}{n}\exp\left(-K_1\frac{L(m)\|f\|_\infty}{2C^2(a,\varphi)}\right)\right.$$
$$\left. + \frac{49\times 4C^2(a,\varphi)}{K_1 n C^2(1/2)}\exp\left(-\frac{K_1 C(1/2)}{14}\frac{\sqrt{\|f\|_\infty\|g\|_\infty}}{C(a,\varphi)}\sqrt{n}\right)\right).$$

Therefore under [A5], we get, for C a positive constant,

$$\sum_{m\in\mathcal{M}_n}\mathbb{E}\left[\left(\sup_{t\in B_m}|\nu_{n,4}(t)|^2 - 4\frac{\|f\|_\infty\|g\|_\infty L(m)}{n}\right)_+\right] \leq \frac{C}{n}. \tag{28}$$

As a consequence, gathering (25)-(26)-(27) and (28) gives the result of Lemma 6.1 for C a positive finite constant, depending on a, $\|f\|_\infty$, $\|g\|_\infty$ and C_φ. $\square$

7 Proofs of Section 3

7.1 Preliminary tools

In the sequel we make an extensive use of the following:

- The Young Inequality: for $u \in L^p$ and $v\in L^q$, $1\leq p,q\leq r\leq\infty$,

$$\|u\star v\|_r \leq \|u\|_p\|v\|_q, \quad \frac{1}{r}+1 = \frac{1}{p}+\frac{1}{q}, \tag{29}$$

- The Bernstein inequality: For i.i.d. random variables Z_i, set $S_n = \frac{1}{n}\sum_{i=1}^n (Z_i - \mathbb{E}[Z_i])$. If $\mathbb{E}[Z_1^2]\leq\mathfrak{v}$ and $|Z_1|\leq\mathfrak{b}$ a.s. then with probability larger than $1-2e^{-\lambda}$, for any $\lambda>0$,

$$|S_n| \leq \sqrt{\frac{2\mathfrak{v}\lambda}{n}} + \frac{\lambda}{n}\mathfrak{b}. \tag{30}$$

- Deriving a bound in expectation from a bound on probability: If $\mathbb{P}(Z\geq\frac{\kappa\lambda}{n})\leq n^2e^{-\lambda}$ for all $\lambda>0$, then it holds

$$\mathbb{E}[Z_+] \leq c\frac{\log(n)}{n}. \tag{31}$$

  Indeed, for all positive A we have

$$\mathbb{E}[Z_+] = \int_0^\infty \mathbb{P}(Z\geq x)dx = \frac{\kappa}{n}\int_0^\infty \mathbb{P}(Z\geq\frac{\kappa\lambda}{n})d\lambda \leq \frac{\kappa}{n}\left(A + 2n^2e^{-A}\right),$$

  and $A = 2\log(n)$ gives the result.

7.2 Proof of Proposition 3.1

First we note that

$$\mathbb{E}\left(\widetilde{(fg)}_{h,h_o}(x)\right) = \mathbb{E}(f_{h_o}(Y_1)K_h(Y_1-x)) = (f_{h_o}g)\star K_h(x).$$

Then, write the bias variance decomposition:

$$\mathbb{E}\left(\|\widetilde{(fg)}_{h,h_o} - fg\|^2\right) = \underbrace{\int[(f_{h_o}g)\star K_h(x) - fg(x)]^2dx}_{:=\mathfrak{B}} + \underbrace{\mathbb{E}\left(\|\widetilde{(fg)}_{h,h_o} - (f_{h_o}g)\star K_h\|^2\right)}_{:=\mathfrak{V}}.$$

Let us first study $\mathfrak{B}$. We have

$$\mathfrak{B} \leq 2\int [(f_{h_o}g)\star K_h(x) - (fg)\star K_h(x)]^2 dx + 2\int [(fg)\star K_h(x) - fg(x)]^2 dx := 2\mathfrak{B}_1 + 2\mathfrak{B}_2.$$

The term $2\mathfrak{B}_2$ is the first rhs term of (13). Next, we have

$$\mathfrak{B}_1 = \int [(f_{h_o} - f)g]\star K_h(x))^2 dx = \|[(f_{h_o} - f)g]\star K_h\|^2 \leq \|(f_{h_o} - f)g\|^2 \|K_h\|_1^2$$

by applying Young Inequality (29) with $r = p = 2$ and $q = 1$. We get

$$\mathfrak{B}_1 \leq \int (f_{h_o} - f)^2(x) g^2(x) dx (\int |K(u)|du)^2 \leq \|g\|_\infty^2 \|K\|_1^2 \|f_{h_o} - f\|^2$$

which gives the second term of (13). Next, we split $\mathfrak{V} = \mathfrak{V}_1 + \mathfrak{V}_2$ where

$$\mathfrak{V}_1 := \mathbb{E}\left\{\int \left[\frac{1}{n}\sum_{i=1}^n \widetilde{f}_{h_o}(Y_i)K_h(Y_i - x) - (\widetilde{f}_{h_o}g)\star K_h(x)\right]^2 dx\right\}$$

$$\mathfrak{V}_2 := \mathbb{E}\left\{\int \left[(\widetilde{f}_{h_o} - f_{h_o})g)\star K_h(x)\right]^2 dx\right\}.$$

First, we have

$$\mathfrak{V}_1 = \frac{1}{n}\int \mathrm{Var}\left(\widetilde{f}_{h_o}(Y_1)K_h(Y_1 - x)\right) dx \leq \frac{1}{n}\int \mathbb{E}\left[\left(\widetilde{f}_{h_o}(Y_1)K_h(Y_1 - x)\right)^2\right] dx.$$

Then, we write that

$$\mathbb{E}\left[\left(\widetilde{f}_{h_o}(Y_1)K_h(Y_1 - x)\right)^2\right] = \mathbb{E}\left[(\widetilde{f}_{h_o} - f_{h_o})^2(Y_1)K_h^2(Y_1 - x)\right] + \mathbb{E}\left[f_{h_o}^2(Y_1)K_h^2(Y_1 - x)\right]$$

where

$$\begin{aligned}
\int \mathbb{E}\left[f_{h_o}^2(Y_1)K_h^2(Y_1 - x)\right] dx &= \int\int f_{h_o}^2(u)K_h^2(u - x)g(u)dudx = \int K_h^2(v)dv \int f_{h_o}^2(u)g(u)du \\
&\leq \frac{\|K\|^2}{h}\|g\|_\infty \|f\star K_{h_o}\|^2 \\
&\leq \frac{\|K\|^2}{h}\|g\|_\infty \|f\|^2 \|K_{h_o}\|_1^2 = \frac{\|K\|^2\|K\|_1^2\|f\|^2\|g\|_\infty}{h}
\end{aligned}$$

and

$$\begin{aligned}
\int \mathbb{E}\left[(\widetilde{f}_{h_o} - f_{h_o})^2(Y_1)K_h^2(Y_1 - x)\right] dx &= \int\int \mathbb{E}\left[(\widetilde{f}_{h_o} - f_{h_o})^2(u)\right] K_h^2(u - x)g(u)dudx \\
&= \|K_h\|^2 \int \mathbb{E}\left[(\widetilde{f}_{h_o} - f_{h_o})^2(u)\right] g(u)du \\
&\leq \frac{\|K\|^2}{h}\|g\|_\infty \int \mathrm{Var}(\widetilde{f}_{h_o}(u))du = \frac{\|K\|^2}{h}\|g\|_\infty \frac{\|K\|^2}{nh_o} \\
&\leq \frac{\|K\|^4}{h}\|g\|_\infty \quad \text{as } \frac{1}{nh_o} \leq 1.
\end{aligned}$$

As a consequence,

$$\mathfrak{V}_1 \leq \frac{\|K\|^2\|g\|_\infty(\|K\|^2 + \|K\|_1^2\|f\|^2)}{nh}.$$

Next, we bound $\mathfrak{V}_2$ as follows.

$$\begin{aligned}
\mathfrak{V}_2 &= \int \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n}(K_{h_o}(X_i-.)g)\star K_h(x)\right]dx = \frac{1}{n}\int \mathrm{Var}\left((K_{h_o}(X_1-.)g)\star K_h(x)\right)dx \\
&\leq \frac{1}{n}\int \mathbb{E}\left[((K_{h_o}(X_1-.)g)\star K_h(x))^2\right]dx = \frac{1}{n}\int\int \left[((K_{h_o}(u-.)g)\star K_h(x))^2\right]dx f(u)du \\
&= \frac{1}{n}\int \|(K_{h_o}(u-.)g)\star K_h\|^2 f(u)du \\
&\leq \frac{1}{n}\int \|(K_{h_o}(u-.)g)\|_1^2\|K_h\|^2 f(u)du \quad \text{by the Young Inequality (29)} \\
&\leq \frac{\|K\|^2}{nh}\int (\|g\|_\infty \|K_{h_o}\|_1)^2 f(u)du = \frac{\|K\|^2\|K\|_1^2\|g\|_\infty^2}{nh}.
\end{aligned}$$

Finally we get,

$$\mathfrak{V} \leq \frac{\|K\|^2\|g\|_\infty(\|K\|^2 + 2\|K\|_1^2\|f\|^2)}{nh},$$

which is the last term of Inequality (13) and ends the proof of Proposition 3.1 . □

7.3 Proof of Theorem 3.1

For simplicity we write in the sequel $\widetilde{(fg)}_{h,h'} = K_{h'} \star \widetilde{(fg)}_h$. The proof starts by decompositions which are standard when studying Goldenschluger and Lepski (2011) methods and bounds. For $\kappa' \geq \kappa$, we get (see Comte (2017), sec 4.2)

$$\mathbb{E}\left(\|\widetilde{(fg)}_{\check{h}} - fg\|^2\right) \leq 3\mathbb{E}(\|\widetilde{(fg)}_h - fg\|^2) + 6\kappa' V(h) + 6\mathbb{E}(A(h)). \tag{32}$$

Recall that $u_h = u \star K_h$ and set also $u_{h,h'} = u \star K_h \star K_{h'} = u_{h',h}$. Then, to start the study of $A(h)$, we write

$$\|\widetilde{(fg)}_{h'} - \widetilde{(fg)}_{h,h'}\|^2 \leq 3(\|\widetilde{(fg)}_{h'} - (fg)_{h'}\|^2 + \|(fg)_{h'} - (fg)_{h,h'}\|^2 + \|\widetilde{(fg)}_{h,h'} - (fg)_{h,h'}\|^2).$$

The bound on the middle term

$$\|(fg)_{h'} - (fg)_{h,h'}\|^2 \leq \|K\|_1^2\|(fg)_h - (fg)\|^2$$

refers to an adequate bias term. Next

$$\|\widetilde{(fg)}_{h'} - (fg)_{h'}\|^2 \leq 2(\|\widetilde{(fg)}_{h'} - (f_{h_o}g)_{h'}\|^2 + \|(f_{h_o}g)_{h'} - (fg)_{h'}\|^2)$$

and we have the bound $\|(f_{h_o}g)_{h'} - (fg)_{h'}\|^2 \leq \|K\|_1^2\|g\|_\infty\|f_{h_o} - f\|^2$. Now we notice that $\|\widetilde{(fg)}_{h'} - (f_{h_o}g)_{h'}\|^2 = \sup_{t\in\mathcal{B}(0,1)} \nu_n^2(t)$ where $\mathcal{B}(0,1)$ is a countable set of square integrable functions with $\|t\| = 1$ and the empirical process is defined by

$$\nu_n(t) = \langle \widetilde{(fg)}_{h'} - (f_{h_o}g)_{h'}, t\rangle.$$

Therefore we have

$$\|\widetilde{(fg)}_{h'} - (fg)_{h'}\|^2 \leq 2\left(\sup_{t\in\mathcal{B}(0,1)} \nu_n^2(t) + \|K\|_1^2\|g\|_\infty^2\|f_{h_o} - f\|^2\right).$$

Analogously for the last term we get $\|(f_{h_o}g)_{h,h'} - (fg)_{h,h'}\|^2 \leq \|K\|_1^4\|g\|_\infty^2\|f_{h_o} - f\|^2$. More generally, all h, h' terms are handled like the h or h' terms with an additional factor $\|K\|_1^2$ in all bounds. Therefore

$$\|\widetilde{(fg)}_{h,h'} - (fg)_{h,h'}\|^2 \leq 2\left(\sup_{t\in\mathcal{B}(0,1)} \bar{\nu}_n^2(t) + \|K\|_1^4\|g\|_\infty^2\|f_{h_o} - f\|^2\right)$$

with

$$\bar{\nu}_n(t) = \langle \widetilde{(fg)}_{h,h'} - (f_{h_o}g)_{h,h'}, t\rangle.$$

Reminding the definition of $A(h)$ given by (14), we have

$$\begin{aligned}\mathbb{E}(A(h)) \quad \leq \quad & 6\mathbb{E}\left(\sup_{h'\in\mathcal{H}_n}\sup_{t\in\mathcal{B}(0,1)} \nu_n^2(t) - \frac{\kappa}{12}V(h')\right) + 6\mathbb{E}\left(\sup_{h'\in\mathcal{H}_n}\sup_{t\in\mathcal{B}(0,1)} \bar{\nu}_n^2(t) - \frac{\kappa}{12}V(h')\right) \\ & + 6\|K\|_1^4\|g\|_\infty^2\|f_{h_o} - f\|^2 + 6\|K\|_1^2\|(fg)_h - (fg)\|^2. \qquad (33)\end{aligned}$$

Thus, the result of Theorem 3.1 holds if we prove that, for two constants $\mathfrak{c}_1, \mathfrak{c}_2$, we have

$$\sum_{h'\in\mathcal{H}_n} \mathbb{E}\left(\sup_{t\in\mathcal{B}(0,1)} \nu_n^2(t) - \mathfrak{c}_1 V(h')\right) \leq \frac{C}{n} \quad \text{and} \quad \sum_{h'\in\mathcal{H}_n} \mathbb{E}\left(\sup_{t\in\mathcal{B}(0,1)} \bar{\nu}_n^2(t) - \mathfrak{c}_2 V(h')\right) \leq \frac{C}{n}. \quad (34)$$

Indeed, plugging (34) in (33) and the result in (32) is the result of Theorem 3.1.

We prove the first bound of (34), the second one can be checked similarly (with additional factors $\|K\|_1^2$). Define $B(y) := \{|\widetilde{f}_{h_o}(y)| \leq b_0\}$ and $b_0 = 2\|f\|_\infty\|K\|_1$, we split $\nu_n(t) = \sum_{i=1}^3 \nu_{n,i}(t)$ where

$$\nu_{n,1}(t) = \frac{1}{n}\sum_{i=1}^n \widetilde{f_{h_o}}(Y_i)\mathbf{1}_{B(Y_i)}\langle K_h(Y_i - \cdot), t\rangle - \langle [\widetilde{f}_{h_o}\mathbf{1}_B g] \star K_h, t\rangle,$$

$$\nu_{n,2}(t) = \frac{1}{n}\sum_{i=1}^n \widetilde{f_{h_o}}(Y_i)\mathbf{1}_{B(Y_i)^c}\langle K_h(Y_i - \cdot), t\rangle - \langle [\widetilde{f}_{h_o}\mathbf{1}_{B^c} g] \star K_h, t\rangle,$$

$$\nu_{n,3}(t) = \langle [\widetilde{f}_{h_o} - f_{h_o})g] \star K_h, t\rangle = \frac{1}{n}\sum_{i=1}^n \psi_t(X_i), \;\; \psi_t(X_i) := \langle (K_{h_o}(X_i - \cdot) - f_{h_o})g, t \star K_h\rangle.$$

Note that, as K is even,

$$\mathbb{E}\left(\widetilde{f_{h_o}}(Y_i)\mathbf{1}_{B(Y_i)}\langle K_h(Y_i - \cdot), t\rangle \,|\mathbf{X}\right) = \langle [\widetilde{f}_{h_o}\mathbf{1}_B g] \star K_h, t\rangle = \langle [\widetilde{f}_{h_o}\mathbf{1}_B g], t \star K_h\rangle.$$

The terms being similar to the model selection case, we only give a sketch of proof concerning the key bounds associated to the three terms.

Study of $\nu_{n,1}$. First, we compute the bounds required to apply Talagrand Inequality conditional to $\mathbf{X}$. Recall that $b_0 = 2\|f\|_\infty\|K\|_1$.

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in\mathcal{B}(0,1)} \nu_{n,1}^2(t) \,|\mathbf{X}\right) \quad & \leq \quad \frac{1}{n}\int \mathrm{Var}\left(\widetilde{f_{h_o}}(Y_1)\mathbf{1}_{B(Y_1)}K_h(Y_1 - v) \,|\mathbf{X}\right) dv \\ & \leq \quad \frac{1}{n}\int\left(\int (\widetilde{f_{h_o}}(y))^2\mathbf{1}_{B(y)}K_h^2(y - v)g(y)dy\right) dv \\ & \leq \quad \frac{b_0^2}{n}\iint K_h^2(y - v)g(y)dy\, dv \leq \frac{b_0^2\|K\|^2}{nh} := H_1^2.\end{aligned}$$

$$\sup_{y,t} |\widetilde{f_{h_o}}(y)\mathbf{1}_{B(y)}\langle K_h(y - \cdot), t\rangle| \leq b_0\|K_h\| = b_0\|K\|/\sqrt{h} := b_1.$$

$$\begin{aligned}\sup_t \mathrm{Var}\left((\widetilde{f_{h_o}}(Y_1))^2\mathbf{1}_{B(Y_1)}\langle K_h(Y_1 - \cdot), t\rangle^2|\mathbf{X}\right) \quad & \leq \quad b_0^2 \sup_t \mathbb{E}\left(\langle K_h(Y_1 - \cdot), t\rangle^2\right) \\ & \leq \quad b_0^2\|g\|_\infty \sup_t \|K_h \star t\|^2 \\ & \leq \quad b_0^2\|g\|_\infty\|K\|_1^2 := v_1.\end{aligned}$$

Then Talagrand Inequality implies that, under [B4],

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\sup_{t\in\mathcal{B}(0,1)}\nu_{n,1}^2(t)-4\frac{b_0^2\|K\|^2}{nh}\right)\le\sum_{h\in\mathcal{H}_n}\mathbb{E}\left(\sup_{t\in\mathcal{B}(0,1)}\nu_{n,1}^2(t)-4\frac{b_0^2\|K\|^2}{nh}\right)_+\le\frac{C}{n}.$$

Study of $\nu_{n,2}$. First, by noticing that $|\widetilde{f}_{h_o}(x)|\le|\widetilde{f}_{h_o}(x)-f_{h_o}(x)|+\|f\|_\infty\|K\|_1$ we get

$$\mathbb{P}((B(x))^c)=\mathbb{P}(|\widetilde{f}_{h_o}(x)|>b_0)\le\mathbb{P}(|\widetilde{f}_{h_o}(x)-f_{h_o}(x)|>\|f\|_\infty\|K\|_1)$$

and by Bernstein Inequality

$$\mathbb{P}(|\widetilde{f}_{h_o}(x)|>b_0)\le 2\exp\left(-nh_o\frac{b_0^2}{2(\|f\|_\infty\|K\|^2+\|K\|_\infty b_0)}\right)\le 2n^{-p}\tag{35}$$

for $nh_0\ge\mathfrak{b}_1\log(n)$ and $\mathfrak{b}_1\ge 3p\|K\|_\infty/(2\|f\|_\infty\|K\|_1)$, which is ensured under [B3]. Then

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in\mathcal{B}(0,1)}\nu_{n,2}^2(t)\right)&\le\frac{1}{n}\mathbb{E}\left(\int\widetilde{f}_{h_0}^2(y)\mathbf{1}_{B(y)^c}\int K_h^2(y-u)dug(y)dy\right)\\&\le\frac{\|K\|_1^2\|K\|_\infty^2}{nh_0h}\int\mathbb{P}(B^c(y))g(y)dy\le\frac{\|K\|_1^2\|K\|_\infty^2}{2n^{p-1}}\le\frac{C}{n^2},\end{aligned}$$

for $\mathrm{card}(\mathcal{H}_n)\le n$ and $h^{-1}\le n$, $h_o^{-1}\le n$, provided that $p\ge 3$. This implies

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\sup_{t\in\mathcal{B}(0,1)}\nu_{n,2}^2(t)\right)\le\sum_{h\in\mathcal{H}_n}\mathbb{E}\left(\sup_{t\in\mathcal{B}(0,1)}\nu_{n,2}^2(t)\right)\le\frac{C}{n}.$$

Study of $\nu_{n,3}$. We apply Talagrand Inequality with respect to the X_i's. We have

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in\mathcal{B}(0,1)}\nu_{n,3}^2(t)\right)&\le\mathbb{E}\left(\left\|\frac{1}{n}\sum_{i=1}^n[(K_{h_o}(X_i-.)-f_{h_o})g]\star K_h\right\|^2\right)\\&=\mathbb{E}\left(\int\left\{\frac{1}{n}\sum_{i=1}^n[(K_{h_o}(X_i-.)-f_{h_o})g]\star K_h(u)\right\}^2du\right)\\&=\frac{1}{n}\int\mathrm{Var}\left([K_{h_o}(X_1-.)g]\star K_h(u)\right)du\\&\le\frac{1}{n}\mathbb{E}\left[\int\left([K_{h_o}(X_1-.)g]\star K_h(u)\right)^2\right]du\\&\le\frac{1}{n}\mathbb{E}\left[\|K_{h_o}(X_1-.)g\|_1^2\|K_h\|^2\right]\le\frac{\|g\|_\infty^2\|K\|^2\|K\|_1^2}{nh}:=H_3.\end{aligned}$$

Next we have

$$\begin{aligned}\sup_{x,t}|\psi_t(x)|&\le\sup_x\|[(K_{h_o}(x-\cdot)-f_{h_o})g]\star K_h\|^2\le\sup_x\|[(K_{h_o}(x-\cdot)-f_{h_o})g]\|_1^2\|K_h\|^2\\&\le\sup_x\left(\int|K_{h_o}(x-u)-f_{h_o}(u)|\,g(u)du\right)^2\frac{\|K\|^2}{h}\le(2\|K\|_1\|g\|_\infty)^2\frac{\|K\|^2}{h}\end{aligned}$$

so that $b_3=2\|K\|_1\|g\|_\infty\|K\|/\sqrt{h}$. Lastly

$$\begin{aligned}\sup_t\mathbb{E}\left[\langle K_{h_o}(X_1-\cdot)g,t\star K_h\rangle^2\right]&=\sup_t\int\left(\int K_{h_o}(z-u)g(u)t\star K_h(u)du\right)^2f(z)dz\\&\le\sup_t\|f\|_\infty\left\|K_{h_o}\star[g(t\star K_h)]\right\|^2\\&\le\sup_t\|f\|_\infty\|K_{h_o}\|_1^2\|g(t\star K_h)\|^2\le\sup_t\|f\|_\infty\|g\|_\infty^2\|K\|_1^2\|t\star K_h\|^2\\&\le\|f\|_\infty\|g\|_\infty^2\|K\|_1^4:=v_3.\end{aligned}$$

The conclusion follows as for $\nu_{n,1}$ and gathering the previous bounds implied inequality (34) and t Theorem 3.1. □

7.4 Proof of Theorem 3.2

Following the first step in Lacour *et al.* (2017), we write

$$\|\widetilde{(fg)}_{\widetilde{h}} - fg\|^2 \leq \|\widetilde{(fg)}_h - fg\|^2 + (\text{pen}(h) - \psi_n(h)) - (\text{pen}(\widetilde{h}) - \psi_n(\widetilde{h})) \tag{36}$$

with

$$\psi_n(h, h_{\min}) = \langle \widetilde{(fg)}_h - fg, \widetilde{(fg)}_{h_{\min}} - fg\rangle.$$

As in Comte and Marie (2021), we decompose ψ_n in

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

First,

$$\psi_{1,n}(h, , h_{\min}) := \frac{\langle K_h, K_{h_{\min}}\rangle}{n^2}\sum_{i=1}^n \widetilde{f}^2_{h_o}(Y_i) + \frac{U(h, h_{\min})}{n^2} = \text{pen}_1(h) + \frac{U(h, h_{\min})}{n^2},$$

where

$$U_n(h, h') := \sum_{1\leq i\neq j\leq n} \langle \widetilde{f}_{h_o}(Y_i)K_h(Y_i - \cdot) - (fg)_h, \widetilde{f}_{h_0}(Y_j)K_{h'}(Y_j - \cdot) - (fg)_{h'}\rangle. \tag{37}$$

Indeed,

$$\frac{\langle K_h, K_{h_{\min}}\rangle}{n^2}\sum_{i=1}^n \widetilde{f}^2_{h_o}(Y_i) = \frac{1}{n^2}\sum_{i=1}^n \langle \widetilde{f}_{h_o}(Y_i)K_h(Y_i - \cdot), \widetilde{f}_{h_o}(Y_i)K_{h_{\min}}(Y_i - \cdot)\rangle.$$

Second,

$$\begin{aligned}\psi_{2,n}(h, h_{\min}) \quad &:= \quad -\frac{1}{n^2}\left(\sum_{i=1}^n \langle \widetilde{f}_{h_o}(Y_i)K_{h_{\min}}(Y_i - \cdot), (fg)_h\rangle + \sum_{i=1}^n \langle \widetilde{f}_{h_o}(Y_i)K_h(Y_i - \cdot), (fg)_{h_{\min}}\rangle\right) \\ &\qquad + \frac{1}{n}\langle (fg)_h, (fg)_{h_{\min}}\rangle\end{aligned} \tag{38}$$

and lastly

$$\psi_{3,n}(h, h_{\min}) := V_n(h, h_{\min}) + V_n(h_{\min}, h) + \langle (fg)_h - fg, (fg)_{h_{\min}} - fg\rangle$$

with

$$V_n(h, h') := \langle \widetilde{(fg)}_h - (fg)_h, (fg)_{h'} - fg\rangle.$$

We state a series of Lemmas that permit to establish Theorem 3.2.

Lemma 7.1. *Under the assumptions of Theorem 3.2,* $\mathbb{E}\left(\sup_{h,h'\in\mathcal{H}_n} |\psi_{2,n}(h, h')|\right) \leq C/n$*, where* $C = C(f, g, K)$ *is a positive constant depending on* f*,* g*,* K*.*

Lemma 7.2. *Under the assumptions of Theorem 3.2, for every* $\vartheta \in (0, 1)$*, it holds*

$$\mathbb{E}\left(\sup_{h,h'}\{|V_n(h, h')| - \vartheta\|(fg)_{h'} - fg\|^2\}\right) \leq \frac{1}{2\vartheta}\|K\|_1^2\|g\|_\infty\|f_{h_0} - f\|^2 + C\frac{\log(n)}{n}.$$

Lemma 7.3. *Under the assumptions of Theorem 3.2, for every* $\vartheta \in [0, 1]$*, it holds*

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\left\{\frac{|U_n(h, h_{\min})|}{n^2} - \vartheta\frac{c_0(f, g, K)}{nh}\right\}\right) \leq \vartheta\|f_{h_o} - f\|^2 + \frac{C\log(n)}{n}.$$

We deduce from (36) that

$$\begin{aligned}\|\widetilde{(fg)}_{\widetilde{h}} - fg\|^2 \quad &\leq \quad \|\widetilde{(fg)}_h - fg\|^2 + 2\left(\text{pen}_1(h) - \psi_n(h)\right) + 2\text{pen}_2(h) \\ &\qquad -2(\text{pen}_1(\widetilde{h}) - \psi_n(\widetilde{h})) - 2\text{pen}_2(\widetilde{h}).\end{aligned} \tag{39}$$

We have for all positive h

$$\psi_n(h) - \mathrm{pen}_1(h) = \frac{U_n(h,h_{\min})}{n^2} + \psi_{2,n}(h,h_{\min}) + V_n(h,h_{\min}) + V_n(h_{\min},h) + \langle (fg)_h - fg, (fg)_{h_{\min}} - fg\rangle.$$

We note that for all positive θ

$$|\langle (fg)_h - fg, (fg)_{h_{\min}} - fg\rangle| \leq \frac{\theta}{2}\|(fg)_h - fg\|^2 + \frac{1}{2\theta}\|(fg)_{h_{\min}} - fg\|^2.$$

Applying these for $h = \widetilde{h}$ we get

$$\begin{aligned}
&\mathbb{E}\left(\left|\psi_n(\widetilde{h}) - \mathrm{pen}_1(\widetilde{h})\right| - \mathrm{pen}_2(\widetilde{h})\right)\\
\leq\ & \mathbb{E}\left(\left|\frac{U_n(\widetilde{h},h_{\min})}{n^2}\right| - \theta\frac{c_0(f,g,K)}{n\widetilde{h}}\right)\\
&+\mathbb{E}\left(|V_n(\widetilde{h},h_{\min})| - \frac{\theta}{2}\|(fg)_{h_{\min}} - fg\|^2\right) + \mathbb{E}\left(|V_n(h_{\min},\widetilde{h})| - \frac{\theta}{2}\|(fg)_{\widetilde{h}} - fg\|^2\right)\\
&+\mathbb{E}\left(\theta\|(fg)_{\widetilde{h}} - fg\|^2 + (\theta-\kappa)\frac{c_0(f,g,K)}{n\widetilde{h}}\right) + \frac{1}{2}(\theta + \frac{1}{\theta})\|(fg)_{h_{\min}} - fg\|^2 + \frac{C}{n}.
\end{aligned}$$

where we used Lemma 7.1. Now, using Lemmas 7.2 and 7.3, we get

$$\begin{aligned}
\mathbb{E}\left(\left|\psi_n(\widetilde{h}) - \mathrm{pen}_1(\widetilde{h})\right| - \mathrm{pen}_2(\widetilde{h})\right) \leq\ & \frac{1}{2}(\theta + \frac{1}{\theta})\|(fg)_{h_{\min}} - fg\|^2 + c_1(f,g,K,\theta)\|f_{h_o} - f\|^2\\
&+\theta\mathbb{E}\left(\|(fg)_{\widetilde{h}} - fg\|^2 + (1-\frac{\kappa}{\theta})\frac{c_0(f,g,K)}{n\widetilde{h}}\right) + C\frac{\log(n)}{n}.
\end{aligned}$$

Observe that $\|\widetilde{(fg)}_h - fg\|^2 = \|\widetilde{(fg)}_h - (fg)_h\|^2 + \|(fg)_h - fg\|^2 + 2V_n(h,h)$. It follows that for all $\theta \in (0, \frac{1}{2})$,

$$\begin{aligned}
&(1-2\theta)\left(\|(fg)_h - fg\|^2 + (1-\frac{\kappa}{\theta})\frac{c_0(f,g,K)}{nh}\right) - \|\widetilde{(fg)}_h - fg\|^2\\
&\quad = -2\left(V_n(h,h) + \theta\|(fg)_h - fg\|^2\right) + (1-2\theta)(1-\frac{\kappa}{\theta})\frac{c_0(f,g,K)}{nh} - \|\widetilde{(fg)}_h - (fg)_h\|^2\\
&\quad \leq -2\left(V_n(h,h) + \theta\|(fg)_h - fg\|^2\right) \leq 2(|V_n(h,h)| - \theta\|(fg)_h - fg\|^2)
\end{aligned}$$

provided that $1 - \kappa/\theta \leq 0$. Therefore we choose $\kappa \geq \theta$ and apply Lemma 7.2 again. We obtain

$$\begin{aligned}
\mathbb{E}\left(\left|\psi_n(\widetilde{h}) - \mathrm{pen}_1(\widetilde{h})\right| - \mathrm{pen}_2(\widetilde{h})\right) \leq\ & \frac{1}{2}(\theta + \frac{1}{\theta})\|(fg)_{h_{\min}} - fg\|^2 + c_2(f,g,K)\|f_{h_o} - f\|^2\\
&+\frac{\theta}{1-2\theta}\mathbb{E}\left(\|\widetilde{(fg)}_{\widetilde{h}} - fg\|^2\right) + C\frac{\log(n)}{n}.
\end{aligned}$$

Similarly, we get

$$\begin{aligned}
\mathbb{E}\left(|\psi_n(h) - \mathrm{pen}_1(h)| + \mathrm{pen}_2(h)\right) \leq\ & \frac{1}{2}(\theta + \frac{1}{\theta})\|(fg)_{h_{\min}} - fg\|^2 + c_2(f,g,K)\|f_{h_o} - f\|^2\\
&+\theta\left(\|(fg)_h - fg\|^2 + (1+\frac{\kappa}{\theta})\frac{c_0(f,g,K)}{nh}\right) + C\frac{\log(n)}{n}.
\end{aligned}$$

Plugging the last two bounds in the expectation of (39) implies

$$\begin{aligned}
\left(1 - \frac{2\theta}{1-2\theta}\right)\mathbb{E}\left(\|\widetilde{(fg)}_{\widetilde{h}} - fg\|^2\right) \leq\ & \mathbb{E}\left(\|\widetilde{(fg)}_h - fg\|^2\right) + 2\theta\left(\|(fg)_h - fg\|^2 + (1+\frac{\kappa}{\theta})\frac{c_0(f,g,K)}{nh}\right)\\
&+2(\theta + \frac{1}{\theta})\|(fg)_{h_{\min}} - fg\|^2 + 4c_2(f,g,K)\|f_{h_o} - f\|^2 + C\frac{\log(n)}{n}.
\end{aligned}$$

Now applying Proposition 3.1 with a rougher bound on the variance (the constant is larger), we get for $\theta \in (0,1/4)$, and $\kappa \geq \theta$,

$$\left(1-\frac{2\theta}{1-2\theta}\right)\mathbb{E}\left(\|\widetilde{(fg)}_{\widetilde{h}}-fg\|^2\right) \leq 2(1+\theta)\|(fg)_h-fg\|^2+2(1+\theta+\kappa)\frac{c_0(f,g,K)}{nh}$$
$$+2(\theta+\frac{1}{\theta})\|(fg)_{h_{\min}}-fg\|^2+4c_2(f,g,K)\|f_{h_o}-f\|^2+C\frac{\log(n)}{n}.$$

We conclude that for $\kappa \geq 1/4$ and all $\theta \in (0,1/4)$ the result given in Theorem 3.2 holds true. □

7.4.1 Proof of Lemma 7.1

For the study of $\psi_{2,n}$, note that $|\langle (fg)_h,(fg)_{h'}\rangle| \leq \|fg\star K_h\|\|fg\star K_{h'}\| \leq \|K\|_1^2\|fg\|^2$ and observe that

$$\mathbb{E}(|\widetilde{f}_{h_o}(Y_1)|) \leq \sqrt{\mathbb{E}(\widetilde{f}^2_{h_o}(Y_1))} \leq \sqrt{\|g\|_\infty(\|K\|^2+\|K\|_1^2\|f\|^2)}.$$

As a consequence, for all positive h,h',

$$\mathbb{E}\left(\sup_{h,h'\in\mathcal{H}_n}\left|\frac{1}{n}\sum_{i=1}^n\langle\widetilde{f}_{h_o}(Y_i)K_h(Y_i-\cdot),(fg)_{h'}\rangle\right|\right) \leq \mathbb{E}\left(|\widetilde{f}_{h_o}(Y_1)|\sup_{h,h'}|\langle K_h(Y_1-\cdot),(fg)_{h'}\rangle|\right)$$
$$\leq \mathbb{E}(|\widetilde{f}_{h_o}(Y_1)|)\|K\|_1^2\|fg\|_\infty$$
$$\leq \sqrt{\|g\|_\infty(\|K\|^2+\|K\|_1^2\|f\|^2)}\|K\|_1^2\|fg\|_\infty.$$

From the definition of $\psi_{2,n}$ given by (38), il follows that the result of Lemma 7.1 holds with $C=(2??\sqrt{\|g\|_\infty(\|K\|^2+\|K\|_1^2\|f\|^2)}+\|g\|_\infty)\|fg\|_\infty\|K\|_1^2$. □

7.4.2 Proof of Lemma 7.2 and study of $V_n(h,h')$

We decompose $V_n(h,h')=V_{n,1}(h,h')+V_{n,2}(h,h')+V_{n,3}(h,h')$ with

$$V_{n,1}(h,h') = \langle\widetilde{(fg)}_h-(\widetilde{f}_{h_o}g)_h,(fg)_{h'}-fg\rangle,$$
$$V_{n,2}(h,h') = \langle(\widetilde{f}_{h_o}g)_h-(f_{h_o}g)_h,(fg)_{h'}-fg\rangle,$$
$$V_{n,3}(h,h') = \langle(f_{h_o}g)_h-(fg)_h,(fg)_{h'}-fg\rangle.$$

We have for all positive θ

$$|V_{n,3}(h,h')|=|\langle((f_{h_o}g))_h-(fg)_h,(fg)_{h'}-fg\rangle| \leq \frac{\vartheta}{2}\|(fg)_{h'}-fg\|^2+\frac{1}{2\vartheta}\|K\|_1^2\|g\|_\infty^2\|f_{h_0}-f\|^2,$$

implying that

$$\mathbb{E}\left(\sup_{h,h'}\left\{|V_{n,3}(h,h')|-\frac{\vartheta}{2}\|(fg)_{h'}-fg\|^2\right\}\right) \leq \frac{1}{2\vartheta}\|K\|_1^2\|g\|_\infty\|f_{h_0}-f\|^2. \quad (40)$$

Next, we write

$$V_{n,2}(h,h')=\frac{1}{n}\sum_{i=1}^n(Z_i-\mathbb{E}(Z_i)), \quad Z_i-\mathbb{E}(Z_i):=\langle[(K_{h_o}(X_i-\cdot)-f_{h_o}(\cdot)g(\cdot)]\star K_h,(fg)_{h'}-fg\rangle$$

and apply Bernstein Inequality. Using that K is even, the variance bound is obtained by

$$\mathrm{Var}Z_1 \leq \mathbb{E}(Z_1^2)=\int\langle[(K_{h_o}(u-\cdot)\,g(\cdot)]\star K_h,(fg)_{h'}-fg\rangle^2f(u)du$$
$$=\int[K_{h_o}\star(g\,[K_h\star((fg)_{h'}-fg)])(u)]^2\,f(u)du$$
$$\leq \|f\|_\infty\|K_{h_o}\star(gK_h\star((fg)_{h'}-fg))\|^2 \leq \|f\|_\infty\|K_{h_o}\|_1^2\|(gK_h\star((fg)_{h'}-fg))\|^2$$
$$\leq \|f\|_\infty\|g\|_\infty^2\|K\|_1^4\|(fg)_{h'}-fg)\|^2:=v^2_{h,h'}.$$

On the other hand, we have

$$\begin{aligned}
|Z_1| &\leq \sup_z \left| \int K_{h_o}(z-x) g(x) ((fg)_{h'} - fg) \star K_h)(x) dx \right| \\
&\leq \|g\|_\infty \sup_x |((fg)_{h'} - fg) \star K_h)(x)| \int |K_{h_o}(z)| dz \\
&\leq \|g\|_\infty \|K\|_1 \sup_x |((fg)_{h'} - fg)(x)| \int |K_h(u)| du \\
&\leq \|g\|_\infty \|K\|_1^2 (1 + \|K\|_1) \|fg\|_\infty := b_{h,h'}.
\end{aligned}$$

Therefore, Bernstein Inequality (30) implies that with probability larger than $1 - 2e^{-\lambda}$

$$\begin{aligned}
|V_{n,2}(h,h')| &\leq \sqrt{\frac{2\lambda v_{h,h'}^2}{n}} + \frac{\lambda}{n} b_{h,h'} \\
&\leq \frac{\theta}{4} \|(fg)_{h'} - fg)\|^2 + \frac{\lambda}{n} \|f\|_\infty \|g\|_\infty^2 \|K\|_1^2 (2 \frac{\|K\|_1^2}{\theta} + 1 + \|K\|_1).
\end{aligned}$$

This together with (31) and $|\mathcal{H}_n| = n$ leads to

$$\mathbb{E}\left(\sup_{h,h'} \left\{ |V_{n,2}(h,h')| - \frac{\vartheta}{4} \|(fg)_{h'} - fg\|^2 \right\} \right) \leq C \frac{\log(n)}{n}. \tag{41}$$

For $V_{n,1}$, we write $V_{n,1}(h,h') = V_{n,1}^b(h,h') + V_{n,1}^c(h,h')$ with

$$V_{n,1}^b(h,h') = \frac{1}{n} \sum_{i=1}^n \int \left(\widetilde{f}_{h_o}^b(Y_i) K_h(Y_i - x) - (\widetilde{f}_{h_o}^b g) \star K_h(x) \right) ((fg)_{h'} - fg)(x) dx$$

where

$$\widetilde{f}_{h_0}^b(x) = \widetilde{f}_{h_0}(x) \mathbf{1}_{|\widetilde{f}_{h_o}(x)| \leq c_0} \quad c_0 = 2\|f\|_\infty \|K\|_1.$$

We apply Bernstein inequality conditionally to $\mathbf{X}$, with $b_{h,h'} = \|f\|_\infty \|fg\|_\infty \|K\|_1^2 (1 + \|K\|_1)$ and $v_{h,h'}^2 = \|f\|_\infty^2 \|g\|_\infty \|K\|_1^4 \|(fg)_{h'} - fg\|^2$. The orders of $b_{h,h'}$ and $v_{h,h'}$ being the same as for $V_{n,2}$ and independent of $\mathbf{X}$, the result for $V_{n,1}^b$ is :

$$\mathbb{E}\left(\sup_{h,h'} \left\{ |V_{n,1}^b(h,h')| - \frac{\vartheta}{4} \|(fg)_{h'} - fg\|^2 \right\} \right) \leq C \frac{\log(n)}{n}. \tag{42}$$

Lastly, by noticing that $|\widetilde{f}_{h_o}(x)| \leq |\widetilde{f}_{h_o}(x) - f_{h_o}(x)| + \|f\|_\infty \|K\|_1$, we get

$$\mathbb{P}(|\widetilde{f}_{h_o}(x)| > c_0) \leq \mathbb{P}(|\widetilde{f}_{h_o}(x) - f_{h_o}(x)| > \|f\|_\infty \|K\|_1)$$

and by Bernstein Inequality

$$\mathbb{P}(|\widetilde{f}_{h_o}(x)| > c_0) \leq \exp\left(-n h_o \frac{c_0^2}{2(\|f\|_\infty \|K\|^2 + \|K\|_\infty c_0)} \right) \leq 2n^{-p} \tag{43}$$

for $nh_0 \geq c_1 \log(n)$ and $c_1 \geq 3p\|K\|_\infty / (2\|f\|_\infty \|K\|_1)$. Then as $\|(fg)_h - fg\|_\infty \leq \|fg\|_\infty (\|K\|_1 + 1)$, we get that

$$\begin{aligned}
\mathbb{E}\left(\sup_{h,h' \in \mathcal{H}_n} |V_{n,1}^c(h,h')| \right) &\leq \sum_{h,h' \in \mathcal{H}_n} 2 \frac{\|K\|_\infty}{h_o} \|fg\|_\infty (\|K\|_1 + 1) \int \mathbb{P}(|\widetilde{f}_{h_o}(x)| > c_0) g(x) dx \\
&\leq \frac{C(f,g,K) |\mathcal{H}_n|^2}{n^p h_o} \leq \frac{C}{n}.
\end{aligned}$$

Note that the last bound is obtained using $|\mathcal{H}_n| \leq n$, $1/h_o \leq n$ and $p = 4$, which holds under assumption [B3](4). Gathering this with (40), (41), (42) gives the result of Lemma 7.2. □

7.4.3 Proof of Lemma 7.3 and study of $U_n(h,h')$

Warning. For the study of this term, in order to avoid burdensome technicalities, we assume that $\widetilde{f}_{h_0}$ is bounded by $2\|K\|_1\|f\|_\infty$. We proved in the study of V_n (see (43)) that that the probability of the complement is $1/n^4$ under [B3](4).

Recall that $U_n(h,h')$ is defined by (37). We write that

$$\begin{aligned}
& \widetilde{f}_{h_o}(Y_i)K_h(Y_i-\cdot)-(fg)_h \\
= \; & \underbrace{\widetilde{f}_{h_o}(Y_i)K_h(Y_i-\cdot)-(\widetilde{f}_{h_o}g)\star K_h}_{(1)_h}+\underbrace{(\widetilde{f}_{h_o}g)\star K_h-(f_{h_o}g)\star K_h}_{(2)_h}+\underbrace{(f_{h_o}g)\star K_h-(fg)_h}_{(3)_h}
\end{aligned}$$

so that $U_n(h,h_{\min})$ can be splitted into 9 terms, denoted with obvious super-indices (k,ℓ) for $k,\ell\in\{1,2,3\}$. These 9 terms can be reduced to 6 by symmetry arguments, denoted by $U_n^{(i),(j)}(h,h_{\min})$ for $i\le j\in\{1,2,3\}$.

• **Treatment of** $U_n^{(1),(1)}(h,h_{\min})$ we have, by analogy with Lemma 6.2 in Comte and Marie (2021), that, for every $\vartheta\in[0,1]$,

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}_n}\left\{\frac{|U_n^{(1),(1)}|(h,h_{\min})}{n^2}-\frac{\vartheta\|K\|^2\|K\|_1^2\|f\|_\infty^2}{nh}\right\}|\mathbf{X}\right)\le\frac{C\log(n)}{n} \tag{44}$$

and it is easy to see that all bounds do note depend on $\mathbf{X}$ so de-conditioning is straightforward.

• **Treatment of** $U_n^{(3),(3)}(h,h_{\min})$ it is easy to handle thanks to the equality

$$U_n^{(3),(3)}(h,h_{\min})=n(n-1)\langle(f_{h_o}g)\star K_h-(fg)_h,(f_{h_o}g)\star K_{h_{\min}}-(fg)_{h_{\min}}\rangle$$

leading to the bound

$$\frac{|U_n^{(3),(3)}(h,h_{\min})}{n^2}|\le\|[(f_{h_o}-f)g]\star K_h\|\|[(f_{h_o}-f)g]\star K_{h_{\min}}\|\le\|K\|_1^2\|g\|_\infty^2\|f_{h_o}-f\|^2. \tag{45}$$

• **Treatment of** $U_n^{(2),(3)}(h,h_{\min})$ first note that $U_n^{(2),(3)}(h,h_{\min})/n^2=[(n-1)/n]\langle(2)_h,(3)_{h_{\min}}\rangle$ where $\langle(2)_h,(3)_{h_{\min}}\rangle$ is equal to

$$\begin{aligned}
\langle[(\widetilde{f}_{h_o}g)-(f_{h_o}g)]\star K_h,((f_{h_o}-f)g)\star K_{h_{\min}}\rangle &= \langle[(\widetilde{f}_{h_o}g)-(f_{h_o}g)],((f_{h_o}-f)g)\star K_{h_{\min}}\star K_h\rangle\\
&= \langle[(\widetilde{f}_{h_o}g)-(f_{h_o}g)],((f_{h_o}-f)g)\star K_h\star K_{h_{\min}}\rangle\\
&= \langle[(\widetilde{f}_{h_o}g)-(f_{h_o}g)]\star K_{h_{\min}},((f_{h_o}-f)g)\star K_h\rangle
\end{aligned}$$

and thus $\langle(2)_h,(3)_{h_{\min}}\rangle=\langle(2)_{h_{\min}},(3)_h\rangle$, so that $U_n^{(2),(3)}(h,h_{\min})=U_n^{(3),(2)}(h,h_{\min})$. Now, the process can be written as

$$\frac{1}{n}\sum_{i=1}^n(Z_i^{2,3}-\mathbb{E}(Z_i^{2,3})),\quad Z_i^{2,3}:=\langle K_{h_o}(X_i-\cdot)g,[(f_{h_o}-f)g]\star K_h\star K_{h_{\min}}\rangle.$$

To apply Bernstein Inequality, we need to bound the variance and infinite norm of the $Z_i^{2,3}$'s. For the moment of order 2, we have

$$\begin{aligned}
\mathbb{E}[(Z_1^{2,3})^2] &= \int f(x)\left(\int K_{h_o}(x-u)g(u)[(f-f_{h_o})g]\star K_h\star K_{h_{\min}}(u)du\right)^2dx\\
&\le \|f\|_\infty\|K_{h_o}\star[g[(f_{h_o}-f)g]\star K_h\star K_{h_{\min}}]\|^2\\
&\le \|f\|_\infty\|K_{h_o}\|_1^2\|[g[(f_{h_o}-f)g]\star K_h\star K_{h_{\min}}]\|^2\\
&\le \|f\|_\infty\|g\|_\infty^2\|K\|_1^6\|f_{h_o}-f\|^2:=\mathfrak{v}.
\end{aligned}$$

For the upper bound, it holds:

$$\begin{aligned}
& \sup_x \left| \int K_{h_o}(x-u)g(u)[(f-f_{h_o})g] \star K_h \star K_{h_{\min}}(u)du \right| \\
\leq\ & \|g\|_\infty \sup_u |[(f-f_{h_o})g] \star K_h \star K_{h_{\min}}(u)| \sup_x \int |K_{h_o}(x-u)|du \\
\leq\ & \|g\|_\infty \|K\|_1 \sup_v ||[(f-f_{h_o})g] \star K_h(v)| \sup_u \int |K_{h_{\min}}(u)|du \\
\leq\ & \|g\|_\infty^2 \|K\|_1^3 (\|f\|_\infty + \sup_u |f \star K_{h_o}(u)|) \leq \|g\|_\infty^2 \|K\|_1^3 (1+\|K\|_1)\|f\|_\infty := \mathfrak{b}.
\end{aligned}$$

Then Bernstein Inequality implies that with probability larger than $1-2e^{-\lambda}$, $\lambda > 0$, for any $\vartheta \in (0,1)$,

$$|U_n^{(2),(3)}(h,h_{\min})|/n^2 \leq |\langle (2)_h, (3)_{h_{\min}} \rangle| \leq \sqrt{\frac{2\mathfrak{v}\lambda}{n}} + \frac{\lambda}{n}\mathfrak{b} \leq \vartheta \|f-f_{h_o}\|^2 + C(K,f,g)\frac{\lambda}{\vartheta n}.$$

As a consequence, we obtain

$$\mathbb{P}\left(\sup_{h,h'} \left(\left| \frac{U^{(2),(3)}(h,h')}{n^2} \right| - \vartheta \|f-f_{h_o}\|^2 \right) \geq C(K,f,g)\frac{\lambda}{\vartheta n} \right) \leq 2|\mathcal{H}_n|^2 e^{-\lambda}.$$

Then it follows from (31) and $|\mathcal{H}_n| = n$ that

$$\mathbb{E}\left(\sup_{h,h'} \left| \frac{U^{(2),(3)}(h,h')}{n^2} \right| - \vartheta \|f-f_{h_o}\|^2 \right)_+ \leq \frac{C' \log(n)}{n}. \tag{46}$$

• **Treatment of** $U_n^{(1),(3)}(h,h_{\min})$ write that $U_n^{(1),(3)}(h,h_{\min})/n^2 = [(n-1)/n]\langle (1)_h, (3)_{h_{\min}} \rangle$ and $\langle (1)_h, (3)_{h_{\min}} \rangle$ is

$$\frac{1}{n}\sum_{i=1}^n \langle \widetilde{f}_{h_o}(Y_i)K_h(Y_i - \cdot) - (\widetilde{f}_{h_o} g) \star K_h, [(f_{h_o}-f)g] \star K_{h_{\min}} \rangle.$$

We apply Bernstein Inequality conditionally to $\mathbf{X}$, recalling that we consider $\widetilde{f}_{h_o}$ bounded by $2\|f\|_\infty \|K\|_1$.

$$\begin{aligned}
& \mathbb{E}\left(\langle \widetilde{f}_{h_o}(Y_i)K_h(Y_i-\cdot), [(f_{h_o}-f)g] \star K_{h_{\min}} \rangle^2 |\mathbf{X} \right) \\
& = \int \left(\int \widetilde{f}_{h_o}(y)K_h(y-u)[(f_{h_o}-f)g] \star K_{h_{\min}}(u)du \right)^2 g(y)dy \\
\leq\ & 4\|K\|_1^2 \|f\|_\infty^2 \|g\|_\infty \, \||K_h| \star |(f_{h_o}-f)g| \star |K_{h_{\min}}|\|^2 \leq 4\|K\|_1^6 \|f\|_\infty^2 \|g\|_\infty^3 \|f_{h_o}-f\|^2,
\end{aligned}$$

by iterative application of Young Inequality. Next for the infinite norm

$$\begin{aligned}
& \sup_y |\langle \widetilde{f}_{h_o}(y)K_h(y-\cdot), [(f_{h_o}-f)g] \star K_{h_{\min}} \rangle| \\
\leq\ & 2\|f\|_\infty \|K\|_1 \sup_y \int |K_h(y-u)||[(f_{h_o}-f)g] \star K_{h_{\min}}(u)|du \\
\leq\ & 2\|f\|_\infty \|K\|_1 \int |K_h(v)|dv \sup_u |[(f_{h_o}-f)g] \star K_{h_{\min}}(u)| \\
\leq\ & 2\|f\|_\infty \|K\|_1^3 \sup_z |(f_{h_o}-f)(z)g(z)| \leq 2\|f\|_\infty^2 \|g\|_\infty \|K\|_1^3 (1+\|K\|_1).
\end{aligned}$$

The bounds do not depend on $\mathbf{X}$, it holds:

$$\mathbb{E}\left(\sup_{h,h'} \left| \frac{U^{(1),(3)}(h,h')}{n^2} \right| - \vartheta \|f-f_{h_o}\|^2 \right)_+ \leq C\frac{\log(n)}{n} \tag{47}$$

for a constant $C >$ depending on f, g, K. Moreover, the bounds do not depend on $h, h_{\min}$ so the same bound hold for $U_n^{(3),(1)}(h, h_{\min})/n^2$.

• **Treatment of** $U_n^{(1),(2)}(h, h_{\min})$ write $U_n^{(1),(2)}(h, h_{\min})/n^2 = [(n-1)/n]\langle (1)_h, (2)_{h_{\min}}\rangle$ where $\langle (1)_h, (2)_{h_{\min}}\rangle$ is

$$\frac{1}{n}\sum_{i=1}^{n}\langle \widetilde{f}_{h_o}(Y_i)K_h(Y_i-\cdot) - (\widetilde{f}_{h_o}g)\star K_h, [(\widetilde{f}_{h_o}-f_{h_o})g]\star K_{h_{\min}}\rangle.$$

First, we apply a Bernstein Inequality conditionally to $\mathbf{X}$. The variance term is:

$$\mathbb{E}\left(\langle \widetilde{f}_{h_o}(Y_i)K_h(Y_i-\cdot), [(\widetilde{f}_{h_o}-f_{h_o})g]\star K_{h_{\min}}\rangle^2\right) \leq 4\|f\|_\infty\|g\|_\infty\|K\|_1^4\|K_h\star[(\widetilde{f}_{h_o}-f_{h_o})g]\|^2 := \mathfrak{v}.$$

For the upper bound, we get

$$\sup_y |\langle \widetilde{f}_{h_o}(y)K_h(y-\cdot), [(\widetilde{f}_{h_o}-f_{h_o})g]\star K_{h_{\min}}\rangle| \leq 6\|K\|_1^4\|f\|_\infty^2\|g\|_\infty := \mathfrak{b}$$

with usual tricks. Now, we can notice that

$$\mathbb{E}\left(\|K_h\star[(\widetilde{f}_{h_o}-f_{h_o})g]\|^2\right) \leq \frac{\|g\|_\infty\|K\|_1^2\|K\|^2}{nh}. \tag{48}$$

So we write

$$\begin{aligned}
&\mathbb{E}\left[\sup_{h\in\mathcal{H}_n}\left(\frac{U_n^{(1),(2)}(h,h_{\min})}{n^2} - 4\vartheta\frac{\|g\|_\infty\|K\|_1^2\|K\|^2}{nh}\right)\right]\\
\leq\ &\mathbb{E}\left\{\mathbb{E}\left[\sup_{h\in\mathcal{H}_n}\left(\frac{U_n^{(1),(2)}(h,h_{\min})}{n^2} - \vartheta\|K_h\star[(\widetilde{f}_{h_o}-f_{h_o})g]\|^2\right)|\mathbf{X}\right]\right\}\\
&+\vartheta\mathbb{E}\left[\sup_{h\in\mathcal{H}_n}\left(\|K_h\star[(\widetilde{f}_{h_o}-f_{h_o})g]\|^2 - 4\frac{\|g\|_\infty\|K\|_1^2\|K\|^2}{nh}\right)\right]
\end{aligned}$$

The first term is bounded by taking the expectation of the conditional Bernstein, where constants are independent of the X_i, which writes with the terms $\mathfrak{b}, \mathfrak{v}$:

$$\mathbb{E}\left[\sup_{h\in\mathcal{H}_n}\left(\frac{U_n^{(1),(2)}(h,h_{\min})}{n^2} - \vartheta\|K_h\star[(\widetilde{f}_{h_o}-f_{h_o})g]\|^2\right)\right] \leq C\frac{\log(n)}{n}. \tag{49}$$

For the second, we use Talagrand Inequality, relying on the linear process

$$\nu_n(t) = \langle K_h\star[(\widetilde{f}_{h_o}-f_{h_o})g], t\rangle$$

which fulfills $\sup_{t\in\mathcal{B}(0,1)}\nu_n^2(t) = \|K_h\star[(\widetilde{f}_{h_o}-f_{h_o})g]\|^2$ where $\mathcal{B}(0,1)$ is a countable dense subset of $\{t\in\mathbb{L}^2(\mathbb{R}), \|t\|^2 = 1\}$. To apply Talagrand inequality, we compute H^2, v, b. We have from (48) that

$$H^2 = \frac{\|g\|_\infty\|K\|_1^2\|K\|^2}{nh}.$$

Then we compute v^2.

$$\begin{aligned}
\sup_{\|t\|=1}\mathrm{Var}\left(\iint K_h(x-u)K_{h_o}(X_1-u)t(x)dxdu\right) &\leq \sup_{\|t\|=1}\mathbb{E}\left[\left(\iint K_h(x-u)K_{h_o}(X_1-u)t(x)dxdu\right)^2\right]\\
&= \sup_{\|t\|=1}\mathbb{E}\left[(K_{h_o}\star K_h\star t(X_1))^2\right]\\
&\leq \|f\|_\infty\sup_{\|t\|=1}\|K_{h_o}\star K_h\star t\|^2 \leq \|f\|_\infty\|K\|_1^4 := v^2.
\end{aligned}$$

Next, for b, we find

$$\begin{aligned}\sup_{\|t\|=1}\sup_y |K_{h_o}\star K_h\star t(y)| &\le \sup_{\|t\|=1}\sup_y\left[\int t^2(x)dx\int (K_h\star K_{h_o}(y-x))^2dx\right]^{1/2} = \|K_h\star K_{h_o}\| \\ &\le \frac{\|K\|_1\|K\|}{\sqrt{h}} := b.\end{aligned}$$

Talagrang Inequality gives

$$\begin{aligned}&\mathbb{E}\left[\sup_{h\in\mathcal{H}_n}\left(\|K_h\star[(\widetilde{f}_{h_o}-f_{h_o})g]\|^2 - 4\frac{\|g\|_\infty\|K\|_1^2\|K\|^2}{nh}\right)_+\right] \\ &\le \frac{C}{n}\left(\sum_{h\in\mathcal{H}_n}\exp(-c_1/h)+\text{card}(\mathcal{H}_n)\exp(-C_2\sqrt{n})\right).\end{aligned}$$

As a consequence, as under [B4], $\sum_{h\in\mathcal{H}_n}\exp(-c_1/h)\le\Sigma<+\infty$ and $\text{card}(\mathcal{H}_n)\le n$, we get

$$\mathbb{E}\left[\sup_{h\in\mathcal{H}_n}\left(\frac{U_n^{(1),(2)}(h,h_{\min})}{n^2} - 4\vartheta\frac{\|g\|_\infty\|K\|_1^2\|K\|^2}{nh}\right)\right]\le C\frac{\log(n)}{n}. \tag{50}$$

• **Treatment of** $U_n^{(2),(2)}(h,h_{\min})$ write $U_n^{(2),(2)}(h,h_{\min})/n^2 = [(n-1)/n]\langle(2)_h,(2)_{h_{\min}}\rangle$ where $\langle(2)_h,(2)_{h_{\min}}\rangle$ is

$$\langle[(\widetilde{f}_{h_o}-f_{h_o})g]\star K_h, [(\widetilde{f}_{h_o}-f_{h_o})g]\star K_{h_{\min}}\rangle.$$

The decomposition of this term involves first a U-statistics related to X:

$$\begin{aligned}\frac{U_n^{\mathbf{X}}(h,h_{\min})}{n^2} := \frac{1}{n^2}\sum_{1\le i\ne j\le n}\int &\left(\int (K_{h_o}(X_i-u)-f\star K_{h_o}(u))g(u)K_h(x-u)du\right)\\ &\times\left(\int (K_{h_o}(X_j-v)-f\star K_{h_o}(v))g(v)K_{h_{\min}}(x-v)dv\right)dx\end{aligned}$$

and terms corresponding to $i=j$ that are studied separately:

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n\int&\left(\int (K_{h_o}(X_i-u)-f\star K_{h_o}(u))g(u)K_h(x-u)du\right)\\ &\times\left(\int (K_{h_o}(X_i-v)-f\star K_{h_o}(v))g(v)K_{h_{\min}}(x-v)dv\right)dx.\end{aligned}$$

First, developing the latter product leads to the study of the four following terms. Two cross-terms that are are bounded by

$$\begin{aligned}&\frac{1}{n^2}\sum_{i=1}^n\left|\int f\star(K_{h_o}g)\star K_h\star K_{h_{\min}}(x)K_{h_o}(X_i-x)g(x)dx\right| \\ &\le \frac{\|g\|_\infty}{n}\sup_x|f\star(K_{h_o}g)\star K_h\star K_{h_{\min}}(x)|\int |K_{h_o}(z)|dz \\ &\le \frac{\|g\|_\infty\|K\|_1^2}{n}\sup_x|f\star(K_{h_o}g)\star K_h(x)| \le \frac{\|g\|_\infty^2\|K\|_1^3}{n}\sup_x|f\star K_{h_o}(x)| \\ &\le \frac{\|f\|_\infty\|g\|_\infty^2\|K\|_1^4}{n}.\end{aligned}$$

The product of last terms can be written

$$\begin{aligned}\frac{1}{n}\left|\int (f\star K_{h_o}g)\star K_h(x)(f\star K_{h_o}g)\star K_{h_{\min}}(x)dx\right| &\le \frac{1}{n}\|(f\star K_{h_o}g)\star K_h\|\|(f\star K_{h_o}g)\star K_{h_{\min}}\| \\ &\le \frac{1}{n}\|K_h\|_1\|K_{h_{\min}}\|_1\|g\|_\infty^2\|f\star K_{h_o}\|^2 \\ &\le \frac{\|K\|_1^4\|g\|_\infty^2\|f\|^2}{n}.\end{aligned}$$

Finally, the product of the first terms is

$$\begin{aligned}
& \left| \frac{1}{n^2} \sum_{i=1}^{n} \iiint K_{h_o}(X_i - u) K_{h_o}(X_i - v) K_h(x-u) K_{h_{\min}}(x-v) g(u) g(v) dudvdx \right| \\
\leq & \frac{\|g\|_\infty^2}{n^2} \sum_{i=1}^{n} \int |K_{h_o}| \star |K_h|(X_i - x) |K_{h_o}| \star |K_{h_{\min}}|(X_i - x) dx \\
= & \frac{\|g\|_\infty^2}{n} \int |K_{h_o}| \star |K_h|(z) |K_{h_o}| \star |K_{h_{\min}}|(z) dz \\
\leq & \frac{\|g\|_\infty^2}{n} \sup_z |K_{h_o}| \star |K_h|(z) \int |K_{h_o}| \star |K_{h_{\min}}|(z) dz \\
\leq & \frac{\|g\|_\infty^2 \|K\|_1^3 \|K\|_\infty}{nh},
\end{aligned}$$

implying that

$$\sup_h \left(\left| \frac{1}{n^2} \sum_{i=1}^{n} \iiint K_{h_o}(X_i - u) K_{h_o}(X_i - v) K_h(x-u) K_{h_{\min}}(x-v) g(u) g(v) dudvdx \right| - \frac{\|g\|_\infty^2 \|K\|_1^3 \|K\|_\infty}{nh} \right) \leq 0.$$

Let us deal with the U-statistics $U_n^{\mathbf{X}}(h, h_{\min})$. We follow the line of the proof of Lemma 6.2 in Comte and Marie (2021) and write $U_n^{\mathbf{X}}(h, h_{\min}) = \sum_{1 \leq i \neq j \leq n} G_{h,h_{\min}}(X_i, X_j)$ where

$$G_{h,h_{\min}}(X_i, X_j) = \langle [(K_{h_o}(X_i - \cdot) - f_{h_o})g] \star K_h, [(K_{h_o}(X_j - \cdot) - f_{h_o})g] \star K_{h_{\min}} \rangle.$$

Indeed, $G_{h,h_{\min}}(X_i, X_j) = G_{h_{\min},h}(X_i, X_j)$ as for all functions u, v it holds

$$\langle u \star K_h, v \star K_{h_{\min}} \rangle = \langle u, v \star K_{h_{\min}} \star K_h \rangle = \langle u, v \star K_{h_{\min}} \star K_h \rangle = \langle u \star K_{h_{\min}}, v \star K_h \rangle.$$

We apply the deviation inequality for U-statistics of order 2, as in Lacour *et al.* (2017), see Theorem 3.4 in Houdré and Reynaud-Bouret (2003). Following the notations of the aforementionned papers, we have to compute four bounds $\mathfrak{a}_n, \mathfrak{b}_n, \mathfrak{c}_n, \mathfrak{d}_n$.

⋄ First $\mathfrak{a}_n$ is a bound on $\sup_{z,z'} |G_{h,h_{\min}}(z, z')|$.

$$\begin{aligned}
\sup_{z,z'} |G_{h,h_{\min}}(z,z')| & \leq \sup_{z,z'} \left(\sup_x |[(K_{h_o}(z - \cdot) - f_{h_o})g] \star K_{h_{\min}}(x)| \int |[(K_{h_o}(X_i - \cdot) - f_{h_o})g] \star K_h(x)| dx \right) \\
& \leq \sup_{z,z'} \left(\|K_{h_{\min}}\|_\infty \|(K_{h_o}(z - \cdot) - f_{h_o})g\|_1 \int |[(K_{h_o}(X_i - \cdot) - f_{h_o})g](x)| dx \int |K_h(x)| dx \right) \\
& \leq 2 \frac{\|K\|_\infty}{h_{\min}} \|g\|_\infty \|K\|_1 \times 2 \|g\|_\infty \|K\|_1^2 = 4 \frac{\|g\|_\infty^2 \|K\|_1^3 \|K\|_\infty}{h_{\min}} := \mathfrak{a}_n.
\end{aligned}$$

Thus

$$\frac{\mathfrak{a}_n \lambda^2}{n^2} \leq 4\lambda^2 \frac{\|g\|_\infty^2 \|K\|_1^3 \|K\|_\infty}{n}.$$

⋄ Next $\mathfrak{b}_n^2$ is a bound on $n \sup_z \mathbb{E}[G_{h,h_{\min}}^2(z, X_1)]$, we write

$$\begin{aligned}
n \sup_z \mathbb{E}[G_{h,h_{\min}}^2(z, X_1)] & \leq n \sup_z \|[(K_{h_o}(z - \cdot) - f_{h_o})g] \star K_h\|^2 \mathbb{E}(\|[(K_{h_o}(X_1 - \cdot) - f_{h_o})g] \star K_{h_{\min}}\|^2) \\
& \leq \|K_h\|^2 \|K_{h_{\min}}\|^2 \sup_z \|[(K_{h_o}(z - \cdot) - f_{h_o})g]\|_1^2 \mathbb{E}(\|[(K_{h_o}(X_1 - \cdot) - f_{h_o})g]\|_1^2) \\
& \leq 4n \frac{\|K\|^4 \|K\|_1^4 \|g\|_\infty^2}{hh_{\min}} := \mathfrak{b}_n^2.
\end{aligned}$$

We obtain

$$\frac{\mathfrak{b}_n \lambda^{3/2}}{n^2} \leq 2\lambda^{3/2} \frac{\|K\|^2 \|K\|_1^2 \|g\|_\infty}{\sqrt{hh_{\min}} n^{3/2}} \leq \theta \frac{\|K\|^2 \|K\|_1^2 \|g\|_\infty^2}{nh} + \frac{\lambda^3}{\theta} \frac{\|K\|^2 \|K\|_1^2}{n^2 h_{\min}}.$$

⋄ We compute $\mathfrak{c}_n^2$ which is a bound on $n^2\mathbb{E}\left[G^2_{h,h_{\min}}(X_1,X_2)\right]$. Decompose

$$\mathbb{E}\left[G^2_{h,h_{\min}}(X_1,X_2)\right] = \mathbb{E}\left[\langle[(K_{h_o}(X_1\cdot)-f_{h_o})g]\star K_h,[(K_{h_o}(X_2-\cdot)-f_{h_o})g]\star K_{h_{\min}}\rangle^2\right]$$

into four squared terms. First,

$$\langle f_{h_o}g\star K_h, f_{h_o}g\star K_{h_{\min}}\rangle^2 \le \|K\|_1^4\|f_{h_o}g\|^4 \le \|g\|_\infty^4\|K\|_1^8\|f\|^4 \le \|g\|_\infty^4\|K\|_1^8\|f\|_\infty^2.$$

Second

$$\begin{aligned}
\langle f_{h_o}g\star K_h,(K_{h_o}(X_2-\cdot)g)\star K_{h_{\min}}\rangle^2 &\le \left\{\sup_z|f_{h_o}g\star K_h(z)|\int|(K_{h_o}(X_2-\cdot)g)\star K_{h_{\min}}(z)|dz\right\}^2\\
&\le \left\{\|K\|_1\sup_z|(f_{h_o}g)(z)|\|K\|_1\int|K_{h_o}(X_1-u)g(u)|du\right\}^2\\
&\le \left\{\|K\|_1^2\|f\|_\infty\|g\|_\infty\times\|g\|_\infty\|K\|_1^2\right\}^2=(\|f\|_\infty\|g\|_\infty^2\|K\|_1^4)^2.
\end{aligned}$$

The twin term in $h_{\min},h$ has clearly the same bound. Lastly

$$\begin{aligned}
&\mathbb{E}\left[\langle(K_{h_o}(X_1-\cdot)g)\star K_h,(K_{h_o}(X_2-\cdot)g)\star K_{h_{\min}}\rangle^2\right]\\
=&\iint\left(\int K_{h_o}(u-\cdot)g\star K_h(x)K_{h_o}(v-\cdot)g\star K_{h\min}(x)dx\right)^2 f(u)f(v)dudv\\
\le&\ \|g\|_\infty^4\iint\left(\int|K_{h_o}|\star|K_h|(u-x)|K_{h_o}|\star|K_{h\min}|(v-x)dx\right)^2 f(u)f(v)dudv\\
=&\ \|g\|_\infty^4\iint\left[|K_{h_o}|\star|K_h|\star|K_{h_o}|\star|K_{h\min}|(u-v)\right]^2 f(u)f(v)dudv\\
\le&\ \|g\|_\infty^4\|f\|_\infty^2\||K_{h_o}|\star|K_h|\star|K_{h_o}|\star|K_{h\min}|\|^2\\
\le&\ \|g\|_\infty^4\|f\|_\infty^2\|K\|_1^6\frac{\|K\|^2}{h}.
\end{aligned}$$

We get

$$\mathfrak{c}_n^2=\frac{n^2}{h}\|g\|_\infty^4\|f\|_\infty^2\|K\|_1^6\left(\|K\|^2+3\|K\|_1^2\right).$$

Thus , for all positive θ,λ it holds

$$\frac{\mathfrak{c}_n\sqrt{\lambda}}{n^2}\le\theta\frac{\|g\|_\infty^4\|f\|_\infty^2\|K\|_1^6}{nh}+\frac{\lambda(\|K\|^2+3\|K\|_1^2)}{4n\,\theta}.$$

⋄ Lastly, the term $\mathfrak{d}_n$ is a bound on

$$\sup_{a,b}\sum_{1\le i\ne j\le n}\mathbb{E}\left[G_{h,h_{\min}}(X_i,X_j)a_i(X_i)b_j(X_j)\right],$$

where $a_k(\cdot)$, $b_k(\cdot)$ for $k=1,\dots,n$ is such that $\mathbb{E}(\sum_{k=1}^n a_k^2(X_k))\le 1$ and $\mathbb{E}(\sum_{k=1}^n b_k^2(X_k))\le 1$. Using the independence for $i\ne j$ between functions of X_i and functions of X_j, we get that the term inside the sup is less than

$$\langle\sum_{i=1}^n\mathbb{E}\left(|K_{h_o}(X_i-\cdot)g-f_{h_o}g|\star|K_h||a_i(X_i)|\right),\sum_{i=1}^n\mathbb{E}\left(|K_{h_o}(X_j-\cdot)g-f_{h_o}g|\star|K_{h_{\min}}||b_j(X_j)|\right)\rangle. \tag{51}$$

First we have

$$\begin{aligned}
\sum_{i=1}^n\mathbb{E}\left(|K_{h_o}(X_i-\cdot)g-f_{h_o}g|\star|K_h||a_i(X_i)|\right) &\le \sqrt{n}\left\{\sum_{i=1}^n\left[\mathbb{E}\left(|K_{h_o}(X_i-\cdot)g-f_{h_o}g|\star|K_h||a_i(X_i)|\right)\right]^2\right\}^{1/2}\\
&\le \sqrt{n}\left\{\sum_{i=1}^n\mathbb{E}\left[\left(|K_{h_o}(X_i-\cdot)g-f_{h_o}g|\star|K_h|\right)^2\right]\mathbb{E}(a_i^2(X_i))\right\}^{1/2}
\end{aligned}$$

As
$$\mathbb{E}\left[(|K_{h_o}(X_i-\cdot)g - f_{h_o}g|\star|K_h|)^2\right] \le \|f\|_\infty\|g\|_\infty^2[\||K_{h_o}|\star|K_h|\|^2 + \|f_{h_o}\star|K_h|\|^2]$$
we get
$$\sum_{i=1}^n \mathbb{E}\left(|K_{h_o}(X_i-\cdot)g - f_{h_o}g|\star|K_h||a_i(X_i)|\right) \le \sqrt{2n\|f\|_\infty}\|g\|_\infty\|K\|_1\|K_h\|.$$
Plugging this in formula (51), we get

$$\begin{aligned}
&\langle\sum_{i=1}^n \mathbb{E}\left(|K_{h_o}(X_i-\cdot)g - f_{h_o}g|\star|K_h||a_i(X_i)|\right), \sum_{i=1}^n \mathbb{E}\left(|K_{h_o}(X_j-\cdot)g - f_{h_o}g|\star|K_{h_{\min}}||b_j(X_j)|\right)\rangle \\
\le\ & \sqrt{2n\|f\|_\infty}\|g\|_\infty\|K\|_1\|K_h\| \sum_{j=1}^n \mathbb{E}\left(\int |K_{h_o}(X_j-\cdot)g - f_{h_o}g|\star|K_{h_{\min}}(u)|du\,|b_j(X_j)|\right) \\
\le\ & \sqrt{2n\|f\|_\infty}\|g\|_\infty\|K\|_1\|K_h\| \times 2\|K\|_1^2\|g\|_\infty \left(\sum_{j=1}^n \mathbb{E}(b_j(X_j)|)\right) \\
\le\ & 2\sqrt{2n\|f\|_\infty}\|g\|_\infty^2\|K\|_1^2\|K_h\| \times \sqrt{n}.
\end{aligned}$$

Therefore,
$$\mathfrak{d}_n := 2\sqrt{2\|f\|_\infty}\|g\|_\infty^2\|K\|_1^2\|K\|\frac{n}{\sqrt{h}}.$$
It follows that
$$\frac{\mathfrak{d}_n\lambda}{n^2} \le \theta\frac{\|f\|_\infty^2\|g\|_\infty^4\|K\|_1^4\|K\|^2}{nh} + 2\frac{\lambda^2}{n}.$$
Applying the deviation inequality for U-statistics of order 2 (see Lacour *et al.* (2017) and Theorem 3.4 in Houdré and Reynaud-Bouret (2003)) leads thus to
$$\mathbb{E}\left\{\sup_{h\in\mathcal{H}_n}\left(\frac{U_n^{\mathbf{X}}(h,h_{\min})}{n^2} - \theta\frac{\|f\|_\infty^2\|g\|_\infty^4\|K\|_1^4\|K\|^2}{nh}\right)\right\} \le C\frac{\log(n)}{n}.$$

Therefore
$$\mathbb{E}\left\{\sup_{h\in\mathcal{H}_n}\left(\frac{U_n^{(2),(2)}(h,h_{\min})}{n^2} - \theta\frac{\|f\|_\infty^2\|g\|_\infty^4\|K\|_1^4\|K\|^2}{nh}\right)\right\} \le C\frac{\log(n)}{n}. \tag{52}$$

The result of Lemma 7.3 follows by gathering the bounds (44), (49), (50), (52), (47), (46), (45).□

8 Appendix

The Talagrand inequality. The result below follows from the Talagrand concentration inequality given in Klein and Rio (2005) and arguments in Birgé and Massart (1998) (see the proof of their Corollary 2 page 354).

Lemma 8.1. *(Talagrand Inequality) Let $Y_1,\ldots,Y_n$ be independent random variables and let $\mathcal{F}$ be a countable class of uniformly bounded measurable functions. Consider ν_n, the centered empirical process defined by*
$$\nu_n(f) = \frac{1}{n}\sum_{i=1}^n [f(Y_i) - \mathbb{E}(f(Y_i))]$$
for $f\in\mathcal{F}$. Assume there exists three positive constants M, H and v such that

$$\sup_{f\in\mathcal{F}}\|f\|_\infty \le b, \quad \mathbb{E}\Big[\sup_{f\in\mathcal{F}}|\nu_n(f)|\Big] \le H, \quad \sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{k=1}^n \mathrm{Var}(f(Y_k)) \le v^2.$$

Then, for any $\delta > 0$ the following holds

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}}|\nu_n(f)|^2 - 2(1+2\delta)H^2\Big]_+ \leq \frac{4}{K_1}\left(\frac{v^2}{n}\exp\left(-K_1\delta\frac{nH^2}{v^2}\right) + \frac{49b^2}{K_1 n^2 C^2(\delta)}\exp\left(-\frac{K_1 C(\delta)\sqrt{2\delta}}{7}\frac{nH}{b}\right)\right),$$

with $C(\delta) = \sqrt{1+\delta} - 1$ and $K_1 = 1/6$.

By standard density arguments, this result can be extended to the case where $\mathcal{F}$ is a unit ball of a linear normed space, after checking that $f \mapsto \nu_n(f)$ is continuous and $\mathcal{F}$ contains a countable dense family.

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series.* newblock For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.

[2] Belomestny, D., Comte, F. and Genon-Catalot, V. Sobolev-Hermite versus Sobolev nonparametric density estimation on $\mathbb{R}$. *Ann. Inst. Statist. Math.* **71**, no. 1, 29-62,(2019).

[3] B. Bongioanni and J. L. Torrea. Sobolev spaces associated to the harmonic oscillator. *Proc. Indian Acad. Sci. Math. Sci.*, 116(3):337–360, 2006.

[4] Butucea, C., Delmas, J.-F., Dutfoy, A. and Fischer, R. Fast adaptive estimation of log-additive exponential models in Kullback-Leibler divergence. *Electron. J. Stat.* **12** , no. 1, 1256-1298, 2018.

[5] Chinot, G and Lerasle, M. On the robustness of the minimum $\ell 2$ interpolator. To appear in *Bernoulli* arXiv:2003.05838, 2020.

[6] Cohen, Albert; Davenport, Mark A.; Leviatan, Dany On the stability and accuracy of least squares approximations. *Found. Comput. Math.* 13 , no. 5, 819-834, 2013.

[7] Comte, F. *Nonparametric estimation.* Spartacus-Idh. 128 p., Paris, 2017.

[8] Comte, F. and Lacour, C. (2021). Noncompact estimation of the conditional density from direct or noisy data, Preprint Hal-03276251.

[9] Comte, F. and Marie, N. On a Nadaraya-Watson estimator with two bandwidths. *Electron. J. Stat.* **15**(1), 2566-2607, 2021.

[10] Comte, F. and Rebafka, T. Adaptive density estimation in the pile-up model involving measurement errors. *Electronic journal of Statistics* **6**, 2002-2037, 2012.

[11] Goldenshluger, A. and Lepski, O. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** (3), 1608-1632, 2011.

[12] 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.

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

[14] Lacour, C., Massart, P. and Rivoirard, V. Estimator selection: a new method with applications to kernel density estimation. *Sankhya* **A**, *The Indian Journal of Statistics*, **79**(2),298-335, 2017.

[15] Lacour, C. Adaptive estimation of the transition density of a Markov chain. *Ann. Inst. H. Poincaré Probab. Statist.* **43**, no. 5, 571-597, 2007.

[16] Laurent, B. and Massart, P. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28**, no. 5, 1302-1338, 2000.

[17] Lerasle, M. Optimal model selection in density estimation, *Ann. Inst. Henri Poincaré,* **48**, 3 884–908, 2012.

[18] Massart, P. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics.* Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

[19] Tsybakov, A. B. Introduction to nonparametric estimation. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. Springer Series in Statistics. Springer, New York, 2009.