

Adaptive regression with Brownian path covariate

Karine Bertin, Nicolas Klutchnikoff

▶ **To cite this version:**

Karine Bertin, Nicolas Klutchnikoff. Adaptive regression with Brownian path covariate. Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques, 2021, 57 (3), pp.1495-1520. 10.1214/20-AIHP1128 . hal-02332820

HAL Id: hal-02332820

https://univ-rennes2.hal.science/hal-02332820

Submitted on 25 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Adaptive regression with Brownian path covariate

Karine Bertin* Nicolas Klutchnikoff†

September 23, 2019

Abstract

This paper deals with estimation with functional covariates. More precisely, we aim at estimating the regression function m of a continuous outcome Y against a standard Wiener coprocess W. Following Cadre and Truquet (2015) and Cadre, Klutchnikoff, and Massiot (2017) the Wiener-Itô decomposition of $m(W)$ is used to construct a family of estimators. The minimax rate of convergence over specific smoothness classes is obtained. A data-driven selection procedure is defined following the ideas developed by Goldenshluger and Lepski (2011). An oracle-type inequality is obtained which leads to adaptive results.

Keywords: Functional regression, Wiener-Itô chaos expansion, Oracle inequalities, Adaptive minimax rates of convergence.

AMS Subject Classification: 62G08, 62H12

1 Introduction

The problem of regression estimation is one of the most studied in statistics and different models have been considered depending on the nature of the data. In an increasing number of applications, it seems natural to assume that the covariate takes values in a functional space. The book of Ramsay and Silverman (2005) provides an overview on the subject of functional data analysis. In this context, several authors studied linear functional regression models (see for example Müller and Stadtmüller, 2005; Cai and Hall, 2006;

*CIMFAV, Universidad de Valparaíso, General Cruz 222, Valparaíso, Chile, karine.bertin@uv.cl

†Univ Rennes, CNRS, IRMAR – UMR 6625, F-35000 Rennes, France, +33 2 99 14 18 19 nicolas.klutchnikoff@univ-rennes2.fr

Crambes, Kneip, and Sarda, 2009). Non-parametric functional regression models have also been investigated (see Ferraty and Vieu, 2006, and references therein). In this paper, we are interested in such a model where the covariate is a Wiener Process. More precisely, let ε be a real-valued random variable and $W = (W(t) : 0 \leq t \leq 1)$ be a standard Brownian motion independent of ε. We define

$$Y = m(W) + \varepsilon$$

where $m : \mathcal{C} \to \mathbb{R}$ is a mapping defined on the set $\mathcal{C}$ of all continuous functions $w : [0, 1] \to \mathbb{R}$ and we assume that both $m(W)$ and ε are square integrable random variables. Our goal is to estimate the function m using a dataset $(Y_1, W_1), \ldots, (Y_n, W_n)$ of independent realizations of (Y, W).

Since this framework is a specific case of the more general functional regression framework, usual approaches (which mainly consist in extending classical local methods such as k-nearest neighbors, kernel smoothing or local polynomial smoothing) could be used. However, in our context, these methods are known to lead to slow rates of convergence over classical models (see below for detailed references). Taking advantage of the probabilistic properties of the Wiener coprocess, we aim at defining a new family of models as well as dedicated estimation procedures with faster rates of convergence (in both minimax and adaptive minimax senses).

In usual functional approaches the set $\mathcal{C}$ is endowed with a metric d (see for example Ferraty and Vieu, 2006; Ferraty, Mas, and Vieu, 2007; Biau, Cérou, and Guyader, 2010) which allows to extend several nonparametric estimators. For example a simple version of the Nadaraya-Watson estimator is given, for any function $w \in \mathcal{C}$ and any bandwidth $h > 0$, by:

$$\tilde{m}_h(w) = \frac{1}{n} \sum_{i=1}^{n} Y_i \frac{\mathbf{I}_{\{d(W_i,w) \leq h\}}}{\sum_{j=1}^{n} \mathbf{I}_{\{d(W_j,w) \leq h\}}},$$

where $\mathbf{I}$ stands for the indicator function. The properties of these estimators are related to the behavior of a quantity known as the *small ball probability* defined for $w \in \mathcal{C}$ and $h > 0$ by $\varphi_w(h) = \mathbf{P}(d(W, w) \leq h)$. Pointwise risks of such methods can be generally bounded, up to a positive factor by

$$h^\beta + \left(\frac{1}{n\varphi_w(h)} \right)^{1/2}$$

where β denotes the smoothness of the mapping m measured in a Hölder sense. For example, if $0 < \beta \leq 1$ it is assumed that there exists $L > 0$ such that $|m(w) - m(w')| \leq L d(w, w')^\beta$ for any $w, w' \in \mathcal{C}$. Under additional assumptions similar results can be obtained for integrated risks.

The classical assumption $\varphi_w(h) \asymp h^k$ corresponds roughly to the situation where the covariate W lies in some space of finite dimension k (see Azaïs and Fort, 2013). This framework corresponds to the usual nonparametric case. The minimax rates of convergence are then given by $n^{-\beta/(2\beta+k)}$ (see Tsybakov, 2009). However if W lies in a functional space, the behavior of $\varphi_w(h)$ is quite different. In our context, where W is a standard Wiener process, it is well-known (see Li and Shao, 2001) that

$$\log \mathbf{P}\left(\sup_{t\in[0,1]} |W(t)| \leq h\right) = \log \varphi_0(h) \underset{h\to 0}{\asymp} -h^{-2}$$

which leads to slower rates of convergence of the form $(\log n)^{-\beta/2}$. We refer the reader to Chagny and Roche (2016) for recent results with different behavior of $\varphi_w(h)$.

In practical situations, since β is unknown, finding adaptive procedures to select the smoothing parameter h is of prime interest. To our best knowledge few papers deal with this problem. Adaptive procedures based on cross validation have been used in Rachdi and Vieu (2007). Chagny and Roche (2016) also propose an adaptation of the method developed by Goldenshluger and Lepski (see Goldenshluger and Lepski, 2011) using an empirical version of the quantity $\varphi_w(h)$. Lower bounds have been investigated by Mas (2012). In all these papers, the pointwise risk is studied in terms of $\varphi_w(h)$ and theoretical properties are obtained assuming a β-Hölder condition on m with respect to the metric d with smoothness $\beta \in (0, 1]$.

In this paper we follow a different strategy. Taking advantage of probabilistic properties of the Wiener process, similarly to the methodology developed by Cadre and Truquet (2015) and Cadre et al. (2017), we consider the Wiener-Itô chaotic decomposition of $m(W)$. Indeed, every random variable that belongs to $\mathbb{L}^2_W = \{\mathfrak{m}(W) \mid \mathfrak{m} : \mathcal{C} \to \mathbb{R} \text{ and } \mathbf{E}(\mathfrak{m}(W))^2 < +\infty\}$ can be decomposed as a sum of multiple stochastic integrals (see Di Nunno, Øksendal, and Proske, 2009, for more details). There exists a unique sequence of functions $(f_\ell)_{\ell\geq 1}$ such that

$$m(W) \overset{\mathbb{L}^2}{=} \mathbf{E}(Y) + \sum_{\ell=1}^{\infty} \frac{1}{\ell!} I_\ell(f_\ell)(W), \tag{1}$$

where f_ℓ belongs to $\mathbf{L}^2_{\mathrm{sym}}(\Delta_\ell)$, the set of symmetric and square integrable real-valued functions defined on $\Delta_\ell = [0, 1]^\ell$ and

$$I_\ell(f_\ell)(W) = \int_{\Delta_\ell} f_\ell \, \mathrm{d}W^{\otimes \ell} = \int_{\Delta_\ell} f_\ell(u_1, \ldots, u_\ell) W(\mathrm{d}u_1) \cdots W(\mathrm{d}u_\ell).$$

The iterated integral $I_\ell(f_\ell)(W)$ is called a chaos of order ℓ. We define a well-adapted family of models assuming that the summation (1) stops at a

finite index L and imposing regularity assumptions on each function f_ℓ for $\ell = 1, \dots, L$ (see Section 2.1 for a precise definition). Kernel-type estimators $\widetilde{f}_\ell$ of f_ℓ can be defined using the Itô's isometry. Using plug-in estimator for each chaos, this lead to a simple estimator of m given by:

$$\hat{m}(W) = \frac{1}{n}\sum_{i=1}^{n} Y_i + \sum_{\ell=1}^{L} \frac{1}{\ell!} I_\ell(\widetilde{f}_\ell)(W).$$

We prove that such dedicated estimators achieve optimal rates of convergence on our models (indeed, these rates are proved to be the minimax ones). Contrary to the classical functional framework (where logarithmic rates are derived) the minimax rates of convergence over our models are polynomial in n with an exponent that depends on the smoothness of the functions f_ℓ. A data-driven procedure, based on the method developed by Goldenshluger and Lepski (2011), is then defined to tune the bandwidths used in the estimation of the functions f_ℓ. The resulting estimator of m satisfies an oracle-type inequality that allows us to derive adaptive results.

The paper is organized as follows. Section 2 presents the model and the studied problem. Section 3 describes the construction of the estimators. Section 4 gives the main results and Section 5 is dedicated to the proofs.

2 Statistical framework

2.1 Model

Let $W = (W(t) : 0 \leq t \leq 1)$ be a standard Brownian motion and let ε be a centered real-valued random variable independent of W. We define:

$$Y = m(W) + \varepsilon$$

where $m : \mathcal{C} \to \mathbb{R}$ is a given mapping. We assume that $m(W)$ as well as ε belong to $\mathbb{L}^2$, the set of square integrable random variables, and that there exists $L \in \mathbb{N}$ such that

$$m(W) = \mathbf{E}(Y) + \sum_{\ell=1}^{L} \frac{1}{\ell!} I_\ell(f_\ell)(W),$$

where for $\ell \in \{1, \dots, L\}$, f_ℓ belongs to $\mathbf{L}^2_{\mathrm{sym}}(\Delta_\ell)$. As mentioned in the introduction we also assume that f_ℓ is a regular function. Below we define precisely the functional classes used to measure the smoothness of each function f_ℓ.

Definition 1 *Set $\ell \in \mathbb{N}$, $s_\ell > 0$, $\Lambda_\ell > 0$ and $M > 0$. The Hölder ball $\mathcal{H}_\ell(s_\ell, \Lambda_\ell, M)$ is the set of all functions $f : \Delta_\ell \to \mathbb{R}$ that satisfy the following properties:*

1. *For any $\alpha = (\alpha_1, \ldots, \alpha_\ell) \in \mathbb{N}^\ell$ such that $|\alpha| = \sum_i \alpha_i \leq \lfloor s_\ell \rfloor = \max\{k \in \mathbb{N} \mid k < s_\ell\}$, the partial derivative $D^\alpha f$ exists where*

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x^{\alpha_1} \ldots \partial x^{\alpha_\ell}}.$$

2. *For any x and y in Δ_ℓ we have:*

$$\sum_{|\alpha| = \lfloor s_\ell \rfloor} |D^\alpha f(x) - D^\alpha f(y)| \leq \Lambda_\ell |x - y|^{s - \lfloor s_\ell \rfloor},$$

   where $|\cdot|$ stands for the Euclidean norm of $\mathbb{R}^\ell$.

3. *We have $\|f\|^2_{\Delta_\ell} \leq M^2\, \ell!$ where:*

$$\|f\|_{\Delta_\ell} = \left(\int_{\Delta_\ell} f^2(u)\, \mathrm{d}u \right)^{1/2}.$$

Equipped with these notations we can define a scale of models for the mapping m. Each model is caraterized by L functional classes $\mathcal{H}_\ell(s_\ell, \Lambda_\ell, M)$ for $\ell = 1, \ldots, L$.

Definition 2 *Set $s = (s_1, \ldots, s_L) \in (0, +\infty)^L$, $\Lambda = (\Lambda_1, \ldots, \Lambda_L) \in (0, +\infty)^L$ and $M > 0$. We denote by $\mathbb{M}_{s,\Lambda,M}$ the subset of $\mathbb{L}^2_W$ defined by:*

$$\mathbb{M}_{s,\Lambda,M} = \mathbb{R} \oplus \left(\bigoplus_{\ell=1}^{L} \frac{1}{\ell!} I_\ell\Big(\mathcal{H}_\ell(s_\ell, \Lambda_\ell, M)\Big)(W) \right)$$

where $I_\ell\Big(\mathcal{H}_\ell(s_\ell, \Lambda_\ell, M)\Big)(W)$ consists of all random variables $I_\ell(f)(W)$ with $f \in \mathcal{H}_\ell(s_\ell, \Lambda_\ell, M)$. We also denote by $\mathfrak{M}_{s,\Lambda,M}$ the corresponding class of mappings $\mathfrak{m} : \mathcal{C} \to \mathbb{R}$ such that $\mathfrak{m}(W) \in \mathbb{M}_{s,\Lambda,M}$.

2.2 Adaptive framework

The observations consist in a n-sample $(Y_1, W_1), \ldots, (Y_n, W_n)$ distributed as and independent of (Y, W). Our main goal is to investigate the adaptive estimation of m, based on these observations, over the scale of mapping classes $\mathcal{M}(s^*) = \{\mathfrak{M}_{s,\Lambda,M} \mid s \in (0, s^*)^L, \Lambda \in (0, +\infty)^L, M > 0\}$ where

$s^* > 0$ is fixed. To measure the accuracy of an arbitrary estimator $\tilde{m}_n = \tilde{m}(\,\cdot\,; (Y_1, W_1), \ldots, (Y_n, W_n))$ of m, we consider the following risk:

$$R_p(\tilde{m}_n, m) = \Big(\mathbf{E}|\tilde{m}_n(W) - m(W)|^p\Big)^{1/p}$$

where $p \geq 2$.

The maximal risk of an arbitrary estimator $\tilde{m}_n$ over a given class of mappings $\mathfrak{M}$ is defined by:

$$R_p(\tilde{m}_n, \mathfrak{M}) = \sup_{\mathfrak{m} \in \mathfrak{M}} R_p(\tilde{m}_n, \mathfrak{m}),$$

whereas the minimax risk is defined, taking the infimum over all possible estimators, by:

$$\Phi_n(\mathfrak{M}, p) = \inf_{\tilde{m}_n} R_p(\tilde{m}_n, \mathfrak{M}).$$

An estimator $\hat{m}_n$ whose maximal risk is asymptotically bounded, up to a multiplicative factor, by $\Phi_n(\mathfrak{M}, p)$ is called minimax over $\mathfrak{M}$. Such an estimator is well-adapted to the estimation over $\mathfrak{M}$ but it can performs poorly over another class of mappings. The problem of adaptive estimation consists in finding a single estimation procedure that is simultaneously minimax over a scale of mapping classes. More precisely our goal is to construct a single estimation procedure m_n^* such that, for any $\mathfrak{M} \in \mathcal{M}(s^*)$, the risk $R_p(m_n^*, \mathfrak{M})$ is asymptotically bounded, up to a multiplicative constant, by $\Phi_n(\mathfrak{M}, p)$. One of the main tools to prove such a result is to find an oracle-type inequality that guarantees that this procedure performs almost as well as the best estimator in a rich family of estimators. Ideally, we would like to have, for any $\mathfrak{m} \in \bigcup \mathfrak{M}_{s,\Lambda,M}$, an inequality of the following form:

$$R_p(m_n^*, \mathfrak{m}) \leq \inf_{\eta \in H} R_p(\hat{m}_{n,\eta}, \mathfrak{m}), \tag{2}$$

where $\{\hat{m}_{n,\eta} \mid \eta \in H\}$ is a family of estimators *well-adapted* to our problem in the following sense: for any $\mathfrak{M} \in \mathcal{M}(s^*)$, there exists $\eta \in H$ such that $\hat{m}_{n,\eta}$ is minimax over $\mathfrak{M}$. However, in many situations, (2) is relaxed and we prove a weaker inequality of the type:

$$R_p(m_n^*, \mathfrak{m}) \leq \Upsilon_{1,p} \inf_{\eta \in H} R_p^*(\mathfrak{m}, \eta) + \Upsilon_{2,p} \left(\frac{\log n}{n}\right)^{1/2}, \tag{3}$$

where $\Upsilon_{1,p}$ and $\Upsilon_{2,p}$ are two positive constants and $R_p^*(\mathfrak{m}, \eta)$ is an appropriate quantity to be determined that can be viewed as a tight upper bound on $R_p(\hat{m}_{n,\eta}, \mathfrak{m})$. Inequalities of the form (3) are called *oracle-type* inequalities.

Theorems 2 and 3 below correspond respectively to an oracle-type inequality and an adaptive result of these types.

3 Estimator construction

In this section we present our estimation procedure. To do so, we first recall classical properties satisfied by Wiener chaos which allow us to construct a family of "simple" estimators that depends on a multivariate tuning parameter. Next we construct a procedure which selects, in a data-driven way, this tuning parameter using the methodology developed by Goldenshluger and Lepski (2011).

3.1 Classical properties of the chaos

Throughout this paper and in the construction of our statistical procedure, we use the following two fundamental properties satisfied by the iterated integrals.

For $\ell, \ell' \in \mathbb{N}$, Itô's isometry (Di Nunno et al., 2009) ensures that, if $g \in \mathbf{L}^2_{\mathrm{sym}}(\Delta_\ell)$ and $g' \in \mathbf{L}^2_{\mathrm{sym}}(\Delta_{\ell'})$, then

$$\mathbf{E}\Big(I_\ell(g)(W) I_{\ell'}(g')(W)\Big) = \delta_{\ell,\ell'}\, \ell! \int_{\Delta_\ell} g(u) g'(u)\, \mathrm{d}u \tag{4}$$

where $\delta_{\ell,\ell'}$ denotes the Kronecker delta.

The hypercontractivity property (Nourdin and Peccati, 2012) will be used to control the concentration of our estimators. Set $q \geq 2$ and $\ell \in \mathbb{N}$. There exists a positive constant $\mathfrak{c}_\ell(q)$ such that for any $g \in \mathbf{L}^2_{\mathrm{sym}}(\Delta_\ell)$ we have:

$$\left(\mathbf{E}|I_\ell(g)(W)|^q\right)^{1/q} \leq \mathfrak{c}_\ell(q) \left(\mathbf{E} I^2_\ell(g)(W)\right)^{1/2}. \tag{5}$$

3.2 A simple family of estimators

Let $\mathbf{k} : \mathbb{R} \to \mathbb{R}$ be a function that satisfies the following properties: $\mathbf{k}$ is continuous inside $[0,1]$, $\mathbf{k}(x) = 0$ for any $x \notin [0,1]$,

$$\int_0^1 \mathbf{k}(x)\, \mathrm{d}x = 1 \qquad \text{and} \qquad \int_0^1 x^s \mathbf{k}(x)\, \mathrm{d}x = 0, \quad s = 1, \ldots, \lfloor s^* \rfloor.$$

Let $\ell \in \{1, \ldots, L\}$. A natural estimator of the function f_ℓ is given, for $h \in (0,1)$, by:

$$\hat{f}^{(\ell)}_h(t) = \frac{1}{n} \sum_{i=1}^n Y_i I_\ell\left(K^{(\ell)}_h(t, \cdot)\right)(W_i), \qquad t \in \Delta_\ell \tag{6}$$

where $K^{(\ell)}_h$ is a multivariate kernel defined by:

$$K^{(\ell)}_h(t,u) = \frac{1}{h^\ell} \prod_{k=1}^\ell \mathbf{k}\left(\varsigma(t_k) \frac{t_k - u_k}{h}\right) \qquad \text{with} \qquad \varsigma(\cdot) = 2 I_{(1/2,1)}(\cdot) - 1.$$

This specific construction allows one to obtain an estimator free of boundary bias (see Bertin, El Kolei, and Klutchnikoff, 2018, for more details).

Indeed, note that for any $t \in \Delta_\ell$ and under regularity assumptions on f_ℓ we have:

$$\begin{aligned} f_\ell(t) &\overset{(h\to 0)}{\approx} \int_{\Delta_\ell} f_\ell(u) K_h^{(\ell)}(t-u)\,\mathrm{d}u \\ &= \mathbf{E}\left(\frac{1}{\ell!} I_\ell(f_\ell)(W) I_\ell\left(K_h^{(\ell)}(t,\cdot)\right)(W)\right) \\ &= \mathbf{E}\left(m(W) I_\ell\left(K_h^{(\ell)}(t,\cdot)\right)(W)\right) \end{aligned}$$

where the last two lines are obtained using (1) and (4). Since ε is centered and independent of W we have:

$$\begin{aligned} f_\ell(t) &\overset{(h\to 0)}{\approx} \mathbf{E}\left(Y I_\ell\left(K_h^{(\ell)}(t,\cdot)\right)(W)\right) \\ &\overset{(n\to+\infty)}{\approx} \hat{f}_h^{(\ell)}(t). \end{aligned}$$

Equipped with these notations we define a family of plugin estimators of the mapping m. For all $\boldsymbol{h} = (h_1, \dots, h_L) \in (0,1)^L$ we set:

$$\hat{m}_{\boldsymbol{h}}(W) = \overline{Y}_n + \sum_{\ell=1}^{L} \frac{1}{\ell!} I_\ell\left(\hat{f}_{h_\ell}^{(\ell)}\right)(W)$$

where $\overline{Y}_n = \sum_{i=1}^n Y_i/n$.

3.3 Selection procedure

Set $\tau > 0$ and $M > 0$. Assume that $\mu_4 = (\mathbf{E}|\varepsilon|^4)^{1/4}$ exists. Let $\ell \in \{1, \dots, L\}$ be fixed and define

$$\mathbf{H}_\ell = \left\{h \in (0,1) : n^{-1/(2\tau+\ell)} \le h \le (\log n)^{-1}\right\} \cap \{e^{-k} : k \in \mathbb{N}\}.$$

Now, define

$$M(\ell, h) = \frac{\nu(\ell)\left(1 + 4\sqrt{\log(1/h^\ell)}\right)}{\sqrt{nh^\ell}} \tag{7}$$

where

$$\nu(\ell) = \left(\mu_4 + \sum_{k=1}^{L} \mathfrak{c}_4(k) M\right) \frac{\sqrt{b_{\ell,2}}}{2} \qquad \text{and} \qquad b_{\ell,2} = \mathfrak{c}_\ell^2(4) 2^\ell \ell! \|\mathbf{k}\|_{[0,1]}^{2\ell},$$

where the constants $\mathfrak{c}_\ell(k)$ are defined in (5). Define for $h \in \mathbf{H}_\ell$

$$B(\ell, h) = \max_{h' \in \mathbf{H}_\ell} \left\{\|\hat{f}_{h'}^{(\ell)} - \hat{f}_{h\vee h'}^{(\ell)}\| - M(\ell, h') - M(\ell, h \vee h')\right\}_+$$

and set

$$\hat{h}_\ell = \arg\min_{h \in \mathbf{H}_\ell} \{B(\ell, h) + M(\ell, h)\}\,. \tag{8}$$

The estimation procedure is the defined by $\hat{m} = \hat{m}_{\hat{\boldsymbol{h}}}$ where $\hat{\boldsymbol{h}} = \left(\hat{h}_1, \ldots, \hat{h}_L\right)$.

Remark 1 *This selection rule follows the principles and the ideas developed by Goldenshluger and Lepski in a series of papers (see Goldenshluger and Lepski, 2011, 2014, among others). The quantity $M(\ell, h)$, which is called a* majorant *in the papers cited above, is a penalized version of the standard deviation of the estimator $\hat{f}_h^{(\ell)}$ while the quantity $B(\ell, h)$ is, in some sense, closed to its bias term, see (24). Finding tight* majorants *is the key point of the method since $\hat{h}_\ell$ is chosen in (8) in order to realize an empirical trade-off between these two quantities.*

It is worth noting that the procedure depends on a hyperparameter $\tau > 0$ which can be chosen arbitrary small. The introduction of this parameter is due to technical reasons, see (28) in the proof of Lemma 2. This additional assumption (we would like to take $\tau = 0$) implies some restrictions on Theorem 3 below.

4 Main results

Our first result proves that the minimax rate of convergence over the class $\mathfrak{M}_{s,\Lambda,M}$ is of the same order as:

$$\phi_n(s, \Lambda, M) = \max\left\{ \Lambda_\ell^{\ell/(2s_\ell+\ell)} \left(\frac{M^2}{n}\right)^{s_\ell/(2s_\ell+\ell)} \mid \ell = 1, \ldots, L \right\}.$$

Theorem 1 *Set $p \geq 2$, $L \in \mathbb{N}$, $s \in (0, s^*)^L$, $\Lambda \in (0, +\infty)^L$, $M > 0$ and assume that $\mathbf{E}|\varepsilon|^p < +\infty$. Define $\boldsymbol{h}_n(s, \Lambda, M) = \left(h_n^{(\ell)}(s, \Lambda, M)\right)_{\ell=1,\ldots,L} \in (0, 1)^L$ where:*

$$h_n^{(\ell)}(s, \Lambda, M) = \left(\frac{M^2}{\Lambda_\ell^2\, n}\right)^{\frac{1}{2s_\ell+\ell}}, \qquad \ell = 1, \ldots, L.$$

There exist two positive constants κ_* and κ^* that depend only on L and s^* such that

$$\limsup_{n \to +\infty} \phi_n^{-1}(s, \Lambda, M) R_p\left(\hat{m}_{\boldsymbol{h}_n(s,\Lambda,M)}, \mathfrak{M}_{s,\Lambda,M}\right) \leq \kappa^* \tag{9}$$

and

$$\liminf_{n \to +\infty} \phi_n^{-1}(s, \Lambda, M) \Phi_n\left(\mathfrak{M}_{s,\Lambda,M}, p\right) \geq \kappa_*. \tag{10}$$

Note that this result also ensures that the family of estimators constructed in Section 3.2 is *well-adapted* to our problem. The next result states an *oracle-type* inequality satisfied by our data-driven estimator $\hat{m}$.

Theorem 2 *Assume that for any $\ell = 1, \dots, L$, $\|f_\ell\|_{\Delta_\ell}^2 \leq M^2 \ell!$ and that for any $q \geq 1$ the moment $\mu_q = (\mathbf{E}|\varepsilon|^q)^{1/q}$ exists. Then:*

$$R_2(\hat{m}, m) \leq \Upsilon_1 \sum_{\ell=1}^{L} \inf_{h \in \mathbf{H}_\ell} \left[\max_{\substack{h' \in \mathbf{H}_\ell \\ h' \leq h}} \|\mathbf{E}\hat{f}_{h'}^{\ell} - f_\ell\|_2 + M(\ell, h) \right] + \Upsilon_2 \left(\frac{\log n}{n} \right)^{1/2},$$

where Υ_1 and Υ_2 are two positive constants.

Using Theorems 1 and 2 we can derive our last result: the data-driven estimation procedure is adaptive, up to a logarithmic factor, over the scale $\{\mathfrak{M}_{s,\Lambda,M} : s \in [\tau, s^*)^L, \Lambda \in (0, +\infty)^L, M > 0\}$.

Theorem 3 *Assume that for any $q \geq 1$ the moment $\mu_q = (\mathbf{E}|\varepsilon|^q)^{1/q}$ exists. For any $s \in [\tau, s^*)^L$, any $\Lambda \in (0, +\infty)^L$, any $M > 0$, we have*

$$\limsup_{n \to +\infty} \tilde{\phi}_n^{-1}(s) R_2(\hat{m}, \mathfrak{M}_{s,\Lambda,M}) \leq \kappa^{**}$$

where κ^{**} is a positive constant and

$$\tilde{\phi}_n(s) = \max \left\{ \left(\frac{\log n}{n} \right)^{s_\ell/(2s_\ell + \ell)} \mid \ell = 1, \dots, L \right\}.$$

Remark 2 *While the minimax rate of convergence over $\mathfrak{M}(s, \Lambda, M)$ is given in Theorem 1 for any risk $R_p(\cdot, \cdot)$ with $p \geq 2$, Theorems 2 and 3 are only stated for the risk $R_2(\cdot, \cdot)$. We made this choice to avoid too technical details in the proofs but similar results could be obtained for $p > 2$ using the hypercontractivity property. Note that in Theorem 2, the quantity*

$$\max_{\substack{h' \in \mathbf{H}_\ell \\ h' \leq h}} \|\mathbf{E}\hat{f}_{h'}^{\ell} - f_\ell\|_2$$

is a tight upper bound of the bias term of the estimator $\hat{f}_h^\ell$.

This result ensures that our data-driven procedure is adaptive, up to a logarithmic factor, over a large scale of mapping classes since τ can be chosen as close to 0 as one wants.

The presence of the extra logarithmic factor in the adaptive rate of convergence is not usual for $p = 2$. This term is introduced in the definition of $M(\ell, h)$ to control the deviation of the estimator (6) based on the variables $I_\ell\left(K_h^{(\ell)}(t, \cdot)\right)(W_i)$. See (29) for more details.

5 Proofs

We first consider some notation and lemmas. Define for $i \in \{1,\dots,n\}$, $\ell \in \{1,\dots,L\}$ and $h \in (0,1)$

$$\xi_{i,\ell}(t,h) = I_\ell\Big(K_h^\ell(t-\cdot)\Big)(W_i), \quad \text{and} \quad \xi_\ell(t,h) = I_\ell\Big(K_h^\ell(t-\cdot)\Big)(W)$$

and

$$\Theta_{i,\ell} = I_\ell(f_\ell)(W_i) \quad \text{and} \quad \Theta_{i,\ell} = I_\ell(f_\ell)(W).$$

Lemma 1 *We have, for any $\ell \in \{1,\dots,L\}$, $h \in \mathbf{H}_\ell$ and $r \geq 1$*

$$\Big(\mathbf{E}\xi_\ell^{2r}(t,h)\Big)^{1/r} \leq b_{\ell,r}h^{-\ell} \quad \textit{with} \quad b_{\ell,r} = \mathfrak{c}_\ell^2(2r)2^\ell \ell! \|\mathbf{k}\|_{[0,1]}^{2\ell}.$$

Moreover for $\varphi > 0$ and $q \geq 1$

$$\mathbf{E}\left(|\xi_\ell(t,h)|^r \mathbf{I}_{|\xi_\ell(t,h)|>\varphi}\right) \leq (b_{\ell,r})^{r/2}(b_{\ell,q})^{q/2}\varphi^{-q}h^{-\ell(r+q)/2}.$$

Lemma 2 *Let $\ell \in \{1,\dots,L\}$ and $h \in \mathbf{H}_\ell$. Let $\chi, \chi_1, \dots, \chi_n$ be i.i.d random variables such that, for any $r \geq 1$*

$$\Big(\mathbf{E}|\chi|^{2r}\Big)^{1/(2r)} \leq a_r < +\infty.$$

Define

$$\mathbf{U}(t) = \frac{1}{n}\sum_{i=1}^n \Big\{\chi_i\xi_{i,\ell}(t,h) - \mathbf{E}\left(\chi_i\xi_{i,\ell}(t,h)\right)\Big\}$$

and

$$T = (1+\delta)\frac{a_2\sqrt{b_{\ell,2}}}{2\sqrt{nh^\ell}} \qquad \textit{with} \qquad \delta = 4\sqrt{\log(1/h^\ell)}.$$

Then there exists a positive constant $C > 0$ such that

$$\mathbf{E}\left\{\|\mathbf{U}\|_2 - T\right\}_+^2 \leq Cn^{-1}. \tag{11}$$

The following Lemma recalls the Bousquet's version of Talagrand's concentration inequality (see Bousquet, 2002; Boucheron, Lugosi, and Massart, 2013).

Lemma 3 (Bousquet's inequality) *Let $X_1,\dots,X_n$ be independent identically distributed random variables. Let $\mathcal{S}$ be a countable set of functions and define $Z = \sup_{s\in\mathcal{S}}\sum_{i=1}^n s(X_i)$. Assume that, for all $i = 1,\dots,n$ and $s \in \mathcal{S}$, we have $\mathbf{E}s(X_i) = 0$ and $s(X_i) \leq 1$ almost surely. Assume also that $v = 2\mathbf{E}Z + \sup_{s\in\mathcal{S}}\sum_{i=1}^n \mathbf{E}(s(X_i))^2 < \infty$. Then we have for all $t > 0$*

$$\mathbf{P}\left(Z - \mathbf{E}Z \geq t\right) \leq \exp\left\{-\frac{t^2}{2(v+\frac{t}{3})}\right\}.$$

5.1 Proof of Theorem 1

Set $p \geq 2$, $L \in \mathbb{N}$, $s \in (0, s^*)^L$, $\Lambda \in (0, +\infty)^L$ and $M > 0$. For the sake of readability we denote $\boldsymbol{h} = \boldsymbol{h}_n(s, \Lambda, M)$, $h_\ell = h_n^{(\ell)}(s, \Lambda, M)$, $K_\ell = K_{h_\ell}^{(\ell)}$, $\xi_\ell(t) = \xi_\ell(t, h_\ell)$ and $\hat{f}_\ell = \hat{f}_{h_\ell}^{(\ell)}$. This proof is decomposed into two parts. We first prove the upper bound (9) and then the lower bound (10).

5.1.1 Proof of the upper bound

Decomposition of the risk. Using the triangle inequality we have:

$$\begin{aligned} R_p(\hat{m}_{\boldsymbol{h}}, m) &= \Big(\mathbf{E}|\hat{m}_{\boldsymbol{h}}(W) - m(W)|^p\Big)^{1/p} \\ &\leq \Big(\mathbf{E}|\bar{Y}_n - \mathbf{E}(Y)|^p\Big)^{1/p} + \sum_{\ell=1}^{L} \frac{1}{\ell!} \Big(\mathbf{E}\Big|I_\ell\Big(\hat{f}_\ell - f_\ell\Big)(W)\Big|^p\Big)^{1/p} \\ &\leq \Big(\mathbf{E}|\bar{Y}_n - \mathbf{E}(Y)|^p\Big)^{1/p} + \sum_{\ell=1}^{L} \frac{\mathfrak{c}_\ell(p)}{\ell!} \Big(\mathbf{E}\Big|I_\ell\Big(\hat{f}_\ell - f_\ell\Big)(W)\Big|^2\Big)^{1/2} \end{aligned}$$

Last lines comes from the hypercontractivity property. Now, using Itô's isometry, we obtain:

$$\begin{aligned} R_p(\hat{m}_{\boldsymbol{h}}, m) &\leq \Big(\mathbf{E}|\bar{Y}_n - \mathbf{E}(Y)|^p\Big)^{1/p} + \sum_{\ell=1}^{L} \frac{\mathfrak{c}_\ell(p)}{\sqrt{\ell!}} \Big(\mathbf{E}\|\hat{f}_\ell - f_\ell\|_{\Delta_\ell}^2\Big)^{1/2} \\ &\leq \Big(\mathbf{E}|\bar{Y}_n - \mathbf{E}(Y)|^p\Big)^{1/p} + \sum_{\ell=1}^{L} \frac{\mathfrak{c}_\ell(p)}{\sqrt{\ell!}} \Big(B(\ell) + V(\ell)\Big), \end{aligned} \tag{12}$$

where the *bias term* $B(\ell)$ and the *stochastic term* $V(\ell)$ are defined by:

$$B(\ell) = \|\mathbf{E}\hat{f}_\ell - f_\ell\|_{\Delta_\ell} \quad \text{and} \quad V(\ell) = \Big(\mathbf{E}\|\hat{f}_\ell - \mathbf{E}\hat{f}_\ell\|_{\Delta_\ell}^2\Big)^{1/2}.$$

Study of the constant term. Remark that

$$\begin{aligned} \Big(\mathbf{E}|\bar{Y}_n - \mathbf{E}Y|^p\Big)^{1/p} &= \left(\mathbf{E}\left|\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathbf{E}Y_i)\right|^p\right)^{1/p} \\ &\leq \frac{C_{1,p}}{n^{1-1/p}} \left(\mathbf{E}|Y - \mathbf{E}Y|^p\right)^{1/p} + \frac{C_{2,p}\sigma_Y}{n^{1/2}} \end{aligned} \tag{13}$$

where the last line is obtained using Rosenthal's inequality. Here $C_{1,p}$ and $C_{2,p}$ denote two positive constants while σ_Y stands for the standard deviation

of Y (which is finite since both $m(W)$ and ε belong to $\mathbb{L}^2$). Moreover we have

$$|Y - \mathbf{E}Y|^p = \left|\sum_{\ell=1}^{L} \frac{\Theta_\ell}{\ell!} + \varepsilon\right|^p$$
$$\leq (L+1)^{p-1} \left(\sum_{\ell=1}^{L} \left(\frac{|\Theta_\ell|}{\ell!}\right)^p + |\varepsilon|^p\right)$$

which implies, using the hypercontractivity property combined with Itô's isometry, that

$$(\mathbf{E}|Y - \mathbf{E}Y|^p)^{1/p} \leq (L+1)^{1-1/p} \left(\sum_{\ell=1}^{L} \frac{1}{\ell!} \left(\mathbf{E}\,|\Theta_\ell|^p\right)^{1/p} + \left(\mathbf{E}|\varepsilon|^p\right)^{1/p}\right)$$
$$\leq (L+1)^{1-1/p} \left(\sum_{\ell=1}^{L} \frac{\mathfrak{c}_\ell(p)}{\ell!} \left(\mathbf{E}\Theta_\ell^2\right)^{1/2} + \left(\mathbf{E}|\varepsilon|^p\right)^{1/p}\right)$$

Using hypercontractivity we obtain

$$(\mathbf{E}|Y - \mathbf{E}Y|^p)^{1/p} \leq (L+1)^{1-1/p} \left(M \sum_{\ell=1}^{L} \frac{\mathfrak{c}_\ell(p)}{\sqrt{\ell!}} + \left(\mathbf{E}|\varepsilon|^p\right)^{1/p}\right) \tag{14}$$

where the right hand side of the above inequality is a finite quantity denoted by $s_Y(p)$. Taking together (13) and (14) we finally obtain

$$\left(\mathbf{E}|\bar{Y}_n - \mathbf{E}Y|^p\right)^{1/p} \leq \frac{C_{1,p} s_Y(p)}{n^{1-1/p}} + \frac{C_{2,p}\sigma_Y}{n^{1/2}}$$
$$\leq \frac{\kappa_0}{n^{1/2}}, \tag{15}$$

with $\kappa_0 = C_{1,p} s_Y(p) + C_{2,p}\sigma_Y$.

Study of the bias term. Set $\ell \in \{1, \ldots, L\}$ and note that:

$$\mathbf{E}\hat{f}_\ell(t) = \mathbf{E}\left(Y\xi_\ell(t)\big)(W)\right) = \mathbf{E}\left(m(W)\xi_\ell(t)\right) = \frac{1}{\ell!}\mathbf{E}\left(\Theta_\ell \xi_\ell(t)\right)$$

Using Itô's isometry we thus obtain:

$$\mathbf{E}\hat{f}_\ell(t) = \int_{\Delta_\ell} f_\ell(u) K_\ell(t-u)\,\mathrm{d}u.$$

Since $f_\ell \in \mathcal{H}_\ell(s_\ell, \Lambda_\ell, M)$, we obtain:

$$B(\ell) \leq \mathfrak{b}_\ell(\mathbf{k}, s)\,\Lambda_\ell h_\ell^{s_\ell} \tag{16}$$

Previous inequality follows from Taylor's expansion and classical arguments (see Bertin et al., 2018). The expression of $\mathfrak{b}_\ell(\mathbf{k}, s)$ can be computed:

$$\mathfrak{b}_\ell(\mathbf{k}, s) = 2^\ell m_\ell \sum_{|\alpha|=m_\ell} \prod_{i=1}^{\ell} \int_0^1 |\mathbf{k}(y)| y^{\alpha_i+\gamma_\ell} \, \mathrm{d}y.$$

Here, $\alpha = (\alpha_1, \ldots, \alpha_\ell) \in (\mathbb{N} \cup \{0\})^\ell$, $|\alpha| = \alpha_1 + \ldots + \alpha_\ell$, and $s_\ell = m_\ell + \gamma_\ell$ with $m_\ell \in \mathbb{N} \cup \{0\}$ and $0 < \gamma_\ell \leq 1$.

Study of the stochastic term $V(\ell)$ Set $\ell \in \{1, \ldots, L\}$. We have:

$$\Big(V(\ell)\Big)^2 = \frac{\mathbf{V}_1(\ell) + \mathbf{V}_2(\ell)}{n},$$

where

$$\mathbf{V}_1(\ell) = \int_{\Delta_\ell} \mathbf{E}\left(\varepsilon^2 \xi_\ell^2(t)\right) \, \mathrm{d}t$$

and

$$\mathbf{V}_2(\ell) = \int_{\Delta_\ell} \mathrm{Var}\left(m(W)\xi_\ell(t)\right) \, \mathrm{d}t.$$

Since W and ε are independent Lemma 1 implies:

$$\begin{aligned} \mathbf{V}_1(\ell) &= \int_{\Delta_\ell} \mathbf{E}\left(\varepsilon^2\right) \mathbf{E}\left(\xi_\ell^2(t)\right) dt \\ &\leq \frac{\mu_2^2 b_{\ell,1}}{h_\ell^\ell}. \end{aligned}$$

Now recall some usual conventions. First we assume that $0! = 1$ and, following Di Nunno et al. (2009), we define $f_0 = \mathbf{E}(Y)$ while $I_0(\cdot)(W)$ stands for the identity application. Using these conventions we have:

$$\begin{aligned} \mathbf{V}_2(\ell) &= \int_{\Delta_\ell} \mathbf{E}\left(\sum_{k=1}^{L} \frac{1}{k!} \Theta_k \xi_\ell(t)\right)^2 \mathrm{d}t \\ &= \sum_{k,k'=1}^{L} \frac{1}{k!k'!} \int_{\Delta_\ell} \mathbf{E}\left(\Theta_k \Theta_{k'} \xi_\ell^2(t)\right) \, \mathrm{d}t. \end{aligned}$$

Using Cauchy-Schwarz inequality we obtain:

$$\mathbf{V}_2(\ell) \leq \sum_{k,k'=1}^{L} \frac{\left(\mathbf{E}\Theta_k^4\right)^{1/4} \left(\mathbf{E}\Theta_{k'}^4\right)^{1/4}}{k!k'!} \int_{\Delta_\ell} \left(\mathbf{E}\xi_\ell^4(t)\right)^{1/2} \, \mathrm{d}t.$$

Now, using Lemma 1 and denoting by convention $\mathfrak{c}_0(q) = 1$ for any $q > 0$, we obtain:

$$\mathbf{V}_2(\ell) \leq M^2 \left((\ell!)\mathfrak{c}_\ell^2(4) \sum_{k,k'=1}^{L} \mathfrak{c}_k(4)\mathfrak{c}_{k'}(4) \right) \|\mathbf{k}\|_{\Delta_1}^{2\ell} h_\ell^{-\ell}$$

Now, define

$$\mathfrak{v}_\ell(\mathbf{k}, L) = \left((\ell!)\mathfrak{c}_\ell^2(4) \sum_{k,k'=1}^{L} \mathfrak{c}_k(4)\mathfrak{c}_{k'}(4) + \sigma^2 \right) \|\mathbf{k}\|_{\Delta_1}^{2\ell}$$

Taking the bounds on $\mathbf{V}_1(\ell)$ and $\mathbf{V}_2(\ell)$ together we obtain

$$\Big(V(\ell)\Big)^2 \leq \frac{\mathfrak{v}_\ell(\mathbf{k}, L) M^2}{n h_\ell^\ell}. \tag{17}$$

Control of the minimax risk Taking together (16) and (17) and the definition of h_ℓ, we obtain:

$$\begin{aligned}
\mathbf{E}\|\hat{f}_{\ell,h_\ell} - f_\ell\|_{\Delta_\ell}^2 &\leq \mathfrak{b}_\ell^2(\mathbf{k}, s)\, \Lambda_\ell^2 h_\ell^{2s_\ell} + \frac{\mathfrak{v}_\ell(\mathbf{k}, L) M^2}{n h_\ell^\ell} \\
&\leq \kappa_\ell(\mathbf{k}, s, L) \Lambda_\ell^{2\ell/(2s_\ell+\ell)} \left(\frac{M^2}{n} \right)^{s_\ell/(2s_\ell+\ell)}
\end{aligned}$$

where $\kappa_\ell(\mathbf{k}, s, L) = \Big(\mathfrak{b}_\ell^2(\mathbf{k}, s) + \mathfrak{v}_\ell(\mathbf{k}, L)\Big)$. Combining the above result with (12), we obtain the following bound on the risk:

$$\begin{aligned}
R_p(\hat{m}_{\boldsymbol{h}}, m) &\leq \frac{\kappa_0}{\sqrt{n}} + \sum_{\ell=1}^{L} \frac{\mathfrak{c}_\ell(p)\kappa_\ell(\mathbf{k}, s, L)}{\ell!} \Lambda_\ell^{2\ell/(2s_\ell+\ell)} \left(\frac{M^2}{n} \right)^{s_\ell/(2s_\ell+\ell)} \\
&\leq \kappa^* \max \left\{ \Lambda_\ell^{2\ell/(2s_\ell+\ell)} \left(\frac{M^2}{n} \right)^{s_\ell/(2s_\ell+\ell)} \;\middle|\; \ell = 1, \ldots, L \right\}
\end{aligned}$$

where κ^* is a positive constant that depends only on L and s^*. Note that last line is valid for n large enough. This ends the proof of the upper bound. Now, let us prove the lower bound.

5.1.2 Proof of the lower bound

Method. We fix $s \in (0, s^*)^L$, $\Lambda \in (0, +\infty)^L$ and $M > 0$. To prove the lower bound over the space $\mathfrak{M}(s, \Lambda, M)$, we define

$$\ell = \underset{k=1,\ldots,L}{\arg\max}\, \Lambda_k^{k/(2s_k+k)} \left(\frac{M^2}{n} \right)^{s_k/(2s_k+k)}$$

and we follow the strategy developed by Cadre et al. (2017). In particular Lemma 6.1 of this paper implies (using Itô's isometry combined with Theorem 2.5 in Tsybakov (2009)) that the problem boils down to find a finite family of functions $\{g_\omega\}_{\omega\in\mathcal{W}}$ with cardinal $|\mathcal{W}| \geq 2$ that satisfies the following assumptions:

(i) the null function $0 \in \{g_\omega\}_{\omega\in\mathcal{W}}$.

(ii) for any $\omega \in \mathcal{W}$, the function $g_\omega \in \mathcal{H}_\ell(s_\ell, \Lambda_\ell, M)$

(iii) there exists $\kappa_* > 0$ such that for $\omega \neq \omega'$, $\|g_\omega - g_{\omega'}\|_{\Delta_\ell} \geq 2\kappa_*\phi_n(s, \Lambda, M)$

(iv) there exists $0 < \alpha < 1/8$ such that

$$\frac{n}{|\mathcal{W}|}\sum_{\omega\in\mathcal{W}} \|g_\omega\|^2_{\Delta_\ell} \leq 2\alpha\log(|\mathcal{W}|).$$

Under these assumptions, the lower-bound (10) holds.

Notation. Now we consider $0 < \alpha < 1/8$,

$$\kappa_*^2 = \frac{1}{32}\left(\frac{\|\psi\|^2_{[-1,1]}}{2}\right)^L c_1^2$$

and

$$c_1 = \min\left\{1, (2L\lambda^*)^{-1}, \left(\alpha\log(2)(8\cdot 2^L M^2)^{-1}\right)^{1/2}\right\}.$$

Here, we construct a finite set of functions used in the rest of the proof. We consider the function $\psi : \mathbb{R} \to \mathbb{R}$ defined, for any $u \in \mathbb{R}$ by

$$\psi(u) = \exp(-1/(1-u^2))\mathbf{I}_{(-1,1)}(u).$$

Note that, since the function ψ is infinitely differentiable with compact support, we have:

$$\lambda^* = \max_{0<s<s^*}\sup_{x\neq y}\frac{\left|\psi^{(\lfloor s\rfloor)}(x) - \psi^{(\lfloor s\rfloor)}(y)\right|}{|x-y|^s} < +\infty.$$

We consider the bandwidth

$$h = \left(\frac{M^2}{\Lambda_\ell^2 n}\right)^{\frac{1}{2s_\ell+\ell}}$$

and we set $R = 1/(2h)$. We assume, without loss of generality, that R is an integer and $nh^\ell \geq 1$. Let $\mathcal{R} = \{0, \ldots, R-1\}^\ell$ and define, for any $r = (r_1, \ldots, r_\ell) \in \mathcal{R}$, the function $\phi_r : \Delta_\ell \to \mathbb{R}$ by:

$$\phi_r(y) = \prod_{i=1}^{\ell} \psi\left(\frac{y_i - x_i^{(r)}}{h}\right).$$

where $x_i^{(r)} = (2r_i + 1)h$. Finally, for any $w : \mathcal{R} \to \{0, 1\}$ we define:

$$g_w = \rho_n \sum_{r \in \mathcal{R}} w(r)\phi_r$$

where

$$\rho_n = c_1 \Lambda_\ell h^{s_\ell} = c_1 \frac{M}{\sqrt{nh^\ell}}.$$

Proof of (ii). Set $w : \mathcal{R} \to \{0, 1\}$. The following property can be readily verified:

$$\|g_w\|_{\Delta_\ell}^2 = |w| \, \|\psi\|_{[-1,1]}^{2\ell} \rho_n^2 h^\ell \qquad \text{where} \qquad |w| = \left(\sum_{r \in \mathcal{R}} w(r)\right) \leq R^\ell = \frac{1}{2^\ell h^\ell}.$$

This implies that

$$\|g_w\|_{\Delta_\ell}^2 \leq \left(\frac{\|\psi\|_{[-1,1]}^2}{2}\right)^\ell \rho_n^2 \leq c_1^2 \frac{M^2}{nh^\ell} \leq \ell! M^2. \tag{18}$$

Moreover note that, for any $y \in \Delta_\ell$ and $\alpha = (\alpha_1, \ldots, \alpha_\ell)$ such that $|\alpha| = \lfloor s_\ell \rfloor$, we have:

$$D^\alpha \phi_r(y) = \frac{1}{h^{|\alpha|}} \prod_{i=1}^{\ell} \psi^{(\alpha_i)}\left(\frac{y_i - x_i^r}{h}\right)$$

which implies that, for any $z \in \Delta_\ell$ we have

$$\begin{aligned}
|D^\alpha \phi_r(y) - D^\alpha \phi_r(z)| &\leq \frac{\|\psi\|_\infty^{\ell-1}}{h^{|\alpha|}} \sum_{i=1}^{\ell} \left| \psi^{(\alpha_i)}\left(\frac{y_i - x_i^r}{h}\right) - \psi^{(\alpha_i)}\left(\frac{z_i - x_i^r}{h}\right) \right| \\
&\leq \frac{\lambda^*}{h^{s_\ell}} \sum_{i=1}^{\ell} |y_i - z_i|^{s_\ell - \lfloor s_\ell \rfloor} \\
&\leq \frac{\ell \lambda^*}{h^{s_\ell}} |y - z|^{s_\ell - \lfloor s_\ell \rfloor}
\end{aligned}$$

This also implies, since the function ψ vanishes outside $(-1,1)$, that

$$\begin{aligned}|D^\alpha g_w(y) - D^\alpha g_w(z)| &\leq (2\ell\lambda^*)\,\frac{\rho_n}{h^{s_\ell}}|y-z|^{s_\ell - \lfloor s_\ell \rfloor} \\ &\leq c_1\,(2\ell\lambda^*)\,\Lambda_\ell |y-z|^{s_\ell - \lfloor s_\ell \rfloor} \\ &\leq \Lambda_\ell |y-z|^{s_\ell - \lfloor s_\ell \rfloor}. \end{aligned} \tag{19}$$

Using (18) and (19), we deduce that g_w belongs to $\mathcal{H}_\ell(s_\ell, \Lambda_\ell, M)$ and then (ii) is fulfilled.

Proof of (i) and (iii). Using Lemma 2.9 of Tsybakov (2009), there exists a set $\mathcal{W} \subset \{w : \mathcal{R} \to \{0,1\}\}$ such that the null function belongs to $\mathcal{W}$, $\log_2 |\mathcal{W}| \geq R^\ell/8$ and

$$\forall w \neq w' \in \mathcal{W},\ \sum_{r\in\mathcal{R}} |w(r) - w'(r)| \geq R^\ell/8.$$

Let $w, w' \in \mathcal{W}$ such that $w \neq w'$. We have

$$\begin{aligned}\|g_w - g_{w'}\|_{\Delta_\ell}^2 =& \rho_n^2 \sum_{r\in\mathcal{R}} (w(r) - w'(r))^2 \|\phi_r\|^2 \\ =& \rho_n^2 \sum_{r\in\mathcal{R}} |w(r) - w'(r)| h^\ell \|\psi\|^{2\ell} \\ \geq& \rho_n^2 h^\ell \|\psi\|^{2\ell} R^\ell/8 \\ \geq& \frac{1}{8}\left(\frac{\|\psi\|_{[-1,1]}^2}{2}\right)^L c_1^2 \Lambda_\ell^2 h^{2s_\ell} \\ \geq& 4\kappa_*^2 \phi_n^2(s,\Lambda,M).\end{aligned}$$

Then Assumptions (i) and (iii) are fulfilled.

Proof of (iv). Using (18), we deduce that using the definition of c_1

$$\begin{aligned}\frac{n}{|\mathcal{W}|}\sum_{\omega\in\mathcal{W}} \|g_\omega\|_{\Delta_\ell}^2 \leq& c_1^2 M^2 h^{-\ell} \\ \leq& \frac{\alpha}{8}\,(2h)^{-\ell}\log(2) \\ \leq& \alpha \log|\mathcal{W}|.\end{aligned}$$

Then Assumption (iv) is fulfilled.

5.2 Proof of Theorem 2

We have using (12) and (15) that

$$\left(\mathbf{E}|\hat{m}(W)-m(W)|^2\right)^{1/2} \leq \kappa_0 n^{-1/2} + \sum_{\ell=1}^{L} \frac{1}{\sqrt{\ell!}} \left(\mathbf{E}\left\|\hat{f}^{(\ell)}_{\hat{h}_\ell} - f_\ell\right\|^2_{\Delta_\ell}\right)^{1/2}. \quad (20)$$

Let $\ell \in \{1,\dots,L\}$. Let $h \in \mathbf{H}_\ell$. We have

$$\begin{aligned}
\left\|\hat{f}^{(\ell)}_{\hat{h}_\ell} - f_\ell\right\|_{\Delta_\ell} &\leq \left\|\hat{f}^{(\ell)}_{h} - f_\ell\right\|_{\Delta_\ell} + \left\|\hat{f}^{(\ell)}_{\hat{h}_\ell \vee h} - \hat{f}^{(\ell)}_{h}\right\|_{\Delta_\ell} + \left\|\hat{f}^{(\ell)}_{\hat{h}_\ell} - \hat{f}^{(\ell)}_{\hat{h}_\ell \vee h}\right\|_{\Delta_\ell} \\
&\leq \left\|\hat{f}^{(\ell)}_{h} - f_\ell\right\|_{\Delta_\ell} + B(\ell,h) + M(\ell,\hat{h}_\ell) \\
&\quad + B(\ell,\hat{h}_\ell) + M(\ell,h) + 2M(\ell,\hat{h}_\ell \vee h) \\
&\leq \left\|\hat{f}^{(\ell)}_{h} - f_\ell\right\|_{\Delta_\ell} + 4(B(\ell,h) + M(\ell,h)).
\end{aligned}$$

Then we have

$$\left(\mathbf{E}\left\|\hat{f}^{(\ell)}_{\hat{h}_\ell} - f_\ell\right\|^2_{\Delta_\ell}\right)^{1/2} \leq \left(\mathbf{E}\left\|\hat{f}^{(\ell)}_{h} - f_\ell\right\|^2_{\Delta_\ell}\right)^{1/2} + 4M(\ell,h) + 4\left(\mathbf{E}B^2(\ell,h)\right)^{1/2}. \quad (21)$$

Note that using (17), we have

$$\left(\mathbf{E}\left\|\hat{f}^{(\ell)}_{h} - f_\ell\right\|^2_{\Delta_\ell}\right)^{1/2} \leq C\left(\left\|\mathbf{E}\hat{f}^{(\ell)}_{h} - f_\ell\right\|_{\Delta_\ell} + M(\ell,h)\right). \quad (22)$$

In the following, we will demonstrate that

$$\left(\mathbf{E}B^2(\ell,h)\right)^{1/2} \leq C\left(\max_{h'\leq h}\left\|\mathbf{E}\hat{f}^{(\ell)}_{h'} - f_\ell\right\|_{\Delta_\ell} + M(\ell,h)\right) + O\left(\frac{\log n}{n}\right). \quad (23)$$

Combining (21) with (22) and (23), we obtain that

$$\left(\mathbf{E}\left\|\hat{f}^{(\ell)}_{\hat{h}_\ell} - f_\ell\right\|^2_{\Delta_\ell}\right)^{1/2} \leq C\left(\max_{h'\leq h}\left\|\mathbf{E}\hat{f}^{(\ell)}_{h'} - f_\ell\right\|_{\Delta_\ell} + M(\ell,h)\right) + O\left(\frac{\log n}{n}\right).$$

Theorem 2 is then a direct consequence of the above inequality and (20).

Proof of (23) Now let us control $B(\ell,h)$ for $h \in \mathbf{H}_\ell$. We have

$$\begin{aligned}
B(\ell,h) &= \max_{h'\leq h}\left\{\|\hat{f}^{(\ell)}_{h'} - \hat{f}^{(\ell)}_{h\vee h'}\| - M(\ell,h') - M(\ell,h\vee h')\right\}_+ \\
&\leq \max_{h'\leq h}\left\{\{\|\hat{f}^{(\ell)}_{h'} - \mathbf{E}\hat{f}^{(\ell)}_{h'}\| - M(\ell,h')\}_+\right\} + \{\|\hat{f}^{(\ell)}_{h} - \mathbf{E}\hat{f}^{(\ell)}_{h}\| - M(\ell,h)\}_+ \\
&\quad + 2\max_{h'\leq h}\|\mathbf{E}\hat{f}^{(\ell)}_{h'} - f_\ell\|.
\end{aligned}$$

Then

$$\left(\mathbf{E}B^2(\ell,h)\right)^{1/2} \leq 2\max_{h'\leq h}\|\mathbf{E}\hat{f}^{(\ell)}_{h'} - f_\ell\| + 2\left(\#(\mathbf{H}_\ell)\max_{h'\leq h}A_\ell(h')\right)^{1/2} \quad (24)$$

where

$$A_\ell(h') = \mathbf{E}\left[\left\{\left\|\hat{f}^{(\ell)}_{h'} - \mathbf{E}\hat{f}^{(\ell)}_{h'}\right\|_2 - M(\ell,h')\right\}^2_+\right].$$

We have

$$\hat{f}^{(\ell)}_{h'} - \mathbf{E}\hat{f}^{(\ell)}_{h'} = \sum_{k=1}^{L}\mathbf{U}_{k,\ell} + \mathbf{U}_\ell$$

where

$$k!\mathbf{U}_{k,\ell}(t) = \frac{1}{n}\sum_{i=1}^{n}\left\{\Theta_{i,k}\xi_{i,\ell}(t,h') - \mathbf{E}\left(\Theta_{i,k}\xi_{i,\ell}(t,h')\right)\right\}$$

and

$$\mathbf{U}_\ell(t) = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\xi_{i,\ell}(t,h').$$

Using these notations we have:

$$A_\ell(h') \leq 2^{L+1}\left(\sum_{k=0}^{L}\mathbf{E}\left\{\|\mathbf{U}_{k,\ell}\|_2 - T(k,\ell)\right\}^2_+ + \mathbf{E}\left\{\|\mathbf{U}_\ell\|_2 - T(\ell)\right\}^2_+\right),$$

where

$$T(\ell) = \frac{\mu_4\left(1+4\sqrt{\log(1/(h')^\ell)}\right)}{\sqrt{n(h')^\ell}}$$

and

$$T(k,\ell) = \frac{\mathfrak{c}_4(k)Mk!\left(1+4\sqrt{\log(1/(h')^\ell)}\right)}{\sqrt{n(h')^\ell}}.$$

Now note that

$$\left(\mathbf{E}\varepsilon_i^4\right)^{1/4} = \mu_4 < \infty$$

and for $k \in \{0,\ldots,L\}$

$$\left(\mathbf{E}\Theta^4_{i,k}\right)^{1/4} \leq \mathfrak{c}_4(k)Mk!.$$

Using Lemma 2 with $U = U_\ell$ (respectively $U = k!U_{k,\ell}$), $T = T(\ell)$ (respectively $T = T(k,\ell)$) and $\chi_i = \varepsilon_i$ (respectively $\chi = \Theta_{i,k}$), we deduce that for all for $h' \leq h$

$$A_\ell(h') \leq Cn^{-1}.$$

This implies that

$$\#(\mathbf{H}_\ell)\max_{h'\leq h}A_\ell(h') \leq C(\log n)n^{-1}. \quad (25)$$

Now (24) and (25) entail (23).

5.3 Proof of Theorem 3.

Let $s = (s_1, \dots, s_L) \in [\tau, s^*)^L$, $\Lambda \in (0, +\infty)^L$, $M > 0$ and $f \in \mathfrak{M}_{s,\Lambda,M}$. Define for $\ell \in \{1, \dots, L\}$

$$k_\ell = \left\lfloor \frac{1}{2s_\ell + \ell} \log\left(\frac{n}{\log n}\right) \right\rfloor.$$

For n large enough, we have $h_\ell = e^{-k_\ell} \in \mathbf{H}_\ell$. Using (16) and (7), Theorem 2 implies that

$$\begin{aligned}
R_2(\hat{m}, m) &\leq C \sum_{\ell=1}^{L} \left\{ \max_{h' \leq h_\ell} (h')^{s_\ell} + M(\ell, h_\ell) \right\} + \Upsilon_2 \left(\frac{\log n}{n}\right)^{1/2} \\
&\leq C \sum_{\ell=1}^{L} \left\{ h_\ell^{s_\ell} + \left(\frac{\log n}{n h_\ell^\ell}\right)^{1/2} \right\} + \Upsilon_2 \left(\frac{\log n}{n}\right)^{1/2} \\
&\leq C \tilde{\phi}_n(s) + \Upsilon_2 \left(\frac{\log n}{n}\right)^{1/2} \\
&\leq C \tilde{\phi}_n(s).
\end{aligned}$$

where C may depend on s, Λ and M. Since C does not depend on $m \in \mathfrak{M}(s, \Lambda, M)$, this ends the proof.

5.4 Proof of Lemma 1.

We have

$$\begin{aligned}
\left(\mathbf{E}|\xi_\ell(t,h)|^{2r}\right)^{1/r} &\leq \mathfrak{c}_\ell^2(2r)\mathbf{E}|\xi_\ell(t,h)|^2 \\
&\leq \mathfrak{c}_\ell^2(2r)\ell! \int_{\Delta_\ell} \left(K_h^{(\ell)}(t,u)\right)^2 \mathrm{d}u \\
&\leq \mathfrak{c}_\ell^2(2r)\ell! 2^\ell \|\mathbf{k}\|_{[0,1]}^{2\ell} h^{-\ell} \\
&\leq b_{\ell,r} h^{-\ell}.
\end{aligned}$$

Moreover we have

$$\begin{aligned}
\mathbf{E}\left(|\xi_\ell(t,h)|^r \mathbf{I}_{|\xi_\ell(t,h)|>\varphi}\right) &\leq \left(\mathbf{E}|\xi_\ell(t,h)|^{2r}\right)^{1/2} \left(\mathbf{P}\left(|\xi_\ell(t,h)| > \varphi\right)\right)^{1/2} \\
&\leq (b_{\ell,r} h^{-\ell})^{r/2} \left(\frac{\mathbf{E}|\xi_\ell(t,h)|^{2q}}{\varphi^{2q}}\right)^{1/2} \\
&\leq (b_{\ell,r})^{r/2} (b_{\ell,q})^{q/2} h^{-\ell(r+q)/2} \varphi^{-q}.
\end{aligned}$$

5.5 Proof of Lemma 2

Since h is fixed, we simplify the notation and use in the proof $\xi_{i,\ell}(t,h) = \xi_{i,\ell}(t)$ and $\xi_\ell(t,h) = \xi_\ell(t)$. Now, we have for $k \geq 1$:

$$\mathbf{U} = \bar{\eta}_0(t) + \eta_1(t) + \eta_2(t) + \eta_3(t)$$

where

$$\begin{aligned}
\bar{\eta}_0(t) &= \eta_0(t) - \mathbf{E}\eta_0(t) \qquad \eta_0(t) = \frac{1}{n}\sum_{i=1}^{n} \chi_i \mathbf{I}_{|\chi_i| \leq \varphi(n)} \xi_{i,\ell}(t) \mathbf{I}_{|\xi_{i,\ell}| \leq \psi(n)} \\
\eta_1(t) &= \mathbf{E}\eta_0(t) - \mathbf{E}(\chi \xi_\ell(t)) \\
\eta_2(t) &= \frac{1}{n}\sum_{i=1}^{n} \chi_i \mathbf{I}_{|\chi_i| > \varphi(n)} \xi_{i,\ell}(t) \\
\eta_3(t) &= \frac{1}{n}\sum_{i=1}^{n} \chi_i \mathbf{I}_{|\chi_i| \leq \varphi(n)} \xi_{i,\ell}(t) \mathbf{I}_{|\xi_{i,\ell}(t)| > \psi(n)}.
\end{aligned}$$

where for any $\ell = 1, \ldots, L$:

$$\psi(n) = n^\alpha \quad \text{with} \quad \alpha = \alpha(\tau, \ell) = \frac{\tau + \ell}{2\tau + 4\ell}$$

and

$$\varphi(n) = n^\beta \quad \text{with} \quad \beta = 1/2 - \alpha > 0.$$

Note that both α and β are positive numbers.

5.5.1 Control of η_1

We have using Cauchy-Schwarz, Markov inequality

$$\begin{aligned}
|\eta_1(t)| &= \left| -\mathbf{E}\left(\chi \xi_\ell(t) \mathbf{I}_{|\chi| > \varphi(n)}\right) - \mathbf{E}\left(\chi \xi_\ell(t) \mathbf{I}_{|\chi| \leq \varphi(n)} \mathbf{I}_{|\xi_\ell(t)| > \psi(n)}\right) \right| \\
&\leq \mathbf{E}\left(|\chi \xi_\ell(t)| \mathbf{I}_{|\chi| > \varphi(n)}\right) + \mathbf{E}\left(|\chi \xi_\ell(t)| \mathbf{I}_{|\chi| \leq \varphi(n)} \mathbf{I}_{|\xi_\ell(t)| > \psi(n)}\right) \qquad (26)
\end{aligned}$$

Note that we have

$$\begin{aligned}
\mathbf{E}\left(|\chi \xi_\ell(t)| \mathbf{I}_{|\chi| > \varphi(n)}\right) &\leq \left(\mathbf{E}\left(|\chi|^2 \mathbf{I}_{|\chi| > \varphi(n)}\right) \mathbf{E}|\xi_\ell(t)|^2\right)^{1/2} \\
&\leq b_{\ell,1} h^{-\ell} a_2 \left(\mathbf{P}\left(|\chi| > \varphi(n)\right)\right)^{1/4} \\
&\leq b_{\ell,1} a_2 \left(\mathbf{E}|\chi|\right)^{8/\beta})^{1/4} \\
&\leq b_{\ell,1} a_2 (a_{4/\beta})^{2/\beta} h^{-\ell} (\varphi(n))^{-2/\beta} \\
&\leq C n^{-1}. \qquad (27)
\end{aligned}$$

Moreover since $h \in \mathbf{H}_\ell$, using Lemma 1 with $q = \frac{8\tau+6\ell}{\tau}$, we have

$$\begin{aligned}\left(\mathbf{E}\left(|\chi\xi_\ell(t)|\mathbf{I}_{|\chi|\leq\varphi(n)}\mathbf{I}_{|\xi_\ell(t)|>\psi(n)}\right)\right)^2 \leq & a_1^2\mathbf{E}|\xi_\ell(t)|^2\mathbf{I}_{|\xi_\ell(t)|>\psi(n)} \\ & \leq (b_{\ell,2})(b_{\ell,q})^{q/2}(\psi(n))^{-q}h^{-\ell(2+q)/2} \\ & \leq Cn^{-2}. \end{aligned} \tag{28}$$

Now using (26), (27) and (28), we finally obtain

$$\mathbf{E}\left(\|\eta_1\|_{\Delta_\ell}^2\right) \leq Cn^{-1}.$$

5.5.2 Control of η_2

Define

$$\mu_n(t) = \mathbf{E}\left(\chi\mathbf{I}_{|\chi|>\varphi(n)}\xi_\ell(t)\right).$$

We have

$$\mathbf{E}\left(\|\eta_2\|_{\Delta_\ell}^2\right) \leq 2(A+B)$$

where using Lemma 1 with $q = 4/\beta$

$$\begin{aligned} A &= \frac{1}{n}\int_{\Delta_\ell}\mathbf{E}\left(\chi\mathbf{I}_{|\chi|>\varphi(n)}\xi_\ell(t) - \mu_n(t)\right)^2 dt \\ &\leq \frac{1}{n}\int_{\Delta_\ell}\left(\mathbf{E}\left(\chi^4\mathbf{I}_{|\chi|>\varphi(n)}\right)\mathbf{E}\xi_\ell^4(t)\right)^{1/2} dt \\ &\leq n^{-1}a_4^2\mathbf{P}\left(|\chi|>\varphi(n)\right)^{1/4} b_{\ell,2}h^{-\ell} \\ &\leq a_4^2 b_{\ell,2}\mathbf{P}\left(|\chi|>\varphi(n)\right)^{1/4} \\ &\leq a_4^2 b_{\ell,2}(a_{q/2})^{q/4}\left(\varphi(n)\right)^{-q/4} \\ &\leq Cn^{-1} \end{aligned}$$

and following (27)

$$B = \frac{1}{n}\int_{\Delta_\ell}\mu_n^2(t)dt \leq Cn^{-1}.$$

5.5.3 Control of η_3

Define

$$\nu_n(t) = \mathbf{E}\left(\chi\mathbf{I}_{|\chi|\leq\varphi(n)}\xi_\ell(t)\mathbf{I}_{|\xi_\ell(t)|>\psi(n)}\right).$$

We have

$$\mathbf{E}\left(\|\eta_3\|_{\Delta_\ell}^2\right) \leq 2(A+B)$$

where

$$A = \frac{1}{n}\int_{\Delta_\ell} \mathbf{E}\left(\chi \mathbf{I}_{|\chi|\leq\varphi(n)}\xi_\ell(t)\mathbf{I}_{|\xi_\ell(t)|>\psi(n)} - \nu_n(t)\right)^2 dt$$

and

$$B = \frac{1}{n}\int_{\Delta_\ell} \nu_n^2(t)dt$$

Note that using similar arguments as above with $q = (8\tau + 4\ell)/\tau$,

$$\begin{aligned} A &\leq \frac{1}{n}\int_{\Delta_\ell} \left(\mathbf{E}\left(\chi^4\right)\mathbf{E}\left(\xi_\ell^4(t)\mathbf{I}_{|\xi_\ell(t)|>\psi(n)}\right)\right)^{1/2} dt \\ &\leq a_2^2 b_{\ell,4}(b_{\ell,q})^{q/4}(\psi(n))^{-q/2}h^{-\ell q/4} \\ &\leq Cn^{-1} \end{aligned}$$

and following (28) $B \leq Cn^{-1}$.

5.5.4 Control of $\bar{\eta}_0$

We have to bound

$$\begin{aligned} \mathbf{E}\left\{\|\bar{\eta}_0\|_2 - T\right\}_+^2 &\leq \int_0^{+\infty} \mathbf{P}\left(\|\bar{\eta}_0\|_2 - T > \sqrt{u}\right) du \\ &\leq 2\int_0^{+\infty} u\mathbf{P}\left(\|\bar{\eta}_0\|_2 > u + T\right) du. \end{aligned}$$

Note that, using duality arguments, there exists a countable set $\mathcal{S}$ of functions $s \in \mathbb{L}^2(\Delta_\ell)$ such that $\|s\|_2 \leq 1$ and

$$\begin{aligned} \|\bar{\eta}_0\|_2 &= \sup_{s\in\mathcal{S}} \int_{\Delta_\ell} s(t)\bar{\eta}_0(t)dt \\ &= \frac{2\varphi(n)\psi(n)}{n} Z = \frac{2Z}{\sqrt{n}} \end{aligned}$$

where

$$Z = \sup_{s\in\mathcal{S}} \sum_{i=1}^n X_{i,s}$$

and, for $s \in \mathcal{S}$, we have:

$$X_{i,s} = \int_{\Delta_\ell} s(t)X_i(t)dt \qquad \text{with} \qquad X_i(t) = g_i(t) - \mathbf{E}g_i(t)$$

and

$$g_i(t) = \frac{1}{2\sqrt{n}}\chi_i \mathbf{I}_{\{|\chi_i|\leq\varphi(n)\}}\xi_{i,\ell}(t)\mathbf{I}_{\{|\xi_{i,\ell}(t)|\leq\psi(n)\}}.$$

Note that we have both $\mathbf{E}(X_{i,s}) = 0$ and $\|X_{i,s}\|_\infty \leq 1$. Now, let us control:

$$v = 2\mathbf{E}\left(\sup_{s\in\mathcal{S}} \sum_{i=1}^{n} X_{i,s}\right) + n \sup_{s\in\mathcal{S}} \mathbf{E}X_{1,s}^2.$$

Using Cauchy-schwarz's inequality and Fubini's theorem we obtain:

$$\begin{aligned}\mathbf{E}X_{1,s}^2 &\leq \left(\int_{\Delta_\ell} s^2(t)dt\right)\left(\int_{\Delta_\ell} \mathbf{E}X_i^2(t)dt\right) \\ &\leq \int_{\Delta_\ell} \mathbf{E}g_i^2(t)dt\end{aligned}$$

We have

$$\int_{\Delta_\ell} \mathbf{E}g_i^2(t)dt \leq \frac{a_2^2 b_{\ell,2}}{4nh^\ell}$$

and

$$\begin{aligned}\mathbf{E}\left(\sup_{s\in\mathcal{S}} \sum_{i=1}^{n} X_{i,s}\right) &= \mathbf{E}\left\|\sum_{i=1}^{n} X_i(\cdot)\right\|_{\Delta_\ell} \\ &\leq \left(\mathbf{E}\left\|\sum_{i=1}^{n} X_i(\cdot)\right\|_{\Delta_\ell}^2\right)^{1/2} \\ &\leq \frac{1}{4}\left(\int_{\Delta_\ell} \mathbf{E}\chi^2\xi_\ell^2(t)\mathrm{d}t\right)^{1/2} \\ &\leq \frac{a_2\sqrt{b_{\ell,2}}}{4h^{\ell/2}}\end{aligned}$$

Combining the previous results we have:

$$v \leq \theta_n + \sqrt{\theta_n} \quad \text{with} \quad \theta_n = \frac{a_2^2 b_{\ell,2}}{4h^\ell}.$$

Define

$$T = (1+\delta)\frac{a_2\sqrt{b_{\ell,2}}}{2\sqrt{nh^\ell}} = (1+\delta)\sqrt{\frac{\theta_n}{n}}.$$

We have:

$$\mathbf{P}\left(\|\bar{\eta}_0\|_2 > u + T\right) \leq \mathbf{P}\left(Z - \mathbf{E}Z > \frac{\sqrt{n}u}{2} + \frac{\delta}{2}\sqrt{\theta_n}\right).$$

Define:

$$\mathfrak{a} = \frac{\delta\sqrt{n\theta_n}}{2}, \qquad \mathfrak{b} = \delta^2\theta_n/4, \qquad \mathfrak{c} = \frac{\sqrt{n}}{3}, \quad \mathfrak{d} = 2\theta_n + 2\sqrt{\theta_n}(1+\delta/6).$$

Using Bousquet's inequality we have:

$$\mathbf{P}\left(Z - \mathbf{E}Z > \frac{\sqrt{n}u}{2} + \frac{\delta}{2}\sqrt{\theta_n}\right) \leq C_n(u)D_n(u)$$

where

$$C_n(u) = \exp\left(-\frac{nu^2}{4(\mathfrak{c}u + \mathfrak{d})}\right)$$

and

$$D_n(u) = \exp\left(-\frac{\mathfrak{a}u + \mathfrak{b}}{\mathfrak{c}u + \mathfrak{d}}\right). \tag{29}$$

Since $\mathfrak{a}\mathfrak{d} - \mathfrak{b}\mathfrak{c} > 0$, we have, $D_n(u) \leq D_n(0)$, that is:

$$D_n(u) \leq \exp\left(-\frac{\mathfrak{b}}{\mathfrak{d}}\right) \leq \exp\left(-\frac{\delta^2\theta_n}{4(2\theta_n + 2\sqrt{\theta_n}(1+\delta/6))}\right).$$

Since for n large enough $(1+\delta/6) \leq \sqrt{\theta_n}$ we have $D_n(u) \leq Ch^\ell$. Moreover we have doing the change of variables $v = \sqrt{nh^\ell}u$

$$\begin{aligned}
\mathbf{E}\left\{\|\bar{\eta}_0\|_2 - T\right\}_+^2 &\leq 2\int_0^{+\infty} u\mathbf{P}\left(\|\bar{\eta}_0\|_2 > u + T\right) du \\
&\leq Ch^\ell \int_0^{+\infty} uC_n(u)du \\
&\leq \frac{C}{n}\int_0^{+\infty} v\exp\left(-\frac{Cv^2}{1+v}\right) \mathrm{d}v.
\end{aligned}$$

This implies that:

$$\mathbf{E}\left\{\|\bar{\eta}_0\|_2 - T\right\}_+^2 \leq Cn^{-1}.$$

Combining results of Sections 5.5.1, 5.5.2, 5.5.3 and 5.5.4, we obtain (11)

Acknowledgements

The authors have been supported by Fondecyt projects 1171335 and 1190801, and Mathamsud 18-MATH-07.

References

Jean-Marc Azaïs and Jean-Claude Fort. Remark on the finite-dimensional character of certain results of functional statistics. *C. R. Math. Acad. Sci. Paris*, 351(3-4):139–141, 2013. ISSN 1631-073X. doi: 10.1016/j.crma.2013.02.004. URL https://doi.org/10.1016/j.crma.2013.02.004.

Karine Bertin, Salima El Kolei, and Nicolas Klutchnikoff. Adaptive density estimation on bounded domains. *Accepted for publication in Annales de l'institut Henri Poincaré: Probabilités et Statistiques*, 2018.

Gérard Biau, Frédéric Cérou, and Arnaud Guyader. Rates of convergence of the functional k-nearest neighbor estimate. *IEEE Trans. Inform. Theory*, 56(4):2034–2040, 2010. ISSN 0018-9448. doi: 10.1109/TIT.2010.2040857. URL https://doi.org/10.1109/TIT.2010.2040857.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. ISBN 978-0-19-953525-5. doi: 10.1093/acprof:oso/9780199535255.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199535255.001.0001. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

Olivier Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris*, 334(6):495–500, 2002. ISSN 1631-073X. doi: 10.1016/S1631-073X(02)02292-6. URL https://doi.org/10.1016/S1631-073X(02)02292-6.

Benoît Cadre and Lionel Truquet. Nonparametric regression estimation onto a Poisson point process covariate. *ESAIM Probab. Stat.*, 19:251–267, 2015. ISSN 1292-8100. doi: 10.1051/ps/2014023. URL https://doi.org/10.1051/ps/2014023.

Benoît Cadre, Nicolas Klutchnikoff, and Gaspar Massiot. Minimax regression estimation for Poisson coprocess. *ESAIM Probab. Stat.*, 21:138–158, 2017. ISSN 1292-8100. doi: 10.1051/ps/2017004. URL https://doi.org/10.1051/ps/2017004.

T. Tony Cai and Peter Hall. Prediction in functional linear regression. *Ann. Statist.*, 34(5):2159–2179, 2006. ISSN 0090-5364. doi: 10.1214/009053606000000830. URL https://doi.org/10.1214/009053606000000830.

Gaëlle Chagny and Angelina Roche. Adaptive estimation in the functional nonparametric regression model. *J. Multivariate Anal.*, 146:105–118, 2016.

ISSN 0047-259X. doi: 10.1016/j.jmva.2015.07.001. URL https://doi.org/10.1016/j.jmva.2015.07.001.

Fabienne Comte and Jan Johannes. Adaptive functional linear regression. *Ann. Statist.*, 40(6):2765–2797, 2012. ISSN 0090-5364. doi: 10.1214/12-AOS1050. URL https://doi.org/10.1214/12-AOS1050.

Christophe Crambes, Alois Kneip, and Pascal Sarda. Smoothing splines estimators for functional linear regression. *Ann. Statist.*, 37(1):35–72, 2009. ISSN 0090-5364. doi: 10.1214/07-AOS563. URL https://doi.org/10.1214/07-AOS563.

Giulia Di Nunno, Bernt Øksendal, and Frank Proske. *Malliavin calculus for Lévy processes with applications to finance*. Universitext. Springer-Verlag, Berlin, 2009. ISBN 978-3-540-78571-2. doi: 10.1007/978-3-540-78572-9. URL https://doi.org/10.1007/978-3-540-78572-9.

Frédéric Ferraty and Philippe Vieu. *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York, 2006. ISBN 0-387-30369-3; 978-0387-30369-7. Theory and practice.

Frédéric Ferraty, André Mas, and Philippe Vieu. Nonparametric regression on functional data: inference and practical aspects. *Aust. N. Z. J. Stat.*, 49 (3):267–286, 2007. ISSN 1369-1473. doi: 10.1111/j.1467-842X.2007.00480.x. URL https://doi.org/10.1111/j.1467-842X.2007.00480.x.

Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3):1608–1632, 2011. ISSN 0090-5364. doi: 10.1214/11-AOS883. URL http://dx.doi.org/10.1214/11-AOS883.

Alexander Goldenshluger and Oleg Lepski. On adaptive minimax density estimation on R^d. *Probab. Theory Related Fields*, 159(3-4):479–543, 2014. ISSN 0178-8051. doi: 10.1007/s00440-013-0512-1. URL http://dx.doi.org/10.1007/s00440-013-0512-1.

Oleg Lepski. Adaptive estimation over anisotropic functional classes via oracle approach. *Ann. Statist.*, 43(3):1178–1242, 2015. ISSN 0090-5364. doi: 10.1214/14-AOS1306. URL https://doi.org/10.1214/14-AOS1306.

W. V. Li and Q.-M. Shao. Gaussian processes: inequalities, small ball probabilities and applications. In *Stochastic processes: theory and methods*, volume 19 of *Handbook of Statist.*, pages 533–597. North-Holland,

Amsterdam, 2001. doi: 10.1016/S0169-7161(01)19019-X. URL https://doi.org/10.1016/S0169-7161(01)19019-X.

André Mas. Lower bound in regression for functional data by representation of small ball probabilities. *Electron. J. Stat.*, 6:1745–1778, 2012. ISSN 1935-7524. doi: 10.1214/12-EJS726. URL https://doi.org/10.1214/12-EJS726.

Hans-Georg Müller and Ulrich Stadtmüller. Generalized functional linear models. *Ann. Statist.*, 33(2):774–805, 2005. ISSN 0090-5364. doi: 10.1214/009053604000001156. URL https://doi.org/10.1214/009053604000001156.

Ivan Nourdin and Giovanni Peccati. *Normal approximations with Malliavin calculus*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2012. ISBN 978-1-107-01777-1. doi: 10.1017/CBO9781139084659. URL https://doi.org/10.1017/CBO9781139084659. From Stein's method to universality.

M. Rachdi and P. Vieu. Nonparametric regression for functional data: automatic smoothing parameter selection. *J. Statist. Plann. Inference*, 137(9): 2784–2801, 2007. ISSN 0378-3758. doi: 10.1016/j.jspi.2006.10.001. URL https://doi.org/10.1016/j.jspi.2006.10.001.

J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition, 2005. ISBN 978-0387-40080-8; 0-387-40080-X.

Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. ISBN 978-0-387-79051-0. doi: 10.1007/b13794. URL https://doi.org/10.1007/b13794. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.