Adaptive Density Estimation on Bounded Domains

Karine Bertin * Salima El Kolei † Nicolas Klutchnikoff ‡

January 23, 2017

Abstract

We study the estimation, in $\mathbb{L}_p$-norm, of density functions defined on $[0,1]^d$. We construct a new family of kernel density estimators that do not suffer from the so-called boundary bias problem and we propose a data-driven procedure based on the Goldenshluger and Lepski approach that jointly selects a kernel and a bandwidth. We derive two estimators that satisfy oracle-type inequalities and that are also proved to be adaptive over a scale of anisotropic or isotropic Sobolev-Slobodetskii and Hölder classes. The main interest of the isotropic procedure is to obtain adaptive results without any restriction on the smoothness parameter.

Keywords. Multivariate kernel density estimation, Bounded data, Boundary bias, Adaptive estimation, Oracle inequality, Sobolev-Slobodetskii classes.

AMS Subject Classification. 62G05, 62G20.

1 Introduction

In this paper we study the classical problem of the estimation of a density function $f : \Delta_d \to \mathbb{R}$ where $\Delta_d = [0,1]^d$. We observe n independent and identically distributed random variables $X_1, \ldots, X_n$ with density f. In this context, an estimator is a measurable map $\tilde{f} : \Delta_d^n \to \mathbb{L}_p(\Delta_d)$ where $p \geq 1$ is a fixed parameter. The accuracy of $\tilde{f}$ is measured using the risk:

$$R_n^{(p,q)}(\tilde{f}, f) = \left(\mathbf{E}_f^n \|\tilde{f} - f\|_p^q\right)^{1/q},$$

where q is also a fixed parameter greater than or equal to 1 and $\mathbf{E}_f^n$ denotes the expectation with respect to the probability measure $\mathbf{P}_f^n$ of the observations. Moreover the $\mathbb{L}_p$-norm of a function $g : \Delta_d \to \mathbb{R}$ is defined by

$$\|g\|_p = \left(\int_{\Delta_d} |g(t)|^p dt\right)^{1/p}.$$

The density estimation problem is widely studied and we refer the reader to Devroye and Gyorfi (1985) and Silverman (1986) for a broadly picture of this domain of statistics. One

*CIMFAV, Universidad de Valparaíso, General Cruz 222, Valparaíso, Chile, tel/fax: 0056322303623

†ENSAI, UBL, Campus de Ker Lann, Rue Blaise Pascal - BP 37203 35172 Bruz cedex

‡IRMAR, Université de Rennes 2, CNRS, UBL Campus Villejean Place du recteur Henri le Moal CS 24307, 35043 Rennes cedex, France

of the most popular ways to estimate a density function is to use kernel density estimates introduced by Rosenblatt et al. (1956) and Parzen (1962). Given a kernel K (that is a function $K : \mathbb{R}^d \to \mathbb{R}$ such that $\int_{\mathbb{R}^d} K(x)\,\mathrm{d}x = 1$) and a bandwidth vector $h = (h_1, \ldots, h_d)$, such an estimator writes:

$$\widehat{f}_h(t) = \frac{1}{nV_h} \sum_{k=1}^{n} K\left(\frac{t - X_k}{h}\right), \qquad t \in \Delta_d$$

where $V_h = \prod_{i=1}^{d} h_i$ and u/v stands for the coordinate-wise division of the vectors u and v.

It is commonly admitted that *bandwidth selection* is the main point to estimate accurately the density function f and a lot of popular selection procedures are proposed in the literature. Among others let us point out the cross validation (see Rudemo, 1982; Bowman, 1984; Chiu, 1991) as well as the procedure developed by Goldenshluger and Lepski in a series of papers in the last few years (see Goldenshluger and Lepski, 2008, 2011, 2014, for instance) and fruitfully applied in various context.

Dealing with bounded data, the so-called *boundary bias problem* has also to be taken into account. Indeed, classical kernels suffer from a sever bias term when the underlying density function does not vanish near the boundary of their support. To overcome this drawback, several procedures have been developed: Schuster (1985), Silverman (1986) and Cline and Hart (1991) studied the reflection of the data near the boundary as well as Marron and Ruppert (1994) who proposed a previous transformation of the data. Müller (1991), Lejeune and Sarda (1992), Jones (1993), Müller and Stadtmüller (1999) and Botev et al. (2010) proposed to construct kernels which take into account the shape of the support of the density. In the same spirit, Chen (1999) studied a new class of kernels constructed using a reparametrization of the family of Beta distributions. For these methods, practical choices of bandwidth or cross-validation selection procedures have generally been proposed. Nevertheless few papers study the theoretical properties of *bandwidth selection* procedures in this context. Among others, we point out Bouezmarni and Rombouts (2010)—who study the behavior of Beta kernels with a cross validation selection procedure in a multivariate setting in the specific case of a twice differentiable density. Bertin and Klutchnikoff (2014) study a selection rule based on the Lepski's method (see Lepski, 1991) in conjunction with Beta kernels in a univariate setting and prove that the associated estimator is adaptive over Hölder classes of smoothness smaller than or equal to two.

In this paper, we aim at constructing estimation procedures that address both problems (*boundary bias* and *bandwidth selection*) simultaneously and with optimal adaptive properties in $\mathbb{L}_p$ norm ($p \geq 1$) over a large scale of function classes. To tackle the boundary bias problem, we construct a family of kernel estimators based on new asymmetric kernels, the shape of which adapts to the position of the estimation point in Δ_d. We propose two different data-driven procedures based on the Goldenshluger and Lepski approach that satisfy oracle-type inequalities (see Theorems 1 and 3). The first procedure, based on a fixed kernel, consists in selecting a bandwidth vector. It is proved (see Theorem 2) to be adaptive over anisotropic Sobolev-Slobodetskii classes with smoothness parameters $(s_1, \ldots, s_d) \in (0, \infty)^d$ smaller than the order of the kernel and with the optimal rate $n^{-\overline{s}/(2\overline{s}+1)}$ with $\overline{s} = \left(\sum_{i=1}^{d} 1/s_i\right)^{-1}$. The second procedure jointly selects a kernel (and its order) and a univariate bandwidth. Theorem 4 states that this procedure is adaptive over isotropic Sobolev-Slobodetskii classes without any restriction on the smoothness parameter $s > 0$ and achieves the optimal rate $n^{-s/(2s+d)}$. These function classes are quite large. As a consequence our adaptive results remain valid on classical Hölder classes. Such adaptive results without restrictions on the smoothness of the function to be estimated and with the optimal rates $n^{-2s/(2s+d)}$ or $n^{-\overline{s}/(2\overline{s}+1)}$ have been established only for ellipsoid function classes as in Asin and Johannes (2016), among others. For bounded data, we also mention Rousseau et al. (2010) or Autin et al. (2010) that construct adaptive estimators based on Bayesian mixtures of Beta and wavelets respectively but with an extra logarithmic term factor in the rate of convergence. Additionally note also that Beta kernel density estimators are minimax only for small smoothness (see Bertin and Klutchnikoff, 2011) and consequently neither allow us to obtain such adaptive results.

The rest of the paper is organized as follows. We construct in Section 2 the two statistical procedures. The main results of the paper are stated in Section 3 whereas their proofs are postponed to Section 4.

2 Statistical procedures

We define in Section 2.1 a large family of kernels estimators that are well-adapted to the estimation of bounded data. Two subfamilies are then considered in Sections 2.2 and 2.3 and a unique data-driven procedure is proposed in Section 2.4.

2.1 Boundary kernel estimators

Set $h^* = \exp(-\sqrt{\log n})$ and define the set of bandwidth vectors $\mathcal{H}_n = \{h = (h_1, \dots, h_d) \in (0, h^*]^d : nV_h \geq (\log n)^{2d+1}\}$ where $V_h = \prod_{i=1}^d h_i$. Define also:

$$\mathcal{W} = \left\{ W : \Delta_1 \to \mathbb{R} : \sup_{u \in \Delta_1} |W(u)| < +\infty, \quad \int_{\Delta_1} W(u)\, \mathrm{d}u = 1 \right\}.$$

For any $h \in \mathcal{H}_n$ and $W = (W_1, \dots, W_d) \in \mathcal{W}^d$, we consider:

$$\tilde{f}_{W,h}(t) = \frac{1}{n} \sum_{j=1}^n \mathcal{K}_{W,h}(t, X_j), \quad t \in \Delta_d$$

where, for $x \in \Delta_d$,

$$\mathcal{K}_{W,h}(t,x) = \prod_{i=1}^d \left(\frac{1}{h_i} W_i \left(\sigma(t_i) \frac{t_i - x_i}{h_i} \right) \right) \qquad \text{with} \qquad \sigma(u) = 2\mathbf{I}_{(1/2,1)}(u) - 1$$

where $\mathbf{I}_{(a,b)}(u)$ stands for the indicator function on the interval (a, b).

Remark 1. The family of kernel estimators $\{\tilde{f}_{W,h} : W \in \mathcal{W}^d, h \in \mathcal{H}_n\}$ is well-adapted to the estimation of densities defined on Δ_d. Indeed, it is easily seen that, if f is a continuous function, the pointwise bias of these estimators tends to 0 as h goes to 0 even if the estimation point belongs to the boundary of Δ_d. The shape of the kernel adapts to the estimation point thanks to the introduction of the function σ. In the next sections we construct two subfamilies of $\{\tilde{f}_{W,h} : W \in \mathcal{W}^d, h \in \mathcal{H}_n\}$ specifically designed for the estimation of *isotropic* or *anisotropic* functions.

2.2 Isotropic family of estimators

For $\ell \in \mathbb{N}$, we define:

$$h(\ell) = (e^{-\ell}, \dots, e^{-\ell}) \quad \text{and} \quad m(\ell) = \left\lceil \frac{\log n}{2\ell} + \frac{1}{2} \right\rceil,$$

where $\lceil b \rceil$ stands for the smallest integer larger than or equal to b. We define

$$\mathcal{L}_{\text{iso}} = \{\ell \in \mathbb{N} : h(\ell) \in \mathcal{H}_n\}.$$

For any $\ell \in \mathcal{L}_{\text{iso}}$, we consider $W(\ell) = (w_{m(\ell)}, \dots, w_{m(\ell)}) \in \mathcal{W}^d$ where the univariate kernel w_m is defined, for any $m \in \mathbb{N}$, by:

$$w_m(u) = \sum_{r=0}^{m} a_r^{(m)} u^r, \qquad u \in \Delta_1. \tag{1}$$

Here $a^{(m)} = H_m^{-1} e_0^{(m)}$ where H_m denotes the Hilbert matrix of order $m+1$ and $e_0^{(m)} = (1\,0 \dots 0)^\top \in \mathbb{R}^{m+1}$. Figure 1 represents the kernels ω_m for different values of m.

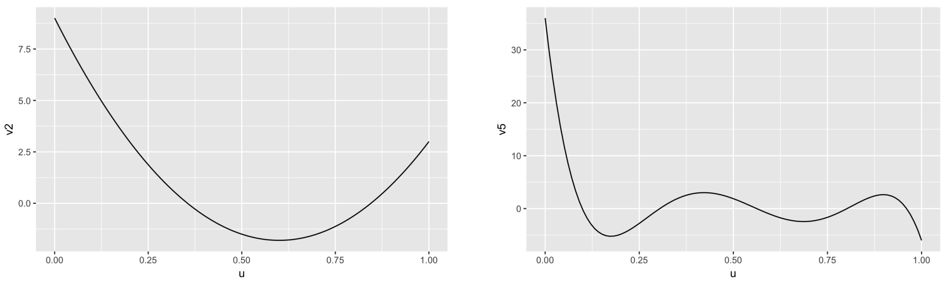

Figure 1: Plots of the kernel ω_2 (left) and ω_5 (right).

We finally define the family of estimators $\{\widehat{f}_\ell^{\text{iso}} : \ell \in \mathcal{L}_{\text{iso}}\}$ where

$$\widehat{f}_\ell^{\text{iso}} = \tilde{f}_{W(\ell),h(\ell)}.$$

Remark 2. The family $\{\widehat{f}_\ell^{\text{iso}} : \ell \in \mathcal{L}_{\text{iso}}\}$ contains kernel density estimators constructed with different kernels and bandwidths. The main idea that leads to this construction is the following: if we consider $\ell \approx \log n/(2s+d)$, then $h(\ell) \approx n^{-1/(2s+d)}$ and $m(\ell) \geq (s+1)$. In other words, the estimator $\widehat{f}_\ell^{\text{iso}}$ is constructed using a kernel of order greater than s and the usual bandwidth (that is, of the classical order) used to estimate functions with smoothness parameter s. The construction of such a class of estimators allows us to obtain adaptive estimators without any restriction on the smoothness parameter (see Theorem 4). However, arbitrary kernels of order m cannot be used to prove Theorem 3 since a control of the $\mathbb{L}_p$-norm of the kernels is required. Lemma 1 explains the construction of w_m.

Lemma 1. *Set $m \in \mathbb{N}$ and $p \in [1, +\infty]$. Let $\mathcal{W}(m) \subseteq \mathcal{W}$ be the family of kernels of order m. That is, $w \in \mathcal{W}(m)$ if:*

$$\int_{\Delta_1} w(u) u^r \, \mathrm{d}u = 0, \qquad r = 1, 2, \dots, m.$$

Then
$$w_m = \arg\min_{w \in \mathcal{W}(m)} \|w\|_2 = (m+1). \tag{2}$$
and
$$\|w_m\|_p \leq \begin{cases} m+1 & \textit{if } p \leq 2 \\ 2(m+1)^{3/2} & \textit{otherwise.} \end{cases}$$

2.3 Anisotropic family of estimators

Let $W^\circ = (W_1^\circ, \dots, W_d^\circ) \in \mathcal{W}^d$ be such that, for any $i = 1, \dots, d$, W_i° is a bounded kernel and consider $h(\ell) = (h_1(\ell), \dots, h_d(\ell))$ defined by:
$$h_i(\ell) = e^{-\ell_i}, \qquad i = 1, \dots, d$$
where $\ell \in \mathcal{L}_{\text{ani}} = \{\ell \in \mathbb{N}^d : h(\ell) \in \mathcal{H}_n\}$. We then define the anisotropic family of estimators $\{\widehat{f}_\ell^{\text{ani}} : \ell \in \mathcal{L}_{\text{ani}}\}$ by
$$\widehat{f}_\ell^{\text{ani}} = \tilde{f}_{W^\circ, h(\ell)}.$$
To make the notation similar to the isotropic case we define $W(\ell) = W^\circ, \forall \ell \in \mathcal{L}_{\text{ani}}$.

Remark 3. This family is more classical than the one constructed in the previous section. All the estimators are defined using the same kernel W° and depend only on a multivariate bandwidth.

2.4 Selection rule

Although the two families differ, the selection procedure is the same in both cases. For the sake of generality, we introduce the following notation: $\mathcal{L}$ is either $\mathcal{L}_{\text{ani}}$ or $\mathcal{L}_{\text{iso}}$ and $\widehat{f}_\ell$ then denotes $\widehat{f}_\ell^{\text{ani}}$ or $\widehat{f}_\ell^{\text{iso}}$. For $\varepsilon \in \{0,1\}^d$, $h \in \mathcal{H}_n$ and $W \in \mathcal{W}^d$ we define:
$$\Delta_{d,\varepsilon} = \prod_{j=1}^d \left(\frac{\varepsilon_j}{2}, \frac{1+\varepsilon_j}{2} \right), \qquad \|W\|_p = \left\| \bigotimes_{i=1}^d W_i \right\|_p$$

$$\widehat{\Lambda}_\varepsilon(W, h, p) = \sqrt{V_h} \left(\int_{\Delta_{d,\varepsilon}} \left(\frac{1}{n} \sum_{j=1}^n \mathcal{K}_{W,h}^2(t, X_j) \right)^{p/2} \mathrm{d}t \right)^{1/p}$$
and
$$\widehat{\Gamma}_\varepsilon(W, h, p) = \begin{cases} 2^{-\frac{d(2-p)}{2p}} \|W\|_2 & \text{if } 1 \leq p \leq 2 \\ C_p^* \left(\widehat{\Lambda}_\varepsilon(W, h, p) + 2\|W\|_p \right) & \text{if } p > 2 \end{cases}$$
where $C_p^* = 15p/\log p$ is the best known constant in the Rosenthal inequality. For any $\ell, \ell' \in \mathcal{L}$ we consider:
$$\widehat{M}_p(\ell) = \frac{1}{\sqrt{nV_{h(\ell)}}} \sum_{\varepsilon \in \{0,1\}^d} \widehat{\Gamma}_\varepsilon(W(\ell), h(\ell), p) \quad \text{and} \quad \widehat{M}_p(\ell, \ell') = \widehat{M}_p(\ell') + \widehat{M}_p(\ell' \wedge \ell)$$

where $\ell \wedge \ell'$ is the vector with coordinates $\ell_i \wedge \ell'_i = \min(\ell_i, \ell'_i)$. Now, for any $\tau > 0$ we define:

$$\widehat{B}_p(\ell) = \max_{\ell' \in \mathcal{L}} \left\{ \|\widehat{f}_{\ell \wedge \ell'} - \widehat{f}_{\ell'}\|_p - (1+\tau)\widehat{M}_p(\ell, \ell') \right\}_+$$

where $x_+ = \max(x, 0)$ denotes the positive part of x.

We then select

$$\widehat{\ell} = \arg\min_{\ell \in \mathcal{L}} \left(\widehat{B}_p(\ell) + (1+\tau)\widehat{M}_p(\ell) \right)$$

which leads to the final plug-in estimator defined by $\widehat{f} = \widehat{f}_{\widehat{\ell}}$. In what follows we denote by $\widehat{f}^{\mathrm{ani}}$ and $\widehat{f}^{\mathrm{iso}}$ the resulting estimators.

Remark 4. This procedure is inspired by the method developped by Goldenshluger and Lepski. Here $\widehat{M}_p(\ell)$ is an empirical version of an upper bound on the standard deviation of $\widehat{f}_\ell$ and $\widehat{B}_p(\ell)$ is linked with the bias term of this estimator, see (14). This implies that $\widehat{f}$ realizes a tradeoff between $\widehat{B}_p(\ell)$ and $(1+\tau)\widehat{M}_p(\ell)$. This can be interpreted as an empirical counterpart of the classical tradeoff between the bias and the standard deviation.

3 Results

Our first result concerns a nonasymptotic oracle-type inequality which proves that the procedure $\widehat{f}^{\mathrm{ani}}$ performs almost as well as the best estimator from the collection $\{\widehat{f}^{\mathrm{ani}}_\ell : \ell \in \mathcal{L}_{\mathrm{ani}}\}$.

Theorem 1. *Assume that $f : \Delta_d \to \mathbb{R}$ is a density function such that $\|f\|_\infty \le F_\infty$. Then there exists a positive constant $\mathfrak{K}_1$ that depends only on F_∞, W°, p, q and τ, such that, for any $n \ge 2$:*

$$R_n^{(p,q)}(\widehat{f}^{\mathrm{ani}}, f) \le \mathfrak{K}_1 \inf_{\ell \in \mathcal{L}_{\mathrm{ani}}} \left\{ \|\mathbf{E}_f^n \widehat{f}^{\mathrm{ani}}_\ell - f\|_p + \max_{\ell' \in \mathcal{L}_{\mathrm{ani}}} \|\mathbf{E}_f^n \widehat{f}^{\mathrm{ani}}_{\ell'} - \mathbf{E}_f^n \widehat{f}^{\mathrm{ani}}_{\ell \wedge \ell'}\|_p + \frac{1}{(nV_{h(\ell)})^{1/2}} \right\}.$$

Our second result is an adaptive minimax bound over a scale of anisotropic Sobolev-Slobodetskii (see Opic and Rákosník, 1991) classes defined below. To our knowledge the minimax risk over such classes has not yet been studied. However they are well-adapted to measure the smoothness of functions defined on a bounded set. Note that the anisotropic Sobolev-Slobodetskii classes contain the classical anisotropic Hölder classes on Δ_d (with the same regularity parameters) but also contain functions with inhomogeneous regularity.

Definition 1. *Set $s = (s_1, \ldots, s_d) \in (0, +\infty)^d$ and $L > 0$. A function $f : \Delta_d \to \mathbb{R}$, belongs to the anisotropic Sobolev-Slobodetskii ball $\mathbb{S}_p(s, L)$ if:*

- *f belongs to $\mathbb{L}_p(\Delta_d)$.*
- *$\|D_i^{\lfloor s_i \rfloor} f\|_p \le L$.*
- *The following property holds:*

$$\sum_{i=1}^d I_i(D_i^{\lfloor s_i \rfloor} f) \le L,$$

where $D_i^k f$ denotes the kth-order partial derivative of f with respect to the variable x_i, $\lfloor s_i \rfloor$ is the largest integer strictly smaller than s_i and

$$I_i(g) = \left(\int_{\Delta_d} \int_{\Delta_1} \frac{\left| g(x) - g(x_1, \dots, x_{i-1}, \xi, x_{i+1}, \dots, x_d) \right|^p}{|x_i - \xi|^{1+p(s_i - \lfloor s_i \rfloor)}} dx d\xi \right)^{1/p}.$$

Theorem 2. *Set $M = (M_1, \dots, M_d) \in \mathbb{N}^d$. Assume that W° is such that W_i° is of order greater than or equal to M_i. For any $s \in \prod_{i=1}^d (0, M_i + 1]$ and $L > 0$, the estimator $\widehat{f}^{\text{ani}}$ satisfies:*

$$\limsup_{n \to +\infty} n^{\frac{\bar{s}}{2\bar{s}+1}} \sup_{f \in \mathbb{S}_{s,p}(L)} R_n^{(p,q)}(\widehat{f}^{\text{ani}}, f) < +\infty,$$

where

$$\frac{1}{\bar{s}} = \sum_{i=1}^d \frac{1}{s_i}.$$

The third result is an oracle-type inequality for the family $\widehat{f}^{\text{iso}}$. This theorem allows us to derive an adaptive result over the isotropic classes $\tilde{\mathbb{S}}_p(s, L)$ (defined below) without any restriction on the parameter $0 < s < +\infty$.

Theorem 3. *Assume that $f : \Delta_d \to \mathbb{R}$ is a density function such that $\|f\|_\infty \le F_\infty$. Then there exists a positive constant $\mathfrak{K}_2$ that depends only on F_∞, p, q and τ, such that, for any $n \ge 2$:*

$$R_n^{(p,q)}(\widehat{f}^{\text{iso}}, f) \le \mathfrak{K}_2 \inf_{\ell \in \mathcal{L}_{\text{iso}}} \left\{ \max_{\ell' \ge \ell} \| \mathbf{E}_f^n \widehat{f}_{\ell'}^{\text{iso}} - f \|_p + \frac{\|W(\ell)\|_{p \vee 2}}{(nV_{h(\ell)})^{1/2}} \right\}.$$

Definition 2. *Set $s > 0$ and $L > 0$. A function $f : \Delta_d \to \mathbb{R}$, belongs to $\tilde{\mathbb{S}}_{s,p}(L)$ if the following properties hold:*

- *for any $\alpha \in \mathbb{N}^d$, such that $|\alpha| \le \lfloor s \rfloor = \max\{i \in \mathbb{N} : i < s\}$, the mixed partial derivatives*

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}$$

  exist and belong to $\mathbb{L}_p(\Delta_d)$.

- *the Gagliardo semi-norm $|f|_{s,p}$ is bounded by L where*

$$|f|_{s,p} = \left(\sum_{|\alpha| = \lfloor s \rfloor} \int_{\Delta_d^2} \frac{|D^\alpha f(y) - D^\alpha f(x)|^p}{|y - x|^{d + p(s - \lfloor s \rfloor)}} \, \mathrm{d}x \, \mathrm{d}y \right)^{1/p}.$$

Theorem 4. *For any $s > 0$ and $L > 0$, the estimator $\widehat{f}^{\text{iso}}$ satisfies:*

$$\limsup_{n \to +\infty} n^{\frac{s}{2s+d}} \sup_{f \in \tilde{\mathbb{S}}_{s,p}(L)} R_n^{(p,q)}(\widehat{f}^{\text{iso}}, f) < +\infty,$$

Remark 5. In Theorems 1 and 3, the right hand sides of the equations can be easily interpreted. In both situations, the term $(nV_{h(\ell)})^{-1/2}$ is of the order of the standard deviation of $\widehat{f}_\ell$. Moreover the terms $\max_{\ell' \in \mathcal{L}_{\text{ani}}} \| \mathbf{E}_f^n \widehat{f}_{\ell'}^{\text{ani}} - \mathbf{E}_f^n \widehat{f}_{\ell \wedge \ell'}^{\text{ani}} \|_p$ and $\max_{\ell' \ge \ell} \| \mathbf{E}_f^n \widehat{f}_{\ell'}^{\text{iso}} - f \|_p$ are linked with the bias of this estimator. More precisely, Proposition 1 and Proposition 2 ensure that these terms have the same behaviour as the bias term $\| \mathbf{E}_f^n \widehat{f}_\ell - f \|_p$ as soon as f belongs to Sobolev-Slobodetskii classes. Theorems 2 and 4 provide new adaptive results since the proposed procedures achieve the minimax rate of convergence over the Sobolev-Slobodetskii classes and classical Hölder classes (lower bounds can be easily obtained using classical techniques introduced by Tsybakov, 2009).

4 Proofs

The proofs of Theorems 1–4 are based on propositions and lemmas which are given below. Before stating these results, we introduce some notation that are used throughout the rest of the paper. For $W = (W_1, \dots, W_d) \in \mathcal{W}^d$, $h \in \mathcal{H}_n$ and $\varepsilon \in \{0,1\}^d$, we define the quantity:

$$\Gamma_\varepsilon(W,h,p) = \begin{cases} 2^{-\frac{d(2-p)}{2p}} \|W\|_2 & \text{if } 1 \le p \le 2 \\ C_p^* \left(\Lambda_\varepsilon(W,h,p) + 2\|W\|_p\right) & \text{if } p > 2 \end{cases}$$

where

$$\Lambda_\varepsilon(W,h,p) = \sqrt{V_h} \left(\int_{\Delta_{d,\varepsilon}} \left(\int_{\Delta_d} \mathcal{K}^2_{W,h}(t,x) f(x)\,\mathrm{d}x \right)^{p/2} \mathrm{d}t \right)^{1/p}.$$

For $g : \Delta_d \to \mathbb{R}$ and $r \ge 1$ we denote

$$\|g\|_{r,\varepsilon} = \left(\int_{\Delta_{d,\varepsilon}} |g(x)|^r \,\mathrm{d}x \right)^{1/r}.$$

The process $\xi_{W,h}$ is defined by

$$\xi_{W,h}(t) = \left(\frac{V_h}{n} \right)^{1/2} \sum_{i=1}^n \left(\mathcal{K}_{W,h}(t, X_i) - \mathbf{E}_f^n \mathcal{K}_{W,h}(t, X_i) \right), \qquad t \in \Delta_d.$$

Finally, for $\ell \in \mathcal{L}$ we define (using the generic notation for the istropic and the anisotropic cases):

$$W^*(\ell) = \left(\frac{(W_1(\ell))^2}{\|W_1(\ell)\|_2^2}, \dots, \frac{(W_d(\ell))^2}{\|W_d(\ell)\|_2^2} \right).$$

Proposition 1 (Anisotropic case)**.** *Set $M = (M_1, \dots, M_d) \in \mathbb{N}^d$. Assume that W° is such that W_i° is of order greater than or equal to M_i. Set $s = (s_1, \dots, s_d) \in \prod_{i=1}^d (0, M_i]$ and $L > 0$. Then, for any $f \in \mathbb{S}_{s,p}(L)$:*

$$\|\mathbf{E}_f^n \widehat{f}_\ell^{\text{ani}} - f\|_p \le 2^{d/p} d \left(\prod_{i=1}^d (M_i + 1) \right) L \sum_{i=1}^d \left(h_i(\ell) \right)^{s_i}. \tag{3}$$

$$\max_{k \in \mathcal{L}_{\text{ani}}} \|\mathbf{E}_f^n \widehat{f}_k^{\text{ani}} - \mathbf{E}_f^n \widehat{f}_{\ell \wedge k}^{\text{ani}}\|_p \le 2^{1+d/p} d \left(\prod_{i=1}^d (M_i + 1) \right) L \sum_{i=1}^d \left(h_i(\ell) \right)^{s_i}. \tag{4}$$

Proposition 2 (Isotropic case)**.** *Set $s > 0$ and $L > 0$. Then for any $\ell \in \mathcal{L}_{\text{iso}}$ we have:*

$$\sup_{f \in \tilde{\mathbb{S}}_{s,p}(L)} \max_{\ell' \ge \ell} \|\mathbf{E}_f^n \widehat{f}_\ell^{\text{iso}} - f\|_p \le \mathfrak{K}_3 \left(\|W(\ell)\|_\infty L (h_1(\ell))^s + \sqrt{\frac{h^*}{n}} \right),$$

where the positive constant $\mathfrak{K}_3$ depends only on d, p, s and L.

Proposition 3. *Set $p, q \ge 1$. Assume that f is such that $\|f\|_\infty \le F_\infty$.*

- *Let $\ell \in \mathcal{L}_{\rm iso}$. There exists a positive constant $\mathfrak{K}_4$ that depends only on p, q, τ and F_∞ such that*

$$\mathbf{E}_f^n \left\{ \|\widehat{f}_\ell^{\rm iso} - \mathbf{E}_f^n \widehat{f}_\ell^{\rm iso}\|_p - (1+\tau)\widehat{M}_p(\ell) \right\}_+^q \leq \mathfrak{K}_4 n^{-q}.$$

- *Let $\ell \in \mathcal{L}_{\rm ani}$. There exists a positive constant $\mathfrak{K}_5$ that depends only on p, q, τ, W° and F_∞ such that*

$$\mathbf{E}_f^n \left\{ \|\widehat{f}_\ell^{\rm ani} - \mathbf{E}_f^n \widehat{f}_\ell^{\rm ani}\|_p - (1+\tau)\widehat{M}_p(\ell) \right\}_+^q \leq \mathfrak{K}_5 n^{-q}.$$

Lemma 2. *Assume that f satisfies $\|f\|_\infty \leq \mathrm{F}_\infty$. For any $W \in \mathcal{W}^d$, $r \geq 1$ and $h \in \mathcal{H}_n$, we have:*

$$\mathbf{E}_f^n \|\xi_{W,h}\|_{r,\varepsilon} \leq \Gamma_\varepsilon(W,h,r) \leq C_0 \|W\|_{2\vee r},$$

where C_0 is an absolute constant that depends only on r and F_∞.

Lemma 3. *Assume that f satisfies $\|f\|_\infty \leq \mathrm{F}_\infty$. We have, for any $\ell \in \mathcal{L}_{\rm iso}$, $h \in \mathcal{H}_n$, $W \in \{W(\ell), W^*(\ell)\}$ and $r \geq 1$,*

$$\mathbf{P}(\|\xi_{W,h}\|_{r,\varepsilon} - \mathbf{E}_f^n \|\xi_{W,h}\|_{r,\varepsilon} \geq \frac{\tau}{2}\Gamma_\varepsilon(W,h,r) + x) \leq \exp\left(-\frac{x^2}{C_2(1+x)}\right) \exp\left(-C_1 \alpha_n(r)\right) \quad (5)$$

where C_1 and C_2 are absolute constants that depend only on r, τ and F_∞ and

$$\alpha_n(r) = \begin{cases} (h^*)^{-d\left(\frac{2}{r}-1\right)} & \text{if } 1 \leq r < 2 \\ (h^*)^{-\frac{d}{r}} & \text{if } r \geq 2. \end{cases}$$

If $\ell \in \mathcal{L}_{\rm ani}$, (5) holds also with C_2 that depends on r, W° and F_∞.

Lemma 4. *Assume that f satisfies $\|f\|_\infty \leq \mathrm{F}_\infty$. We have, for any $\ell \in \mathcal{L}_{\rm iso}$, $h \in \mathcal{H}_n$, $\delta > 0$ and $r \geq 1$,*

$$\mathbf{P}\left(\sum_{\varepsilon \in \{0,1\}^d} \|\xi_{W^*(\ell),h}\|_{r,\varepsilon}^{1/2} \geq \delta 2^d \left(nV_h\right)^{1/4} \right) \leq C_3 \exp\left(-C_1 \alpha_n(r)\right), \quad (6)$$

where C_3 is an absolute constant that depends only on r, δ and F_∞.

If $\ell \in \mathcal{L}_{\rm ani}$, (6) holds also with C_3 that depends on r, δ, W° and F_∞.

We finally state the following lemma that allows us to bound the bias terms which appear in the oracle inequality.

Lemma 5. *Let $h = (h_1, \dots, h_d)$ and $\eta = (\eta_1, \dots, \eta_d)$ be two bandwidths in $\mathcal{H}_n$ such that $\eta_i \in \{0, h_i\}$. Set $W = (w_{M_1}, \dots, w_{M_d}) \in \mathcal{W}^d$ and define:*

$$S_{W,h,\eta}^*(f) = \left(\int_{\Delta_{d,\mathbf{0}}} \left| \int_{\Delta_d} \left(\prod_{j=1}^d w_{M_j}(u_j) \right) \left[f(t + h\cdot u) - f(t + \eta \cdot u) \right] \mathrm{d}u \right|^p \mathrm{d}t \right)^{1/p}$$

where $h \cdot u$ denotes the coordinate-wise product of the vectors h and u. Assume that f belongs to $\mathbb{S}_{s,p}(L)$ and that, for any $i = 1, \dots, d$, the kernel W_i is of order greater than or equal to $\lfloor s_i \rfloor$. Then we have:

$$S_{W,h,\eta}^*(f) \leq d \left(\prod_{i=1}^d (M_i + 1) \right) L \sum_{i \in I} h_i^{s_i}$$

where $I = \{i = 1, \dots, d : \eta_i = 0\}$.

4.1 Proof of Proposition 1

We first prove (3). For the sake of brevity, define $W = W^\circ$ and $h = h(\ell)$.

$$\begin{aligned}\|\mathbf{E}_f^n \widehat{f}^{\mathrm{ani}} - f\|_p^p &= \sum_{\varepsilon\in\{0,1\}^d} \int_{\Delta_{d,\varepsilon}} |\mathbf{E}_f^n \tilde{f}_{W,h}(t) - f(t)|^p \,\mathrm{d}t \\ &= \sum_{\varepsilon\in\{0,1\}^d} \int_{\Delta_{d,\varepsilon}} \left|\int_{\Delta_d} \mathcal{K}_{W,h}(t,x) f(x)\,\mathrm{d}x - f(t)\right|^p \mathrm{d}t \\ &= \sum_{\varepsilon\in\{0,1\}^d} \left(S_{W,h}(f^{(\varepsilon)})\right)^p \end{aligned} \tag{7}$$

where

$$S_{W,h}(f) = \left(\int_{\Delta_{d,\mathbf{0}}} \left|\int_{\Delta_d} \mathcal{K}_{W,h}(t,x) f(x)\,\mathrm{d}x - f(t)\right|^p \mathrm{d}t\right)^{1/p} \tag{8}$$

and

$$f^{(\varepsilon)}(t) = f(\ldots, t_i(1-\varepsilon_i) + (1-t_i)\varepsilon_i, \ldots).$$

Remark that (7) comes from the symmetry properties of the kernel $\mathcal{K}_{W,h}$. Since $f \in \mathbb{S}_{s,p}(L) \iff f^{(\varepsilon)} \in \mathbb{S}_{s,p}(L)$ we obtain

$$\sup_{f\in\mathbb{S}_{s,p}(L)} \|\mathbf{E}_f^n \tilde{f}_{W,h} - f\|_p \le 2^{d/p} \sup_{f\in\mathbb{S}_{s,p}(L)} S_{W,h}(f).$$

Using the notation introduced in Lemma 5 we have:

$$\sup_{f\in\mathbb{S}_{s,p}(L)} \|\mathbf{E}_f^n \tilde{f}_{W,h} - f\|_p \le 2^{d/p} \sup_{f\in\mathbb{S}_{s,p}(L)} S^*_{W,h,0}(f).$$

Equation (3) follows.

Now, let us prove (4). Set $h = h(k)$ and $h' = h(k\wedge\ell) = h(k) \vee h(\ell)$. Similarly to (7) we have:

$$\begin{aligned}\|\mathbf{E}_f^n \tilde{f}_{W,h} - \mathbf{E}_f^n \tilde{f}_{W,h'}\|_p^p &\le 2^d \int_{\Delta_{d,\mathbf{0}}} \left|\int_{\Delta_d} \mathcal{K}_{W,h}(t,x) f(x)\,\mathrm{d}x - \int_{\Delta_d} \mathcal{K}_{W,h'}(t,x) f(x)\,\mathrm{d}x\right|^p \mathrm{d}t \\ &\le 2^d \int_{\Delta_{d,\mathbf{0}}} \left|\int_{\Delta_d} \left(\prod_{i=1}^d W_i(u_i)\right) [f(t + h\cdot u) - f(t + h'\cdot u)]\,\mathrm{d}u\right|^p \mathrm{d}t.\end{aligned}$$

Let $\eta = (\eta_1, \ldots, \eta_d)$ be a bandwidth defined by

$$\eta_i = \begin{cases} 0 & \text{if } h_i < h_i' \\ h_i & \text{if } h_i = h_i'. \end{cases}$$

We have:

$$\|\mathbf{E}_f^n \tilde{f}_{W,h} - \mathbf{E}_f^n \tilde{f}_{W,h'}\|_p^p \le 2^{d+p} \max_{H\in\{h,h'\}} \int_{\Delta_{d,\mathbf{0}}} \left|\int_{\Delta_d} \left(\prod_{i=1}^d W_i(u_i)\right) [f(t + H\cdot u) - f(t + \eta\cdot u)]\,\mathrm{d}u\right|^p \mathrm{d}t.$$

Using Lemma 5, we obtain:

$$\|\mathbf{E}_f^n \tilde{f}_{W,h} - \mathbf{E}_f^n \tilde{f}_{W,h'}\|_p \le 2^{1+d/p} d \left(\prod_{i=1}^d (M_i + 1)\right) L \max_{H\in\{h,h'\}} \sum_{i\in I} H_i^{s_i}$$

where $I = \{i : \eta_i = 0\}$. Since $H_i \le h_i(\ell)$ for any $i \in I$, this allows us to conclude.

4.2 Proof of Proposition 2

In the same way that the proof of Proposition 1, we obtain:

$$\sup_{f \in \tilde{S}_{s,p}(L)} \|\mathbf{E}_f^n \widehat{f}_\ell^{\text{iso}} - f\|_p \le 2^{d/p} \sup_{f \in \tilde{S}_{s,p}(L)} S_{W(\ell),h(\ell)}(f),$$

where $S_{W(\ell),h(\ell)}(f)$ is defined by (8). We introduce the following notation:

$$k = k(\ell, s) = \begin{cases} \lfloor s \rfloor & \text{if } m(\ell) \ge \lfloor s \rfloor \\ m(\ell) & \text{otherwise} \end{cases}$$

and

$$\varsigma = \varsigma(\ell, s) = \begin{cases} s & \text{if } m(\ell) \ge \lfloor s \rfloor \\ m(\ell) + 1 & \text{otherwise} \end{cases}$$

Remark that, using this notation the kernel $w_{m(\ell)}$ is of order greater than or equal to k and $\varsigma \le s$. Moreover, using classical embedding theorems (see Nezza et al. (2012)), there exists a positive constant $\tilde{L}$ that depends only on L, s and p, such that for $\varsigma \in \{2, \dots, \lfloor s \rfloor\}$, we have $\tilde{\mathbb{S}}_{s,p}(L) \subset \tilde{\mathbb{S}}_{\varsigma,p}((\tilde{L}))$. For $\varsigma = s$ we also denote $\tilde{L} = L$.

Now, denoting $h = h(\ell)$ and using a Taylor expansion of f, we obtain:

$$S_{m,h}(f) \le (k \vee 1) \|W(\ell)\|_\infty \left(\sum_{|\alpha| = k} I_\alpha \right)^{1/p}$$

where

$$\begin{aligned} I_\alpha &= h^{pk} \int_{\Delta_{d,\mathbf{0}}} \int_{\Delta_d} \int_0^1 |(D^\alpha f(t + \tau h u) - D^\alpha f(t))|^p \, \mathrm{d}\tau \, \mathrm{d}u \, \mathrm{d}t \\ &\le h^{pk} \int_{\Delta_{d,\mathbf{0}}} \int_{\Delta_d} \int_0^1 \|hu\|_2^{d+p(\varsigma-k)} \frac{|D^\alpha f(t + \tau h u) - D^\alpha f(t)|^p}{\|\tau h u\|_2^{d+p(\varsigma-k)}} \, \mathrm{d}\tau \, \mathrm{d}u \, \mathrm{d}t \\ &\le d^{(d+p)/2} h^{p\varsigma} \int_0^1 \int_{\Delta_{d,\mathbf{0}}} \int_{\Delta_d} \frac{|D^\alpha f(x) - D^\alpha f(t)|^p}{\|x - t\|_2^{d+p(\varsigma-k)}} \, \mathrm{d}x \, \mathrm{d}t \, \mathrm{d}\tau \\ &\le d^{(d+p)/2} \tilde{L}^p h^{p\varsigma}. \end{aligned}$$

We thus obtain

$$\|\mathbf{E}_f^n \widehat{f}_\ell - f\|_p \le \left[C(d,p,s) \|W(\ell)\|_\infty \right] \tilde{L} h^\varsigma \tag{9}$$

where

$$C(d,p,s) = (2^d d^{\frac{d+p}{2}})^{1/p} (\lfloor s \rfloor \vee 1).$$

If $m(\ell) \ge \lfloor s \rfloor$, since $\tilde{L} = L$ and $\varsigma = s$, we deduce from (9) that:

$$\|\mathbf{E}_f^n \widehat{f}_\ell - f\|_p \le \left[C(d,p,s) \|W(\ell)\|_\infty \right] L (h(\ell))^s. \tag{10}$$

Assume now that $m(\ell) < \lfloor s \rfloor$. Then $\varsigma = m(\ell) + 1$ and (9) writes

$$\|\mathbf{E}_f^n \widehat{f}_\ell^{\text{iso}} - f\|_p \le \left[C(d,p,s) \|W(\ell)\|_\infty \right] \tilde{L} (h(\ell))^{m(\ell)+1}.$$

Remark that

$$\begin{aligned}(h(\ell))^{m(\ell)+1} &= \exp(-\ell(m(\ell)+1))\\ &\leq \exp\left(-\ell\left(\frac{\log n}{2\ell}+\frac{1}{2}\right)\right)\\ &\leq \sqrt{\frac{h^*}{n}}.\end{aligned}$$

Thus, using Lemma 1, for $m(\ell) < \lfloor s \rfloor$ we obtain:

$$\|\mathbf{E}_f^n \widehat{f}_\ell^{\text{iso}} - f\|_p \leq \left[C(d,p,s)(\lfloor s \rfloor + 1)^{3d/2}\right] \tilde{L}\sqrt{\frac{h^*}{n}}. \tag{11}$$

Combining (10) and (11) we obtain the proposition.

4.3 Proof of Proposition 3

In the following, $\mathcal{L}$ is either $\mathcal{L}_{\text{ani}}$ or $\mathcal{L}_{\text{iso}}$ and $\widehat{f}_\ell$ then denotes $\widehat{f}_\ell^{\text{ani}}$ or $\widehat{f}_\ell^{\text{iso}}$. Let $\ell \in \mathcal{L}$. We define

$$M_p(\ell) = \frac{1}{\sqrt{nV_{h(\ell)}}} \sum_{\varepsilon \in \{0,1\}^d} \Gamma_\varepsilon(W(\ell), h(\ell), p).$$

First, assume that $1 \leq p \leq 2$. In this case $M_p(\ell) = \widehat{M}_p(\ell)$,which implies that

$$\mathbf{E}_f^n \left\{\|\widehat{f}_\ell - \mathbf{E}_f^n \widehat{f}_\ell\|_p - (1+\tau)\widehat{M}_p(\ell)\right\}_+^q \leq A_{p,q}(\ell)$$

where

$$A_{p,q}(\ell) = \mathbf{E}_f^n \left\{\|\widehat{f}_\ell - \mathbf{E}_f^n \widehat{f}_\ell\|_p - (1+\tau/2)M_p(\ell)\right\}_+^q.$$

Next, assume that $p > 2$. Denoting $\delta = \frac{\tau}{2(1+\tau)}$, we have

$$\begin{aligned}\left\{\|\widehat{f}_\ell - \mathbf{E}_f^n \widehat{f}_\ell\|_p - (1+\tau)\widehat{M}_p(\ell)\right\}_+ &\leq \left\{\|\widehat{f}_\ell - \mathbf{E}_f^n \widehat{f}_\ell\|_p - (1+\tau/2)M_p(\ell)\right\}_+\\ + (1+\tau/2)M_p(\ell)\mathbf{I}_{\bar{\mathcal{D}}_{\delta,\ell}} &\quad + \left\{\|\widehat{f}_\ell - \mathbf{E}_f^n \widehat{f}_\ell\|_p - (1+\tau)\widehat{M}_p(\ell)\right\}_+ \mathbf{I}_{\mathcal{D}_{\delta,\ell}},\end{aligned}$$

where

$$\mathcal{D}_{\delta,\ell} = \left\{\sum_{\varepsilon \in \{0,1\}^d} \|\xi_{W^*(\ell)),h(\ell)}\|_{p/2,\varepsilon}^{1/2} \leq \delta 2^d \left(nV_{h(\ell)}\right)^{1/4}\right\}. \tag{12}$$

This implies:

$$\begin{aligned}&\mathbf{E}_f^n \left\{\|\widehat{f}_\ell - \mathbf{E}_f^n \widehat{f}_\ell\|_p - (1+\tau)\widehat{M}_p(\ell)\right\}_+^q\\ &\leq 3^{q-1}\left(A_{p,q}(\ell) + B_{p,q}(\ell) + C_{p,q}(\ell)\right)\end{aligned}$$

where

$$B_{p,q}(\ell) = (1+\tau/2)^q (M_p(\ell))^q \mathbf{P}_f^n \left(\bar{\mathcal{D}}_{\delta,\ell}\right)$$

$$C_{p,q}(\ell) = \mathbf{E}_f^n \left(\left\{ \|\widehat{f}_\ell - \mathbf{E}_f^n \widehat{f}_\ell\|_p - (1+\tau)\widehat{M}_p(\ell) \right\}_+^q \mathbf{I}_{\mathcal{D}_{\delta,\ell}} \right).$$

Using Lemma 4 with $r = \frac{p}{2}$, we immediately obtain that

$$B_{p,q}(\ell) = \mathcal{O}(n^{-q}).$$

It remains to upper bound $A_{p,q}(\ell)$ for $p, q \geq 1$ and $C_{p,q}(\ell)$ for $q \geq 1$ and $p > 2$.

Control of $A_{p,q}(\ell)$. Remark that

$$\begin{aligned} A_{p,q}(\ell) &\leq \mathbf{E}_f^n \left\{ \sum_{\varepsilon \in \{0,1\}^d} \|\widehat{f}_\ell - \mathbf{E}_f^n \widehat{f}_\ell\|_{p,\varepsilon} - \frac{(1+\tau/2)\Gamma_\varepsilon(W(\ell), h(\ell), p)}{\sqrt{nV_{h(\ell)}}} \right\}_+^q \\ &\leq 2^{d(q-1)} \sum_{\varepsilon \in \{0,1\}^d} \mathcal{I}_{q,\varepsilon}, \end{aligned}$$

where

$$\mathcal{I}_{q,\varepsilon} = \mathbf{E}_f^n \left\{ \|\widehat{f}_\ell - \mathbf{E}_f^n \widehat{f}_\ell\|_{p,\varepsilon} - \frac{(1+\tau/2)\Gamma_\varepsilon(W(\ell), h(\ell), p)}{\sqrt{nV_{h(\ell)}}} \right\}_+^q$$

Thus, using Lemma 2 and Lemma 3 with $r = p$ we can write:

$$\begin{aligned} \left(nV_{h(\ell)}\right)^{q/2} \mathcal{I}_{q,\varepsilon} &= \mathbf{E}_f^n \left\{ \|\xi_{W(\ell),h(\ell)}\|_{p,\varepsilon} - (1+\tau/2)\Gamma_\varepsilon(W(\ell), h(\ell), p) \right\}_+^q \\ &\leq q \int_0^{+\infty} y^{q-1} \mathbf{P}_f^n \left(\|\xi_{W(\ell),h(\ell)}\|_{p,\varepsilon} - (1+\tau/2)\Gamma_\varepsilon(W(\ell), h(\ell), p) > y \right) \mathrm{d}y \\ &\leq q \int_0^{+\infty} y^{q-1} \mathbf{P}_f^n \left(\|\xi_{W(\ell),h(\ell)}\|_{p,\varepsilon} - \mathbf{E}_f^n \|\xi_{W(\ell),h(\ell)}\|_{p,\varepsilon} > \frac{\tau}{2}\Gamma_\varepsilon(W(\ell), h(\ell), p) + y \right) \mathrm{d}y \\ &\leq q \exp(-C_1 \alpha_n(p)) \int_0^{+\infty} y^{q-1} \exp\left(-\frac{y^2}{C_2(1+y)} \right) \mathrm{d}y \\ &\leq C \exp(-C_1 \alpha_n(p)) \end{aligned}$$

where C depends only on C_2 and q.

This implies that

$$A_{p,q}(\ell) = \mathcal{O}(n^{-q}).$$

Control of $C_{p,q}(\ell)$. Recall that $p \geq 2$. Let us remark that

$$\left| \widehat{M}_p(\ell) - M_p(\ell) \right| = \left| \sum_{\varepsilon \in \{0,1\}^d} \frac{C_p^* \|W(\ell)\|_2}{(nV_{h(\ell)})^{1/2}} Z_\varepsilon(\ell, p) \right|$$

where

$$Z_\varepsilon(\ell, p) = \left(\int_{\Delta_{d,\varepsilon}} \left(\mathbf{E}_f^n \mathcal{K}_{W^*(\ell),h(l)}(t, X_1) \right)^{p/2} dt \right)^{1/p} - \left(\int_{\Delta_{d,\varepsilon}} \left(\frac{1}{n} \sum_{i=1}^n \mathcal{K}_{W^*(\ell),h(l)}(t, X_i) \right)^{p/2} dt \right)^{1/p}.$$

We have

$$|Z_\varepsilon(\ell,p)| = \left| \sqrt{\left\| \mathbf{E}_f^n \mathcal{K}_{W^*(\ell),h(l)}(\cdot,X_1) \right\|_{p/2,\varepsilon}} - \sqrt{\left\| \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{W^*(\ell),h(l)}(\cdot,X_i) \right\|_{p/2,\varepsilon}} \right|$$
$$\leq (nV_{h(\ell)})^{-1/4} \|\xi_{W^*(\ell),h(\ell)}(\cdot)\|_{p/2,\varepsilon}^{1/2}.$$

This implies that

$$\left|\widehat{M}_p(\ell) - M_p(\ell)\right| \leq \frac{C_p^* \|W(\ell)\|_2}{(nV_{h(\ell)})^{3/4}} \sum_{\varepsilon\in\{0,1\}^d} \|\xi_{W^*(\ell),h(\ell)}(\cdot)\|_{p/2,\varepsilon}^{1/2}. \tag{13}$$

Thus, under $\mathcal{D}_{\delta,\ell}$ defined in (12) we have

$$\begin{aligned}\left|\widehat{M}_p(\ell) - M_p(\ell)\right| &\leq \frac{2^d C_p^* \|W(\ell)\|_2}{(nV_{h(\ell)})^{3/4}} \delta (nV_{h(\ell)})^{1/4} \\ &\leq \delta M_p(\ell),\end{aligned}$$

$$\widehat{M}_p(\ell) \geq (1-\delta)\, M_p(\ell),$$

and, since $(1-\delta)(1+\tau) = 1+\tau/2$:

$$(1+\tau)\widehat{M}_p(\ell) \geq (1+\tau/2) M_p(\ell).$$

This implies that

$$C_{p,q}(\ell) \leq A_{p,q}(\ell) = \mathcal{O}(n^{-q}).$$

4.4 Proof of Theorem 1

First, we introduce the following notation: for any $\ell,\ell' \in \mathcal{L}$, we denote $\ell \preceq \ell'$ if, for any $i = 1,\dots,d$, we have $\ell_i \leq \ell'_i$. Let $\ell \in \mathcal{L}_{\text{iso}}$ be an arbitrary multiindex. To simplify the notation, we use $\widehat{f}_\ell = \widehat{f}_\ell^{\text{ani}}$ and $\widehat{f} = \widehat{f}^{\text{ani}}$.

Using the definitions of $\widehat{B}_p(\ell)$ and $\widehat{M}_p(\ell)$ we easily obtain:

$$\begin{aligned}\|f - \widehat{f}\|_p &\leq \|f - \widehat{f}_\ell\|_p + \|\widehat{f}_{\widehat{\ell}\wedge\ell} - \widehat{f}_\ell\|_p + \|\widehat{f}_{\widehat{\ell}\wedge\ell} - \widehat{f}_{\widehat{\ell}}\|_p \\ &\leq \|f - \widehat{f}_\ell\|_p + \widehat{B}_p(\widehat{\ell}) + (1+\tau)\widehat{M}_p(\widehat{\ell},\ell) + \widehat{B}_p(\ell) + (1+\tau)\widehat{M}_p(\ell,\widehat{\ell}).\end{aligned}$$

Using the definition of $\widehat{\ell}$, we deduce:

$$\begin{aligned}\|f - \widehat{f}\|_p &\leq \|f - \widehat{f}_\ell\|_p + 2\left(\widehat{B}_p(\ell) + (1+\tau)\widehat{M}_p(\ell)\right) + 2(1+\tau)\widehat{M}_p(\ell\wedge\widehat{\ell}) \\ &\leq \|f - \widehat{f}_\ell\|_p + 2\widehat{B}_p(\ell) + 4(1+\tau)\max_{\ell'\preceq\ell} M_p(\ell') + 4(1+\tau)\max_{\ell'\preceq\ell}\left(\widehat{M}_p(\ell') - M_p(\ell')\right).\end{aligned}$$

This implies that:

$$\begin{aligned}R_n^{(p,q)}(\widehat{f},f) \leq R_n^{(p,q)}(\widehat{f}_\ell,f) &+ 2\left(\mathbf{E}_f^n \widehat{B}_p^q(\ell)\right)^{1/q} \\ &+ 4(1+\tau)\left(\mathbf{E}_f^n \max_{\ell'\preceq\ell} |\widehat{M}_p(\ell') - M_p(\ell')|^q\right)^{1/q} \\ &+ 4(1+\tau)\max_{\ell'\preceq\ell} M_p(\ell').\end{aligned}$$

It remains to bound each term of the right hand side of this inequality.

1. Remark that, using the triangular inequality, we have:

$$\widehat{B}_p(\ell) \leq 2 \max_{\ell' \in \mathcal{L}} \left\{ \|\widehat{f}_{\ell'} - \mathbf{E}_f^n \widehat{f}_{\ell'}\|_p - (1+\tau)\widehat{M}_p(\ell') \right\}_+ + \max_{\ell' \in \mathcal{L}} \|\mathbf{E}_f^n \widehat{f}_{\ell'} - \mathbf{E}_f^n \widehat{f}_{\ell \wedge \ell'}\|_p.$$

This readily implies

$$\begin{aligned}\left(\mathbf{E}_f^n \widehat{B}_p^q(\ell)\right)^{1/q} &\leq 2 \sum_{\ell' \in \mathcal{L}} \left(\mathbf{E}_f^n \left\{ \|\widehat{f}_{\ell'} - \mathbf{E}_f^n \widehat{f}_{\ell'}\|_p - (1+\tau)\widehat{M}_p(\ell') \right\}_+^q \right)^{1/q} \\ &\quad + \max_{\ell' \in \mathcal{L}} \|\mathbf{E}_f^n \widehat{f}_{\ell'} - \mathbf{E}_f^n \widehat{f}_{\ell \wedge \ell'}\|_p \\ &\leq \mathfrak{K}_2^{1/q} (\#\mathcal{L}) n^{-1} + \max_{\ell' \in \mathcal{L}} \|\mathbf{E}_f^n \widehat{f}_{\ell'} - \mathbf{E}_f^n \widehat{f}_{\ell \wedge \ell'}\|_p, \end{aligned} \tag{14}$$

where the last inequality follows immediately from Proposition 3.

2. For $p \leq 2$, we have $\widehat{M}_p(\ell) - M_p(\ell) = 0$.

Let $p > 2$. Here and in the following paragraph, C stands for a constant that depends on p, q, τ, F_∞ and W° and that can change of values from line to line. Using (13), we obtain that

$$|\widehat{M}_p(\ell') - M_p(\ell')| \leq \frac{C_p^* \|W^\circ\|_2}{(nV_{h(\ell)})^{1/2}(nV_{h(\ell')})^{1/4}} \sum_{\varepsilon \in \{0,1\}^d} \|\xi_{W^*(\ell'),h(\ell')}(\cdot)\|_{p/2,\varepsilon}^{1/2}.$$

Now considering the events $\mathcal{D}_{\delta,\ell}$ defined by (12) and since $\#\mathcal{L}$ is bounded by $(\log n)^d$, we have

$$\left(\mathbf{E}_f^n \max_{\ell' \preceq \ell} |\widehat{M}_p(\ell') - M_p(\ell')|^q \right)^{1/q} \leq C \frac{\|W^\circ\|_2}{(nV_{h(\ell)})^{1/2}}.$$

3. By using Lemma 2 we obtain

$$\begin{aligned} M_p(\ell) =& \frac{1}{\sqrt{nV_{h(\ell)}}} \sum_{\varepsilon \in \{0,1\}^d} \Gamma_\varepsilon(W^\circ, h(\ell), p) \\ \leq& \frac{C\ \|W^\circ\|_{p \vee 2}}{\sqrt{nV_{h(\ell)}}}. \end{aligned} \tag{15}$$

This implies that for $\ell' \preceq \ell$

$$4 \max_{\ell' \preceq \ell} M_p(\ell') \leq \frac{C}{\sqrt{nV_{h(\ell)}}}.$$

4.5 Proof of Theorem 2

Set $s \in \prod_{i=1}^d (0, M_i + 1]$. Define $\ell^*(s) = (\ell_1^*(s), \ldots, \ell_d^*(s))$ by:

$$\ell_i^*(s) = \left\lceil \frac{\bar{s}}{s_i(2\bar{s}+1)} \log n \right\rceil, \qquad i = 1 \ldots, d$$

where $\lceil x \rceil$ denotes the least integer greater than or equal to x. Note that $h_i(\ell^*)$ is such that

$$\frac{h_i^*(s)}{e} \leq h_i(\ell^*) \leq h_i^*(s) \tag{16}$$

where

$$h_i^*(s) = n^{-\frac{\bar{s}}{s_i(2\bar{s}+1)}}.$$

This implies that there exists $n_0 = n_0(s, p) \in \mathbb{N}$ such that for any $n \geq n_0$ we have $\ell^* \in \mathcal{L}_{\text{ani}}$. Combining (16) with Proposition 1 and Theorem 1, the result follows.

4.6 Proof of Theorem 3

Let $\ell \in \mathcal{L}_{\text{iso}}$ Note that in the isotropic case, we have a simple inequality

$$\|f - \widehat{f}^{\text{iso}}\|_p \leq \|f - \widehat{f}_\ell^{\text{iso}}\|_p + 4\left(\widehat{B}_p(\ell) + (1+\tau)\widehat{M}_p(\ell)\right),$$

and then

$$\begin{aligned} R_n^{(p,q)}(\widehat{f}^{\text{iso}}, f) \leq R_n^{(p,q)}(\widehat{f}_\ell^{\text{iso}}, f) &+ 4\left(\mathbf{E}_f^n \widehat{B}_p^q(\ell)\right)^{1/q} \\ &+ 4(1+\tau)\left(\mathbf{E}_f^n |\widehat{M}_p(\ell) - M_p(\ell)|^q\right)^{1/q} \\ &+ 4(1+\tau)M_p(\ell). \end{aligned}$$

Following the same arguments of the proof of Theorem 1 (see the second paragraph), we have

$$\left(\mathbf{E}|\widehat{M}_p(\ell) - M_p(\ell)|^q\right)^{1/q} \leq C\frac{\|W(\ell)\|_2}{(nV_{h(\ell)})^{1/2}}. \tag{17}$$

Applying (14), (15) and (17), we deduce the oracle inequality of Theorem 3.

4.7 Proof of Theorem 4

Set $s > 0$. Define:

$$\ell^*(s) = \left\lceil \frac{1}{2s+d} \log n \right\rceil \qquad \text{and} \qquad h^*(s) = n^{-\frac{1}{2s+d}}.$$

Remark that

$$\frac{h^*(s)}{e} \leq h(\ell^*) \leq h^*(s) \tag{18}$$

and

$$s \leq m(\ell^*(s)) \leq s + 1. \tag{19}$$

We note that there exists $n_1 = n_1(s, p)$ such that for any $n \geq n_1$ we have $\ell^*(s) \in \mathcal{L}_{\text{iso}}$.

Note also that, using (19):

$$\max(\|W(\ell^*)\|_{2\wedge p}, \|W(\ell^*)\|_\infty) \leq (s+2)^{3d/2}. \tag{20}$$

Using (18) and (20), Proposition 2 and Theorem 3 lead to conclude the proof.

4.8 Proof of Lemma 1

It is worth noticing that the solution of the minimization problem (2) can be found explicitly since the Lagrangian condition implies that such a kernel is necessarily of the form (1). Now, for $p = 2$, remark that:

$$\|w_m\|_2^2 = (a^{(m)})^\top H_m a^{(m)} = (e_0^{(m)})^\top H_m^{-1} e_0^{(m)} = (m+1)^2.$$

If $p < 2$, using the Cauchy-Schwarz inequality we obtain:

$$\|w_m\|_p \leq m+1.$$

If $p > 2$, we consider $q \in \mathbb{N}^*$ such that $2^{q-1} < p \leq 2^q$. Using Jensen inequality we obtain:

$$\begin{aligned}\|w_m\|_p^p &= \int_0^1 |w_m(u)|^p \,\mathrm{d}u \\ &\leq \left(\int_0^1 |w_m(u)|^{2^q} \,\mathrm{d}u\right)^{\frac{p}{2^q}}.\end{aligned}$$

Now, it can be seen by induction that:

$$\|w_m\|_p \leq 2(m+1)^{3/2}.$$

Indeed, let us denote

$$\begin{aligned}u_q &= \int_0^1 |w_m(u)|^{2^q} \,\mathrm{d}u \\ &= \sum_{i_1,\ldots,i_{2^q}} \frac{a_{i_1}\cdots a_{i_{2^q}}}{1+\sum_{k=1}^{2^q} i_k} \\ &\leq 2^{q-1}(m+1)\left(\sum_{i_1,\ldots,i_{2^{q-1}}} \frac{a_{i_1}\cdots a_{i_{2^{q-1}}}}{1+\sum_{k=1}^{2^{q-1}} i_k}\right)\left(\sum_{j_1,\ldots,j_{2^{q-1}}} \frac{a_{j_1}\cdots a_{j_{2^{q-1}}}}{1+\sum_{k=1}^{2^{q-1}} j_k}\right)\end{aligned}$$

This implies that $v_q = 2^{q+1}u_q$ is such that

$$v_q \leq (m+1)v_{q-1}^2$$

with $v_1 = 4(m+1)^2$. Thus

$$v_q \leq 2^{2^q}(m+1)^{3.2^{q-1}-1}.$$

Then, combining previous results:

$$\|w_m\|_p \leq 2(m+1)^{3/2}.$$

Since $\|w_m\|_\infty = \lim_{p\to\infty} \|w_m\|_p$, this also implies the result in the sup-norm.

4.9 Proof of Lemma 2

For $r \leq 2$, since the Lebesgue measure of $\Delta_{d,\varepsilon}$ equals to 2^{-d}, we have

$$\mathbf{E}_f^n\|\xi_{W,h}\|_{r,\varepsilon} \leq 2^{-\frac{d(2-r)}{2r}}\mathbf{E}_f^n\|\xi_{W,h}\|_{2,\varepsilon} \leq 2^{-\frac{d(2-r)}{2r}}\|W\|_2.$$

Let us now assume that $r > 2$. Using the Rosenthal inequality we have

$$\mathbf{E}_f^n|\xi_{W,h}(t)|^r \leq (C_r^*)^r (V_h)^{r/2} \left\{ \left(\mathbf{E}_f^n \mathcal{K}_{W,h}^2(t, X_1) \right)^{r/2} + 2^{r+1} n^{1-r/2} \mathbf{E}_f^n |\mathcal{K}_{W,h}(t, X_1)|^r \right\}.$$

Using Jensen and Young inequalities we obtain:

$$\begin{aligned}
\mathbf{E}_f^n \|\xi_{W,h}\|_{r,\varepsilon} &\leq \left(\int_{\Delta_{d,\varepsilon}} \mathbf{E}_f^n |\xi_{W,h}(t)|^r \, \mathrm{d}t \right)^{1/r} \\
&\leq C_r^* \left\{ \Lambda_\varepsilon(W, h, r) + 2\|W\|_r (nV_h)^{\frac{1}{r} - \frac{1}{2}} \right\} \\
&\leq C_r^* \left\{ \Lambda_\varepsilon(W, h, r) + 2\|W\|_r \right\}.
\end{aligned}$$

We have

$$\begin{aligned}
\Lambda_\varepsilon(W, h, r) \leq & F_\infty^{1/2} \left(\int_{\Delta_{d,\varepsilon}} \left(V_h \int_{\Delta_d} \mathcal{K}_{W,h}^2(t, x) \, \mathrm{d}x \right)^{r/2} \mathrm{d}t \right)^{1/r} \\
\leq & F_\infty^{1/2} \|W\|_2.
\end{aligned}$$

As a consequence, for all $r \geq 1$, we have

$$\Gamma_\varepsilon(W, h, r) \leq C\|W\|_{r \vee 2}$$

where C depends on F_∞ and r.

4.10 Proof of Lemma 3

In the following, $\mathcal{L}$ is either $\mathcal{L}_{\mathrm{ani}}$ or $\mathcal{L}_{\mathrm{iso}}$. Let $\ell \in \mathcal{L}$. Let $W \in \{W(\ell), W^*(\ell)\}$. We denote by $\mathbb{B}_{r'}$ the unit ball of $\mathbb{L}_{r'}(\Delta_{d,\varepsilon})$ where $1/r + 1/r' = 1$ and, for $\lambda \in \mathbb{B}_{r'}$, we consider $\bar{g}_\lambda$ defined, for $x \in \Delta_d$ by:

$$\bar{g}_\lambda(x) = g_\lambda(x) - \mathbf{E}_f^n g_\lambda(X_1) \quad \text{with} \quad g_\lambda(x) = V_h^{1/2} \int_{\Delta_{d,\varepsilon}} \lambda(t) \mathcal{K}_{W,h}(t, x) \, \mathrm{d}t.$$

The variable $Y = \|\xi_{W,h}\|_{r,\varepsilon}$ satisfies

$$\begin{aligned}
Y &= \sup_{\|\lambda\|_{r',\varepsilon} \leq 1} \int_{\Delta_{d,\varepsilon}} \lambda(t) \xi_{W,h}(t) \, \mathrm{d}t \\
&= \sup_{\|\lambda\|_{r',\varepsilon} \leq 1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{g}_\lambda(X_i)
\end{aligned}$$

Since the set $\mathbb{B}_{r'}$ is a weakly$-*$ separable space, there exists a countable set $(\lambda_k)_{k \in \mathbb{N}}$ such that

$$Y = \sup_{k \in \mathbb{N}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{g}_{\lambda_k}(X_i).$$

Let us first assume that:

$$\|\bar{g}_{\lambda_k}\|_\infty \leq \mathfrak{b}(W, h, r) \tag{21}$$

and

$$\sup_{k\in\mathbb{N}} \mathbf{E}_f^n g_{\lambda_k}^2(X_1) \le \sigma^2(W,h,r), \tag{22}$$

where

$$\sigma^2(W,h,r) = \sigma^2 = \begin{cases} \|W\|_r^2 V_h^{\frac{2}{r}-1} & \text{if } 1 \le r < 2 \\ \mathrm{F}_\infty \|W\|_{2r/(r+2)}^2 V_h^{\frac{2}{r}} & \text{if } r \ge 2. \end{cases}$$

and

$$\mathfrak{b}(W,h,r) = \mathfrak{b} = 2\|W\|_r V_h^{1/r-1/2}.$$

Using the Bousquet inequality, and denoting $\Gamma_\varepsilon = \Gamma_\varepsilon(W,h,r)$, we obtain for any $x>0$:

$$\mathbf{P}\left(Y - \mathbf{E}_f^n Y \ge \frac{\Gamma_\varepsilon \tau}{2} + x\right) \le \exp\left(-\frac{x^2}{2\sigma^2 + \frac{\mathfrak{b}}{\sqrt{n}}\left(\Gamma_\varepsilon\left(\frac{12+\tau}{3}\right) + \frac{2x}{3}\right)}\right)$$
$$\times \exp\left(-\frac{\tau\Gamma_\varepsilon x + \Gamma_\varepsilon^2\tau^2/4}{2\sigma^2 + \frac{\mathfrak{b}}{\sqrt{n}}\left(\Gamma_\varepsilon\left(\frac{12+\tau}{3}\right) + \frac{2x}{3}\right)}\right)$$

Note that, for any $x>0$, we have

$$\frac{\tau\Gamma_\varepsilon x + \Gamma_\varepsilon^2\tau^2/4}{2\sigma^2 + \frac{\mathfrak{b}}{\sqrt{n}}\left(\Gamma_\varepsilon\left(\frac{12+\tau}{3}\right) + \frac{2x}{3}\right)} \ge \frac{\Gamma_\varepsilon^2\tau^2}{4\left(2\sigma^2 + \frac{\mathfrak{b}\Gamma_\varepsilon(12+\tau)}{3\sqrt{n}}\right)}.$$

This inequality holds due to the fact that the homography on the left hand side of the equation is an increasing function. Now, it can be easily proved that there exist absolute positive constants $\mathfrak{c}_1$ and $\mathfrak{c}_2$ that depend only on d, τ, r and F_∞ such that

$$\frac{4\left(2\sigma^2 + \frac{\mathfrak{b}\Gamma_\varepsilon(12+\tau)}{3\sqrt{n}}\right)}{\Gamma_\varepsilon^2\tau^2} \le \mathfrak{c}_1(h^*)^{d\tau_r} + \mathfrak{c}_2(h^*)^{d/r} \tag{23}$$

where $\tau_r = 2/r - 1$ if $1 \le r < 2$ or $\tau_r = 2/r$ if $r \ge 2$. This is a consequence of the fact that if $r \le r'$,

$$\|W\|_r \le \|W\|_{r'}.$$

Using (23) we obtain:

$$\exp\left(-\frac{\tau\Gamma_\varepsilon x + \Gamma_\varepsilon^2\tau^2/4}{2\sigma^2 + \frac{\mathfrak{b}}{\sqrt{n}}\left(\Gamma_\varepsilon\left(\frac{12+\tau}{3}\right) + \frac{2x}{3}\right)}\right) \le \exp(-C_1\alpha_n(r)),$$

where C_1 is an absolute positive constant that depends only on r, τ and F_∞.

Using Lemma 2, (21) and (22), we readily obtain that there exists an absolute constant $\mathfrak{c}_4$ which depends only on F_∞, τ and r such that:

$$2\sigma^2 + \frac{\mathfrak{b}}{\sqrt{n}}\left(\Gamma_\varepsilon\left(\frac{12+\tau}{3}\right) + \frac{2x}{3}\right) \le \mathfrak{c}_4(1+x)\|W\|_{r\vee 2}^2 \begin{cases} (h^*)^{d\left(\frac{2}{r}-1\right)} & \text{if } r<2 \\ (h^*)^{\frac{d}{r}} & \text{if } r \ge 2 \end{cases}$$

Now if $W \in \{W(\ell), W^*(\ell)\}$, with $\ell \in \mathcal{L}_{\mathrm{ani}}$, there exists a constant C_2 that depends on r, τ, F_∞ and W° such that

$$\mathbf{P}\left(Y - \mathbf{E}_f^n Y \ge \frac{\Gamma_\varepsilon\tau}{2} + x\right) \le \exp\left(-\frac{x^2}{C_2(1+x)}\right)\exp\left(-C_1\alpha_n(r)\right). \tag{24}$$

Moreover, if $W \in \{W(\ell), W^*(\ell)\}$, with $\ell \in \mathcal{L}_{\text{iso}}$, using Lemma 1, there exists a constant C_2 that depends on r, τ and F_∞ such that (24) holds. That concludes the lemma.

It remains to prove both (21) and (22).

Proof of (21): Set $k \in \mathbb{N}$:

$$\begin{aligned}\|\bar{g}_{\lambda_k}\|_\infty &\leq 2\|g_{\lambda_k}\|_\infty \\ &\leq 2 \sup_{x\in\Delta_d} V_h^{1/2}\|\lambda_k\|_{r',\varepsilon}\|\mathcal{K}_{W,h}(\cdot,x)\|_{r,\varepsilon} \\ &\leq 2\|W\|_r V_h^{1/r-1/2}\end{aligned}$$

Proof of (22): Assume that $r < 2$. Then, using the Hölder inequality:

$$\begin{aligned}\sup_{k\in\mathbb{N}} \mathbf{E}_f^n g_{\lambda_k}^2(X_1) &= V_h \sup_{k\in\mathbb{N}} \int_{\Delta_d} \left(\int_{\Delta_{d,\varepsilon}} \lambda_k(t)\mathcal{K}_{W,h}(t,x)\,\mathrm{d}t\right)^2 f(x)\,\mathrm{d}x \\ &\leq V_h \sup_{k\in\mathbb{N}} \int_{\Delta_d} \|\mathcal{K}_{W,h}(\cdot,x)\|_{r,\varepsilon}^2\|\lambda_k\|_{r',\varepsilon}^2 f(x)\,\mathrm{d}x \\ &= V_h^{2/r-1}\|W\|_r^2. \end{aligned} \tag{25}$$

Now assume that $r \geq 2$. Using the Young inequality:

$$\begin{aligned}\sup_{k\in\mathbb{N}} \mathbf{E}_f^n g_{\lambda_k}^2(X_1) &\leq \mathrm{F}_\infty V_h \sup_{k\in\mathbb{N}} \int_{\Delta_d} \left(\int_{\Delta_{d,\varepsilon}} \mathcal{K}_{W,h}(t,x)\lambda_k(t)\,\mathrm{d}t\right)^2 \mathrm{d}x \\ &\leq \mathrm{F}_\infty V_h^{2/r}\|W\|_{2r/(r+2)}^2. \end{aligned} \tag{26}$$

Combining (25) and (26), the result follows.

4.11 Proof of Lemma 4

We have

$$\mathbf{P}\left(\sum_{\varepsilon\in\{0,1\}^d} \|\xi_{W^*(\ell),h}\|_{r,\varepsilon}^{1/2} \geq \delta 2^d (nV_h)^{1/4}\right) \leq \sum_{\varepsilon\in\{0,1\}^d} \mathbf{P}_f^n\left(\|\xi_{W^*(\ell),h}\|_{r,\varepsilon} \geq \delta^2(nV_h)^{1/2}\right)$$

For $\ell \in \mathcal{L}_{\text{ani}}$, since $h \in \mathcal{H}_n$ and using Lemma 2, there exists $N_0 = N_0(\delta,\tau,F_\infty,W^\circ)$ such that for any $n \geq N_0$:

$$(1+\tau/2)\Gamma_\varepsilon(W^*(\ell),h,r) \leq \delta^2(nV_h)^{1/2}.$$

For $\ell \in \mathcal{L}_{\text{iso}}$, since $h \in \mathcal{H}_n$, using Lemma 1 and 2, there exists $N_0 = N_0(\delta,\tau,F_\infty)$ such that for any $n \geq N_0$:

$$\begin{aligned}(1+\tau/2)\Gamma_\varepsilon(W^*(\ell),h,r) &\leq (1+\tau/2)C_0\|W^*(\ell)\|_{2\vee r} \\ &\leq (1+\tau/2)C_0\left(\frac{\|W(\ell)\|_{4\vee 2r}}{\|W(\ell)\|_2}\right)^2 \\ &\leq (1+\tau/2)C_0(m(\ell)+1)^d \\ &\leq \delta^2(\log n)^{d+\frac{1}{2}} \\ &\leq \delta^2(nV_h)^{1/2}.\end{aligned}$$

Applying Lemma 3, we obtain the result of the lemma.

4.12 Proof of Lemma 5.

Let $(e_1, \dots, e_d)$ be the canonical basis of $\mathbb{R}^d$ and define

$$v_i(u) = (t_1 + \eta_1 u_1, \dots, t_{i-1} + \eta_{i-1} u_{i-1}, \quad t_i \quad, t_{i+1} + h_{i+1} u_{i+1}, \dots, t_d + h_d u_d).$$

We can write:

$$\begin{aligned} f(t + h \cdot u) - f(t + \eta \cdot u) &= \sum_{i=1}^{d} f(v_i(u) + h_i u_i e_i) - f(v_i(u) + \eta_i u_i e_i) \\ &= \sum_{i \in I} f(v_i(u) + h_i u_i e_i) - f(v_i(u)), \end{aligned}$$

where $I = \{i = 1, \dots, d : \eta_i = 0\}$. Using a Taylor expansion of the function $x \in \mathbb{R} \mapsto f(v_i(u) + x e_i)$ around 0, we obtain:

$$\begin{aligned} f(t + h \cdot u) - f(t + \eta \cdot u) &= \sum_{i \in I} \sum_{k=0}^{\lfloor s_i \rfloor} D_i^k f(v_i(u)) \frac{(h_i u_i)^k}{k!} \\ &\quad + \sum_{i \in I} \frac{(h_i u_i)^{\lfloor s_i \rfloor}}{\lfloor s_i \rfloor !} \int_0^1 \left[D_i^{\lfloor s_i \rfloor} f(v_i(u) + \tau h_i u_i) - D_i^{\lfloor s_i \rfloor} f(v_i(u)) \right] \mathrm{d}\tau. \end{aligned}$$

Using the facts that $v_i(u)$ does not depend on u_i and that $\int_{\Delta_1} W_i(y) y^k \,\mathrm{d}y = 0$ for any $1 \le k \le \lfloor s_i \rfloor$, Fubini's theorem implies that:

$$S^*_{W,h,\eta}(f) = \left(\int_{\Delta_{d,\mathbf{0}}} \left| \int_{\Delta_d} \left(\prod_{j=1}^{d} W_j(u_j) \right) \sum_{i \in I} I_i(u, h) \,\mathrm{d}u \right|^p \,\mathrm{d}t \right)^{1/p}$$

where

$$I_i(t, u, h) = \frac{(h_i u_i)^{\lfloor s_i \rfloor}}{\lfloor s_i \rfloor !} \int_0^1 \left[D_i^{\lfloor s_i \rfloor} f(v_i(u) + \tau h_i u_i) - D_i^{\lfloor s_i \rfloor} f(v_i(u)) \right] \mathrm{d}\tau.$$

Using Jensen's inequality and Fubini's theorem we obtain that:

$$S^*_{W,h,\eta}(f) = \left(d \|W\|_1 \right)^{1-1/p} \left(\int_{\Delta_d} J(u, h) \left| \prod_{j=1}^{d} W_j(u_j) \right| \mathrm{d}u \right)^{1/p},$$

where $J(u, h) = \sum_{i \in I} \int_{\Delta_{d,\mathbf{0}}} |I_i(t, u, h)|^p \,\mathrm{d}t$. Now, we study this last term:

$$J(u, h) \le \sum_{i \in I} \int_{\Delta_{d,\mathbf{0}}} \frac{(h_i u_i)^{1+p s_i}}{(\lfloor s_i \rfloor !)^p} \int_0^1 \frac{\left| D_i^{\lfloor s_i \rfloor} f(v_i(u) + \tau h_i u_i) - D_i^{\lfloor s_i \rfloor} f(v_i(u)) \right|^p}{|\tau h_i u_i|^{1 + p(s_i - \lfloor s_i \rfloor)}} \,\mathrm{d}\tau \,\mathrm{d}t.$$

Using a simple change of variables, we obtain:

$$J(u, h) \le \sum_{i \in I} \frac{(h_i u_i)^{p s_i}}{(\lfloor s_i \rfloor !)^p} \int_{\Delta_d} \int_0^1 \frac{\left| D_i^{\lfloor s_i \rfloor} f(x_1, \dots, x_{i-1}, \xi, x_{i+1}, \dots, x_d) - D_i^{\lfloor s_i \rfloor} f(x) \right|^p}{|\xi - x_i|^{1 + p(s_i - \lfloor s_i \rfloor)}} \,\mathrm{d}\xi \,\mathrm{d}x.$$

Since $u_i \le 1$ and $f \in \mathbb{S}_{s,p}(L)$ we have:

$$S^*_{W,h,\eta}(f) \le d \|W\|_1 \kappa(s) L \left(\sum_{i \in I} h_i^{p s_i} \right)^{1/p}.$$

This implies the result.

5 Acknowledgments

The authors have been supported by Fondecyt project 1141258, Mathamsud project 16-MATH-03 SIDRE and ECOS project C15E05.

References

Nicolas Asin and Jan Johannes. Adaptive non-parametric estimation in the presence of dependence. *arXiv preprint arXiv:1602.00531*, 2016.

Florent Autin, Erwan Le Pennec, and Karine Tribouley. Thresholding methods to estimate copula density. *Journal of Multivariate Analysis*, 101(1):200–222, 2010.

Karine Bertin and Nicolas Klutchnikoff. Minimax properties of beta kernel estimators. *Journal of Statistical Planning and Inference*, 141(7):2287–2297, 2011.

Karine Bertin and Nicolas Klutchnikoff. Adaptive estimation of a density function using beta kernels. *ESAIM: Probability and Statistics*, 18:400–417, 2014.

Zdravko I Botev, Joseph F Grotowski, Dirk P Kroese, et al. Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5):2916–2957, 2010.

Taoufik Bouezmarni and Jeroen V. K. Rombouts. Nonparametric density estimation for multivariate bounded data. *J. Statist. Plann. Inference*, 140(1):139–152, 2010. ISSN 0378-3758. doi: 10.1016/j.jspi.2009.07.013. URL http://dx.doi.org/10.1016/j.jspi.2009.07.013.

Adrian W Bowman. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):353–360, 1984.

Song Xi Chen. Beta kernel estimators for density functions. *Comput. Statist. Data Anal.*, 31(2):131–145, 1999. ISSN 0167-9473. doi: 10.1016/S0167-9473(99)00010-9. URL http://dx.doi.org/10.1016/S0167-9473(99)00010-9.

Shean-Tsong Chiu. Bandwidth selection for kernel density estimation. *The Annals of Statistics*, pages 1883–1905, 1991.

Daren B. H Cline and Jeffrey D Hart. Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics*, 22(1):69–84, 1991. doi: 10.1080/02331889108802286. URL http://www.tandfonline.com/doi/abs/10.1080/02331889108802286.

Luc Devroye and Laszlo Gyorfi. *Nonparametric density estimation: the L1 view*, volume 119. John Wiley & Sons Incorporated, 1985.

A. Goldenshluger and O. 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.

Alexander Goldenshluger and Oleg Lepski. Universal pointwise selection rule in multivariate function estimation. *Bernoulli*, 14(4):1150–1190, 2008. ISSN 1350-7265. doi: 10.3150/08-BEJ144. URL http://dx.doi.org/10.3150/08-BEJ144.

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.

M Chris Jones. Simple boundary correction for kernel density estimation. *Statistics and Computing*, 3(3):135–146, 1993.

Michel Lejeune and Pascal Sarda. Smooth estimators of distribution and density functions. *Comput. Statist. Data Anal.*, 14(4):457–471, 1992. ISSN 0167-9473. doi: 10.1016/0167-9473(92)90061-J. URL http://dx.doi.org/10.1016/0167-9473(92)90061-J.

O. V. Lepski. Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatnost. i Primenen.*, 36(4):645–659, 1991. ISSN 0040-361X. doi: 10.1137/1136085. URL http://dx.doi.org/10.1137/1136085.

James Stephen Marron and David Ruppert. Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 653–671, 1994.

Hans-Georg Müller. Smooth optimum kernel estimators near endpoints. *Biometrika*, 78 (3):pp. 521–530, 1991. ISSN 00063444. URL http://www.jstor.org/stable/2337021.

HG Müller and U Stadtmüller. Multivariate boundary kernels and a continuous least squares principle. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):439–458, 1999.

Eleonora Di Nezza, Giampiero Palatucci, and Enrico Valdinoci. Hitchhiker's guide to the fractional sobolev spaces. *Bulletin des Sciences Mathématiques*, 136(5):521 – 573, 2012. ISSN 0007-4497. doi: http://dx.doi.org/10.1016/j.bulsci.2011.12.004. URL http://www.sciencedirect.com/science/article/pii/S0007449711001254.

B. Opic and J. Rákosník. Estimates for mixed derivatives of functions from anisotropic Sobolev-Slobodeckij spaces with weights. *Quart. J. Math. Oxford Ser. (2)*, 42(167):347–363, 1991. ISSN 0033-5606. doi: 10.1093/qmath/42.1.347. URL http://dx.doi.org/10.1093/qmath/42.1.347.

Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

Murray Rosenblatt et al. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956.

Judith Rousseau et al. Rates of convergence for the posterior distributions of mixtures of betas and adaptive nonparametric estimation of the density. *The Annals of Statistics*, 38(1):146–180, 2010.

Mats Rudemo. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, pages 65–78, 1982.

Eugene F. Schuster. Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics - Theory and Methods*, 14(5):1123–1136, 1985. doi: 10.1080/03610928508828965. URL http://www.tandfonline.com/doi/abs/10.1080/03610928508828965.

Bernard W Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

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 http://dx.doi.org/10.1007/b13794. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.