Adaptive Density Estimation on Bounded Domains

Karine Bertin ∗  Salima El Kolei †  Nicolas Klutchnikoff ‡

January 23, 2017

Abstract

We study the estimation, in $L_p$-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 $\hat{f} : \Delta_d^n \to L_p(\Delta_d)$ where $p \geq 1$ is a fixed parameter. The accuracy of $\hat{f}$ is measured using the risk:

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

where $q$ is also a fixed parameter greater than or equal to 1 and $\mathbb{E}_f^n$ denotes the expectation with respect to the probability measure $P_f^n$ of the observations. Moreover the $L_p$-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 $\int_{\mathbb{R}^d} K(x) \, dx = 1$) and a bandwidth vector $h = (h_1, \ldots, h_d)$, such an estimator writes:

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

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 $L_p$ 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 $\Delta_d$. 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 $n^{-\bar{s}/(2\bar{s}+1)}$ with $\bar{s} = \left(\sum_{i=1}^{d} 1/s_i\right)^{-1}$. 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 $n^{-s/(2s+1)}$ 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 $\mathcal{H}_n = \{ h = (h_1, \ldots, h_d) \in (0, h^*)^d : nV_h \geq (\log n)^{2d+1}\}$ where $V_h = \prod_{i=1}^d h_i$. Define also:

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

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

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

where, for $x \in \Delta_d$,

$$K_{W,h}(t, x) = \prod_{i=1}^d \left( \frac{1}{h_i} W_i \left( \sigma(t_i) \frac{t_i - x_i}{h_i} \right) \right) \quad \text{with} \quad \sigma(u) = 2I_{(1/2, 1)}(u) - 1$$

where $I_{(a,b)}(u)$ 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 $\Delta_d$. 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 $\Delta_d$. The shape of the kernel adapts to the estimation point thanks to the introduction of the function $\sigma$. 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:

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

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

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

For any $\ell \in L_{\text{iso}}$, we consider

$$W(\ell) = (w_{m(\ell)}, \ldots, w_{m(\ell)}) \in \mathcal{W}^d$$

where the univariate kernel $w_m$ is defined, for any $m \in \mathbb{N}$, by:

$$(1) \quad w_m(u) = \sum_{r=0}^{m} a^{(m)}_r u^r, \quad u \in \Delta_1.$$

Figure 1 represents the kernels $\omega_m$ for different values of $m$.

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

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

**Remark 2.** The family $\{ \hat{f}_{\text{iso}}^{\ell} : \ell \in L_{\text{iso}} \}$ 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 $\hat{f}_{\text{iso}}^{\ell}$ 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 $L_p$-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:

$$\int_{\Delta_1} w(u) u^r \, du = 0, \quad r = 1, 2, \ldots, m.$$
Then
\[ w_m = \arg \min_{w \in W(m)} \|w\|_2 = (m + 1). \] (2)
and
\[ \|w_m\|_p \leq \begin{cases} m + 1 & \text{if } p \leq 2 \\ 2(m + 1)^{3/2} & \text{otherwise.} \end{cases} \]

2.3 Anisotropic family of estimators

Let \( W^o = (W^o_1, \ldots, W^o_d) \in W^d \) be such that, for any \( i = 1, \ldots, d, W^o_i \) is a bounded kernel and consider \( h(\ell) = (h_1(\ell), \ldots, h_d(\ell)) \) defined by:
\[
h_i(\ell) = e^{-\ell_i}, \quad i = 1, \ldots, d
\]
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 \( \{ \hat{f}^\text{ani}_\ell : \ell \in \mathcal{L}_{\text{ani}} \} \) by
\[
\hat{f}^\text{ani}_\ell = \hat{f}_{W^o, h(\ell)}.
\]
To make the notation similar to the isotropic case we define \( W(\ell) = W^o, \forall \ell \in \mathcal{L}_{\text{ani}} \).

Remark 3. This family is more classical than the one constructed in the previous section. All the estimators are defined using the same kernel \( W^o \) 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 \( \hat{f}_\ell \) then denotes \( \hat{f}^\text{ani}_\ell \) or \( \hat{f}^\text{iso}_\ell \). For \( \varepsilon \in \{0, 1\}^d \), \( h \in \mathcal{H}_n \) and \( W \in W^d \) we define:
\[
\Delta_{d,\varepsilon} = \prod_{j=1}^d \left( \varepsilon_j \frac{1 + \varepsilon_j}{2} \right), \quad \|W\|_p = \left\| \bigotimes_{i=1}^d W_i \right\|_p
\]
and
\[
\hat{\Lambda}_\varepsilon(W, h, p) = \sqrt{Vh} \left( \int_{\Delta_{d,\varepsilon}} \left( \frac{1}{n} \sum_{j=1}^n K_{W, h}^2(t, X_j) \right)^{p/2} dt \right)^{1/p}
\]
and
\[
\hat{\Gamma}_\varepsilon(W, h, p) = \begin{cases} 2^{d(2-p)/p^2} \|W\|_2 & \text{if } 1 \leq p \leq 2 \\ C_p^* \left( \hat{\Lambda}_\varepsilon(W, h, p) + 2\|W\|_p \right) & \text{if } p > 2 \end{cases}
\]
where \( C_p^* = 15p/\log p \) is the best known constant in the Rosenthal inequality. For any \( \ell, \ell' \in \mathcal{L} \) we consider:
\[
\hat{M}_p(\ell) = \frac{1}{\sqrt{nVh(\ell)}} \sum_{\varepsilon \in \{0, 1\}^d} \hat{\Gamma}_\varepsilon(W(\ell), h(\ell), p) \quad \text{and} \quad \hat{M}_p(\ell, \ell') = \hat{M}_p(\ell') + \hat{M}_p(\ell' \land \ell)
\]
where $\ell \land \ell'$ is the vector with coordinates $\ell_i \land \ell'_i = \min(\ell_i, \ell'_i)$. Now, for any $\tau > 0$ we define:

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

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

We then select

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

which leads to the final plug-in estimator defined by $\hat{f} = \hat{f}_{\hat{\ell}}$. In what follows we denote by $\hat{f}_ani$ and $\hat{f}_iso$ the resulting estimators.

**Remark 4.** This procedure is inspired by the method developed by Goldenshluger and Lepski. Here $\hat{M}_p(\ell)$ is an empirical version of an upper bound on the standard deviation of $\hat{f}_{\ell}$ and $\hat{B}_p(\ell)$ is linked with the bias term of this estimator, see (14). This implies that $\hat{f}$ realizes a tradeoff between $\hat{B}_p(\ell)$ and $(1 + \tau)\hat{M}_p(\ell)$. 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 $\hat{f}_ani$ performs almost as well as the best estimator from the collection $\{\hat{f}_{ani, \ell} : \ell \in L_{ani}\}$.

**Theorem 1.** Assume that $f : \Delta_d \to \mathbb{R}$ is a density function such that $\|f\|_\infty \leq F_\infty$. Then there exists a positive constant $\mathcal{R}_1$ that depends only on $F_\infty$, $W^\circ$, $p$, $q$ and $\tau$, such that, for any $n \geq 2$:

$$R_{n}^{(p, q)}(\hat{f}_ani, f) \leq \mathcal{R}_1 \inf_{\ell \in L_{ani}} \left\{ \|E_{f \ell} \hat{f}_ani - f\|_p + \max_{\ell' \in L_{ani}} \|E_{f \ell} \hat{f}_{ani} - E_{f \ell}\hat{f}_ani\|_p + \frac{1}{(nV_h(\ell))^{1/2}} \right\}.$$ 

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 $\Delta_d$ (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 $S_p(s, L)$ if:

- $f$ belongs to $L_p(\Delta_d)$.
- $\|D_i^{[s_i]} f\|_p \leq L$.
- The following property holds:

$$\sum_{i=1}^{d} I_i (D_i^{[s_i]} f) \leq L,$$
where \( D_i^k f \) denotes the \( k \)-th order partial derivative of \( f \) with respect to the variable \( x_i \), \([s_i]\) is the largest integer strictly smaller than \( s_i \) and

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

**Theorem 2.** Set \( M = (M_1, \ldots, M_d) \in \mathbb{N}^d \). Assume that \( W^s \) is such that \( W_i^s \) 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 \( \hat{f}_{\text{ani}} \) satisfies:

\[
\limsup_{n \to +\infty} n^{\frac{s}{p+q}} \sup_{f \in \mathcal{S}_{p,L}} R_n^{(p,q)}(\hat{f}_{\text{ani}}, f) < +\infty,
\]

where

\[
\frac{1}{s} = \sum_{i=1}^d \frac{1}{s_i}.
\]

The third result is an oracle-type inequality for the family \( \hat{f}_{\text{iso}} \). This theorem allows us to derive an adaptive result over the isotropic classes \( \mathcal{S}_{p,L} \) (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 \( \|f\|_{\infty} \leq F_{\infty} \). Then there exists a positive constant \( R_2 \) that depends only on \( F_{\infty} \), \( p \), \( q \) and \( \tau \), such that, for any \( n \geq 2 \):

\[
R_n^{(p,q)}(\hat{f}_{\text{iso}}, f) \leq R_2 \inf_{\ell \in \mathcal{L} \leq 1} \left\{ \max_{\ell' \geq \ell} \|E_{\ell}^{n f_{\hat{f}_{\text{iso}}}} - f\|_p + \frac{\|W(\ell)\|_{p \tau / 2}}{(n V_{h(\ell)}^{1/2})} \right\}.
\]

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

- for any \( \alpha \in \mathbb{N}^d \), such that \( |\alpha| \leq |s| = \max\{i \in \mathbb{N} : i < s\} \), the mixed partial derivatives

\[
D_{x_1}^{\alpha_1} \cdots D_{x_d}^{\alpha_d} f
\]

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

- the Gagliardo semi-norm \( \|f\|_{s,p} \) is bounded by \( L \) where

\[
\|f\|_{s,p} = \left( \sum_{|\alpha| = |s|} \int_{\Delta_d} \frac{|D_{x_1}^{\alpha_1} \cdots D_{x_d}^{\alpha_d} f(y) - D_{x_1}^{\alpha_1} \cdots D_{x_d}^{\alpha_d} f(x)|^p}{|y - x|^{d+p(s-[s])}} \, dy \right)^{1/p}.
\]

**Theorem 4.** For any \( s > 0 \) and \( L > 0 \), the estimator \( \hat{f}_{\text{iso}} \) satisfies:

\[
\limsup_{n \to +\infty} n^{\frac{s}{p+q}} \sup_{f \in \mathcal{S}_{s,p}(L)} R_n^{(p,q)}(\hat{f}_{\text{iso}}, f) < +\infty,
\]

**Remark 5.** In Theorems 1 and 3, the right hand sides of the equations can be easily interpreted. In both situations, the term \( (n V_{h(\ell)}^{1/2})^{-1/2} \) is of the order of the standard deviation of \( \hat{f}_{\ell} \). Moreover the terms \( \max_{\ell' \leq \ell} \|E_{\ell}^{n f_{\hat{f}_{\text{iso}}}} - E_{\ell'}^{n f_{\hat{f}_{\text{iso}}}}\|_p \) and \( \max_{\ell' \geq \ell} \|E_{\ell}^{n f_{\hat{f}_{\text{iso}}}} - f\|_p \) 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 \( \|E_{\ell}^{n f_{\hat{f}_{\ell}} - f}\|_p \) 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 \( W = (W_1, \ldots, W_d) \in \mathcal{W}_d \), \( h \in \mathcal{H}_n \) and \( \varepsilon \in \{0, 1\}^d \), we define the quantity:

\[
\Gamma_{\varepsilon}(W,h,p) =\begin{cases}
2^{-\frac{d(2-p)}{2p}}\|W\|_2 & \text{if } 1 \leq p \leq 2 \\
C_p^\varepsilon (\Lambda_{\varepsilon}(W,h,p) + 2\|W\|_p) & \text{if } p > 2
\end{cases}
\]

where

\[
\Lambda_{\varepsilon}(W,h,p) = \sqrt{V_h} \left( \int_{\Delta_d,\varepsilon} \left( \int_{\Delta_d} K_{W,h}(t,x) f(x) \, dx \right)^{p/2} \, dt \right)^{1/p}.
\]

For \( g : \Delta_d \to \mathbb{R} \) and \( r \geq 1 \) we denote

\[
\|g\|_{r,\varepsilon} = \left( \int_{\Delta_d,\varepsilon} |g(x)|^r \, dx \right)^{1/r}.
\]

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

\[
\xi_{W,h}(t) = \left( \frac{V_h}{n} \right)^{1/2} \sum_{i=1}^{n} \left( K_{W,h}(t,X_i) - \mathbf{E}_{\varepsilon} K_{W,h}(t,X_i) \right), \quad t \in \Delta_d.
\]

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

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

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

\[
\sup_{f \in S_{s,p}(L)} \max_{\ell \in \mathcal{L}_{\text{ani}}} \|E_f^{n\text{ani}} - f\|_p \leq 2^{d/p} \left( \prod_{i=1}^d (M_i + 1) \right) L \sum_{i=1}^d (h_i(\ell))^{s_i}.
\]

\[
\max_{k \in \mathcal{L}_{\text{ani}}} \|E_f^{n\text{ani}} - E_f^{n\text{ani}}\|_p \leq 2^{1+d/p} \left( \prod_{i=1}^d (M_i + 1) \right) L \sum_{i=1}^d (h_i(\ell))^{s_i}.
\]

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

\[
\sup_{f \in S_{s,p}(L)} \max_{\ell \geq \ell} \|E_f^{n\text{iso}} - f\|_p \leq \mathfrak{R}_3 \left( \|W(\ell)\|_\infty L(h(\ell))^s + \sqrt{\frac{h^*}{n}} \right),
\]

where the positive constant \( \mathfrak{R}_3 \) depends only on \( d, p, s \) and \( L \).

**Proposition 3.** Set \( p, q \geq 1 \). Assume that \( f \) is such that \( \|f\|_\infty \leq F_\infty \).
• Let $\ell \in \mathcal{L}_{iso}$. There exists a positive constant $\mathcal{R}_4$ that depends only on $p$, $q$, $\tau$ and $F_\infty$ such that
\[
\mathbb{E}_f^p \left\{ \left\| \hat{f}_\ell^{iso} - \mathbb{E}_f^{n} \hat{f}_\ell^{iso} \right\|_p - (1 + \tau) \bar{M}_p(\ell) \right\}^q \leq \mathcal{R}_4 n^{-q}.
\]
• Let $\ell \in \mathcal{L}_{ani}$. There exists a positive constant $\mathcal{R}_5$ that depends only on $p$, $q$, $\tau$, $W^o$ and $F_\infty$ such that
\[
\mathbb{E}_f^p \left\{ \left\| \hat{f}_\ell^{ani} - \mathbb{E}_f^{n} \hat{f}_\ell^{ani} \right\|_p - (1 + \tau) \bar{M}_p(\ell) \right\}^q \leq \mathcal{R}_5 n^{-q}.
\]

**Lemma 2.** Assume that $f$ satisfies $\|f\|_\infty \leq F_\infty$. For any $W \in \mathcal{W}^d$, $r \geq 1$ and $h \in \mathcal{H}_n$, we have:
\[
\mathbb{E}_f^p \| \xi_{W,h} \|_{r,\varepsilon} \leq \Gamma_{\varepsilon}(W, h, r) \leq C_0 \|W\|_{2\nu r},
\]
where $C_0$ is an absolute constant that depends only on $r$ and $F_\infty$.

**Lemma 3.** Assume that $f$ satisfies $\|f\|_\infty \leq F_\infty$. We have, for any $\ell \in \mathcal{L}_{iso}$, $h \in \mathcal{H}_n$, $W \in \{W(\ell), W^*(\ell)\}$ and $r \geq 1$,
\[
P(\|\xi_{W,h}\|_{r,\varepsilon} - \mathbb{E}_f^p \|\xi_{W,h}\|_{r,\varepsilon} \geq \tau \Gamma_{\varepsilon}(W, h, r) + x) \leq \exp \left( -\frac{x^2}{C_2(1+x)} \right) \left( \frac{n}{\exp(-C_1 \alpha_n(r))} \right) \quad (5)
\]
where $C_1$ and $C_2$ are absolute constants that depend only on $r$, $\tau$ and $F_\infty$ and
\[
\alpha_n(r) = \begin{cases} (h^*)^{-d(\frac{q}{2} - 1)} & \text{if } 1 \leq r < 2 \\ (h^*)^{-\frac{q}{2}} & \text{if } r \geq 2 \end{cases}
\]

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

**Lemma 4.** Assume that $f$ satisfies $\|f\|_\infty \leq F_\infty$. We have, for any $\ell \in \mathcal{L}_{iso}$, $h \in \mathcal{H}_n$, $\delta > 0$ and $r \geq 1$,
\[
P \left( \sum_{x \in \{0,1\}^d} \|\xi_{W^*(\ell),h}\|_{r,\varepsilon}^{1/2} \geq \delta 2^d \left( n V_h \right)^{1/4} \right) \leq C_3 \exp \left( -C_1 \alpha_n(r) \right), \quad (6)
\]
where $C_3$ is an absolute constant that depends only on $r$, $\delta$ and $F_\infty$.

If $\ell \in \mathcal{L}_{ani}$, (6) holds also with $C_3$ that depends on $r$, $\delta$, $W^o$ and $F_\infty$.

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

**Lemma 5.** Let $h = (h_1, \ldots, h_d)$ and $\eta = (\eta_1, \ldots, \eta_d)$ be two bandwidths in $\mathcal{H}_n$ such that $\eta_i \in \{0, h_i\}$. Set $W = (w_{M_1}, \ldots, w_{M_d}) \in \mathcal{W}^d$ and define:
\[
S_{W,h,\eta}(f) = \left( \int_{\Delta_{d,0}} \left| \int_{\Delta_d} \left( \prod_{j=1}^d w_{M_j}(u_j) \right) \left( f(t + h \cdot u) - f(t + \eta \cdot u) \right) du \right|_p \right)^{1/p}
\]
where $h \cdot u$ denotes the coordinate-wise product of the vectors $h$ and $u$. Assume that $f$ belongs to $S_{s,p}(L)$ and that, for any $i = 1, \ldots, d$, the kernel $W_i$ is of order greater than or equal to $[s_i]$. 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 $I = \{i = 1, \ldots, d : \eta_i = 0\}$. 

9
4.1 Proof of Proposition 1

We first prove (3). For the sake of brevity, define $W = W^0$ and $h = h(\ell)$.

$$
\|E^n f^{\text{ani}} - f\|_p^p = \sum_{\varepsilon\in\{0,1\}^d} \int_{\Delta_{d,\varepsilon}} |E^n f_{W,h}(t) - f(t)|^p \, dt
$$

$$
= \sum_{\varepsilon\in\{0,1\}^d} \int_{\Delta_{d,\varepsilon}} \int_{\Delta_d} K_{W,h}(t,x) f(x) \, dx - f(t) \bigg| \, dt
$$

$$
= \sum_{\varepsilon\in\{0,1\}^d} (S_{W,h}(f^{(\varepsilon)}))^p
$$

where

$$
S_{W,h}(f) = \left( \int_{\Delta_{d,0}} \int_{\Delta_d} K_{W,h}(t,x) f(x) \, dx - f(t) \bigg| \, dt \right)^{1/p}
$$

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 $K_{W,h}$. Since $f \in S_{s,p}(L) \iff f^{(\varepsilon)} \in S_{s,p}(L)$ we obtain

$$
\sup_{f \in S_{s,p}(L)} \|E^n f_{W,h} - f\|_p \leq 2^{d/p} \sup_{f \in S_{s,p}(L)} S_{W,h}(f).
$$

Using the notation introduced in Lemma 5 we have:

$$
\sup_{f \in S_{s,p}(L)} \|E^n f_{W,h} - f\|_p \leq 2^{d/p} \sup_{f \in S_{s,p}(L)} S_{W,h,0}^p(f).
$$

Equation (3) follows.

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

$$
\|E^n f_{W,h} - E^n f_{W,h'}\|_p^p \leq 2^d \int_{\Delta_{d,0}} \int_{\Delta_d} K_{W,h}(t,x) f(x) \, dx - \int_{\Delta_d} K_{W,h'}(t,x) f(x) \, dx \bigg| \, dt
$$

$$
\leq 2^d \int_{\Delta_{d,0}} \int_{\Delta_d} \left( \prod_{i=1}^d W_i(u_i) \right) \left[ f(t + h \cdot u) - f(t + h' \cdot u) \right] du \bigg| \, dt.
$$

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:

$$
\|E^n f_{W,h} - E^n f_{W,h'}\|_p^p \leq 2^{d+p} \max_{H \in \{h,h'\}} \int_{\Delta_{d,0}} \int_{\Delta_d} \left( \prod_{i=1}^d W_i(u_i) \right) \left[ f(t + H \cdot u) - f(t + \eta \cdot u) \right] du \bigg| \, dt.
$$

Using Lemma 5, we obtain:

$$
\|E^n f_{W,h} - E^n f_{W,h'}\|_p \leq 2^{1+d/p} \left( \prod_{i=1}^d (M_i + 1) \right) L \max_{H \in \{h,h'\}} \sum_{i \in I} H_i^{\eta_i}
$$

where $I = \{i : \eta_i = 0\}$. Since $H_i \leq h_i(\ell)$ 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:

$$\sup_{f \in \tilde{S}_{s,p}(L)} \left\| \mathbf{E}_f^{n} \hat{f}_{s_{k}} - f \right\|_p \leq 2^{d/p} \sup_{f \in \tilde{S}_{s,p}(L)} S_{W(\ell),h(\ell)}(f),$$

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

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

and

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

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

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

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

where

$$I_a = h^{pk} \int_{\Delta,0} \int_{\Delta} \int_0^1 \left\| (D^a f(t + \tau hu) - D^a f(t)) \right\|_p^p \, \tau \, d\tau \, du \, dt$$

$$\leq h^{pk} \int_{\Delta,0} \int_{\Delta} \int_0^1 \tau^{d+p(\varsigma-k)} \frac{\|D^a f(t + \tau hu) - D^a f(t)\|_p^p}{\|\tau hu\|_2^{d+p(\varsigma-k)}} \, d\tau \, du \, dt$$

$$\leq d^{(d+p)/2} h^{pk} \int_0^1 \int_{\Delta,0} \int_{\Delta} \frac{|D^a f(x) - D^a f(t)|_p^p}{\|x - t\|_2^{d+p(\varsigma-k)}} \, dx \, dt \, d\tau$$

$$\leq d^{(d+p)/2} \tilde{L}^{p} h^{pk}.$$

We thus obtain

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

(9)

where

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

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

$$\left\| \mathbf{E}_f^{n} \hat{f}_{s_{k}} - f \right\|_p \leq [C(d,p,s)\|W(\ell)\|_\infty] L(h(\ell))^s.$$  

(10)

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

$$\left\| \mathbf{E}_f^{n} \hat{f}_{s_{k}} - f \right\|_p \leq [C(d,p,s)\|W(\ell)\|_\infty] \tilde{L}(h(\ell))^{m(\ell)+1}.$$
Remark that
\[
(h(\ell))^{m(\ell)+1} = \exp(-\ell(m(\ell)+1))
\leq \exp \left(-\ell \left( \frac{\log n}{2\ell} + \frac{1}{2} \right) \right)
\leq \sqrt{\frac{h^s}{n}}.
\]
Thus, using Lemma 1, for \(m(\ell) < \lfloor s \rfloor\) we obtain:
\[
\| \mathbb{E}_f^n \hat{f}_\ell - f \|_p \leq \left[ C(d,p,s)(\lfloor s \rfloor + 1)^{3d/2} \right] \tilde{L} \sqrt{\frac{h^s}{n}}.
\]
Combining (10) and (11) we obtain the proposition.

4.3 Proof of Proposition 3

In the following, \(L\) is either \(L_{ani}\) or \(L_{iso}\) and \(\hat{f}_\ell\) then denotes \(\hat{f}_{ani,\ell}\) or \(\hat{f}_{iso,\ell}\). Let \(\ell \in L\). We define
\[
M_p(\ell) = \frac{1}{\sqrt{n V_h(\ell)}} \sum_{\varepsilon \in \{0, 1\}^d} \Gamma_\varepsilon(W(\ell), h(\ell), p).
\]
First, assume that \(1 \leq p \leq 2\). In this case \(M_p(\ell) = \tilde{M}_p(\ell)\), which implies that
\[
\mathbb{E}_f^n \left\{ \| \hat{f}_\ell - \mathbb{E}_f^n \hat{f}_\ell \|_p - (1 + \tau)\tilde{M}_p(\ell) \right\}_{+}^q \leq A_{p,q}(\ell)
\]
where
\[
A_{p,q}(\ell) = \mathbb{E}_f^n \left\{ \| \hat{f}_\ell - \mathbb{E}_f^n \hat{f}_\ell \|_p - (1 + \tau/2)M_p(\ell) \right\}_{+}^q.
\]
Next, assume that \(p > 2\). Denoting \(\delta = \frac{\tau}{2(1+\tau)}\), we have
\[
\left\{ \| \hat{f}_\ell - \mathbb{E}_f^n \hat{f}_\ell \|_p - (1 + \tau)\tilde{M}_p(\ell) \right\}_{+} \leq \left\{ \| \hat{f}_\ell - \mathbb{E}_f^n \hat{f}_\ell \|_p - (1 + \tau/2)M_p(\ell) \right\}_{+} + (1 + \tau/2)M_p(\ell)I_{\mathcal{D}_{\delta,\ell}} + \left\{ \| \hat{f}_\ell - \mathbb{E}_f^n \hat{f}_\ell \|_p - (1 + \tau)\tilde{M}_p(\ell) \right\}_{+} I_{\mathcal{D}_{\delta,\ell}},
\]
where
\[
\mathcal{D}_{\delta,\ell} = \left\{ \sum_{\varepsilon \in \{0, 1\}^d} \| \xi_{W^*(\ell),h(\ell)} \|_{p/2,\varepsilon}^{1/2} \leq \delta 2^d (n V_h(\ell))^{1/4} \right\}.
\]
This implies:
\[
\mathbb{E}_f^n \left\{ \| \hat{f}_\ell - \mathbb{E}_f^n \hat{f}_\ell \|_p - (1 + \tau)\tilde{M}_p(\ell) \right\}_{+}^q \leq 3^{q-1} (A_{p,q}(\ell) + B_{p,q}(\ell) + C_{p,q}(\ell))
\]
where
\[
B_{p,q}(\ell) = (1 + \tau/2)^q (M_p(\ell))^{q} \mathbb{P}_f^n \left( \mathcal{D}_{\delta,\ell} \right)
\]
\[ C_{p,q}(\ell) = E^n_f \left( \left\{ \| \hat{f}_\ell - E^n_f \hat{f}_\ell \|_p - \frac{(1 + \tau/2) \Gamma_\varepsilon(W(\ell), h(\ell), p)}{\sqrt{nV_{h(\ell)}}} \right\}^q + I_{D_{n,s}} \right). \]

Using Lemma 4 with \( r = \frac{p}{2} \), we immediately obtain that
\[ B_{p,q}(\ell) = O(n^{-q}). \]

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
\[ A_{p,q}(\ell) \leq E^n_f \left\{ \sum_{\varepsilon \in \{0,1\}^d} \| \hat{f}_\ell - E^n_f \hat{f}_\ell \|_{p,\varepsilon} - \frac{(1 + \tau/2) \Gamma_\varepsilon(W(\ell), h(\ell), p)}{\sqrt{nV_{h(\ell)}}} \right\}^q + I_{D_{n,s}}, \]

where
\[ I_{q,\varepsilon} = E^n_f \left\{ \sum_{\varepsilon \in \{0,1\}^d} \| \hat{f}_\ell - E^n_f \hat{f}_\ell \|_{p,\varepsilon} - \frac{(1 + \tau/2) \Gamma_\varepsilon(W(\ell), h(\ell), p)}{\sqrt{nV_{h(\ell)}}} \right\}^q. \]

Thus, using Lemma 2 and Lemma 3 with \( r = p \) we can write:
\[ (nV_{h(\ell)})^{q/2} I_{q,\varepsilon} = E^n_f \left\{ \sum_{\varepsilon \in \{0,1\}^d} \| \hat{f}_\ell - E^n_f \hat{f}_\ell \|_{p,\varepsilon} - \frac{(1 + \tau/2) \Gamma_\varepsilon(W(\ell), h(\ell), p)}{\sqrt{nV_{h(\ell)}}} \right\}^q \]
\[ \leq q \int_0^{+\infty} y^{q-1} P^n_f \left( \sum_{\varepsilon \in \{0,1\}^d} \| \hat{f}_\ell - E^n_f \hat{f}_\ell \|_{p,\varepsilon} - \frac{(1 + \tau/2) \Gamma_\varepsilon(W(\ell), h(\ell), p)}{\sqrt{nV_{h(\ell)}}} > y \right) dy \]
\[ \leq q \int_0^{+\infty} y^{q-1} P^n_f \left( \sum_{\varepsilon \in \{0,1\}^d} \| \hat{f}_\ell - E^n_f \hat{f}_\ell \|_{p,\varepsilon} - \frac{(1 + \tau/2) \Gamma_\varepsilon(W(\ell), h(\ell), p)}{\sqrt{nV_{h(\ell)}}} > \frac{\tau}{2} \right) dy \]
\[ \leq q \exp(-C_1 \alpha_n(p)) \int_0^{+\infty} y^{q-1} \exp \left( -\frac{y^2}{C_2 (1 + y)} \right) dy \]
\[ \leq C \exp(-C_1 \alpha_n(p)) \]

where \( C \) depends only on \( C_2 \) and \( q \).

This implies that
\[ A_{p,q}(\ell) = O(n^{-q}). \]

**Control of \( C_{p,q}(\ell) \).** Recall that \( p \geq 2 \). Let us remark that
\[ \left| \hat{M}_p(\ell) - M_p(\ell) \right| = \left| \sum_{\varepsilon \in \{0,1\}^d} \frac{C^n_p \| W(\ell) \|_2}{(nV_{h(\ell)})^{1/2}} Z_{\varepsilon}(\ell, p) \right| \]

where
\[ Z_{\varepsilon}(\ell, p) = \left( \int_{\Delta_{d,s}} \left( E^n_f K\star(\ell,h(t,X_i)) \right)^{p/2} dt \right)^{1/p} - \left( \int_{\Delta_{d,s}} \left( \frac{1}{n} \sum_{i=1}^n K\star(\ell,h(t,X_i)) \right)^{p/2} dt \right)^{1/p}. \]
We have
\[ |Z_\varepsilon(\ell, p)| = \left| \sqrt{\frac{\|E_f K_{W^*(\ell),h_\ell}(\cdot, X_1)\|_{p/2,\varepsilon}}{\|\sum_{i=1}^n K_{W^*(\ell),h_\ell}(\cdot, X_1)\|_{p/2,\varepsilon}}} - \sqrt{\frac{1}{n} \sum_{i=1}^n F_\varepsilon(\ell, p) - F_\varepsilon(\ell, p)|_{p/2,\varepsilon}} \right| \leq (nV_{h_\ell})^{-1/4} \|\xi_{W^*(\ell),h_\ell}(\cdot)\|_{p/2,\varepsilon}^{1/2}. \]

This implies that
\[ \left| \tilde{M}_\varepsilon(\ell) - M_\varepsilon(\ell) \right| \leq \frac{C_p^*\|W(\ell)\|_2}{(nV_{h_\ell})^{3/4}} \sum_{\varepsilon \in \{0,1\}^d} \|\xi_{W^*(\ell),h_\ell}(\cdot)\|_{p/2,\varepsilon}^{1/2}. \] (13)

Thus, under \( D_{\delta, \ell} \) defined in (12) we have
\[ \left| \tilde{M}_\varepsilon(\ell) - M_\varepsilon(\ell) \right| \leq \frac{2^dC_p^*\|W(\ell)\|_2}{(nV_{h_\ell})^{3/4}} \delta(nV_{h_\ell})^{1/4} \leq \delta M_\varepsilon(\ell), \]

and, since \((1 - \delta)(1 + \tau) = 1 + \tau/2:\)
\[ \left(1 + \tau\right)\tilde{M}_\varepsilon(\ell) \geq \left(1 + \tau/2\right)M_\varepsilon(\ell). \]

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 \( i = 1, \ldots, d \), we have \( \ell_i \leq \ell'_i \). Let \( \ell \in \mathcal{L}_{so} \) be an arbitrary multiindex. To simplify the notation, we use \( \hat{f}_\ell = f_\ell^{\text{ani}} \) and \( \hat{f} = f_\ell^{\text{ani}} \).

Using the definitions of \( \hat{B}_\varepsilon(\ell) \) and \( \hat{M}_\varepsilon(\ell) \) we easily obtain:
\[ \|f - \hat{f}\|_p \leq \|f - \hat{f}_\ell\|_p + \|\hat{f}_{\ell,\ell} - \hat{f}_\ell\|_p + \|\hat{f}_{\ell,\ell} - \hat{f}_\ell\|_p \leq \|f - \hat{f}_\ell\|_p + \hat{B}_\varepsilon(\ell) + (1 + \tau)\hat{M}_\varepsilon(\ell, \ell) + \hat{B}_\varepsilon(\ell) + (1 + \tau)\hat{M}_\varepsilon(\ell, \hat{\ell}). \]

Using the definition of \( \hat{\ell} \), we deduce:
\[ \|f - \hat{f}\|_p \leq \|f - \hat{f}_\ell\|_p + 2\hat{B}_\varepsilon(\ell) + (1 + \tau)\hat{M}_\varepsilon(\ell, \ell) + 2\left(1 + \tau\right)\hat{M}_\varepsilon(\ell, \hat{\ell}) \leq \|f - \hat{f}_\ell\|_p + 2\hat{B}_\varepsilon(\ell) + 4(1 + \tau)\max_{\ell' \preceq \ell} M_\varepsilon(\ell') + 4(1 + \tau)\max_{\ell' \preceq \ell} \left(\hat{M}_\varepsilon(\ell') - M_\varepsilon(\ell')\right). \]

This implies that:
\[ R_n^{(p,q)}(\hat{f}, f) \leq R_n^{(p,q)}(\hat{f}_\ell, f) + 2\left(\mathbf{E}_f^{\ell, q}\hat{B}_\varepsilon(\ell)\right)^{1/q} + 4(1 + \tau)\left(\mathbf{E}_f^{\ell, q}\max_{\ell' \preceq \ell} |\hat{M}_\varepsilon(\ell') - M_\varepsilon(\ell')|^q\right)^{1/q} + 4(1 + \tau)\max_{\ell' \preceq \ell} M_\varepsilon(\ell'). \]

It remains to bound each term of the right hand side of this inequality.
Remark that, using the triangular inequality, we have:

\[
\hat{B}_p(\ell) \leq 2 \max_{\ell' \in \mathcal{L}} \left\{ \| \hat{f}_p - \mathbf{E}_p^0 \hat{f}_p \|_p - (1 + \tau) \hat{M}_p(\ell') \right\} + \max_{\ell' \in \mathcal{L}} \| \mathbf{E}_p^0 \hat{f}_p - \mathbf{E}_p^0 \hat{f}_{\ell' \wedge \ell} \|_p.
\]

This readily implies

\[
\left( \mathbb{E}_p^0 \hat{B}_p^q(\ell) \right)^{1/q} \leq 2 \sum_{\ell' \in \mathcal{L}} \left( \mathbb{E}_p^0 \left\{ \| \hat{f}_p - \mathbf{E}_p^0 \hat{f}_p \|_p - (1 + \tau) \hat{M}_p(\ell') \right\}^q \right)^{1/q} + \max_{\ell' \in \mathcal{L}} \| \mathbf{E}_p^0 \hat{f}_p - \mathbf{E}_p^0 \hat{f}_{\ell' \wedge \ell} \|_p
\]

\[
\leq \frac{1}{n^{1/2}} \left( \# \mathcal{L} \right)^{1/2} \text{n}^{-1} + \max_{\ell' \in \mathcal{L}} \| \mathbf{E}_p^0 \hat{f}_p - \mathbf{E}_p^0 \hat{f}_{\ell' \wedge \ell} \|_p,
\]

(14)

where the last inequality follows immediately from Proposition 3.

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

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

\[
| \hat{M}_p(\ell') - M_p(\ell') | \leq \frac{C_p \| W^o \|_2}{(n V_{h(\ell)})^{1/2} (n V_{h(\ell')}))^{1/4} \sum_{\varepsilon \in \{0,1\}^d} \| \xi_{W^o(\varepsilon), h(\ell')} \|_{p/2, \varepsilon}^{1/2}.
\]

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

\[
\left( \mathbb{E}_p^0 \max_{\ell' \leq \ell} | \hat{M}_p(\ell') - M_p(\ell') |^q \right)^{1/q} \leq C \frac{\| W^o \|_2}{(n V_{h(\ell)})^{1/2}}.
\]

3. By using Lemma 2 we obtain

\[
M_p(\ell) = \frac{1}{\sqrt{n V_{h(\ell)}}} \sum_{\varepsilon \in \{0,1\}^d} \Gamma_{\varepsilon}(W^o, h(\ell), p)
\]

\[
\leq C \frac{\| W^o \|_{p/2}}{\sqrt{n V_{h(\ell)}}}.
\]

(15)

This implies that for \( \ell' \leq \ell \)

\[
4 \max_{\ell' \leq \ell} M_p(\ell') \leq \frac{C}{\sqrt{n V_{h(\ell)}}}.
\]

4.5 Proof of Theorem 2

Set \( s \in \prod_{i=1}^{d} (0, M_i + 1) \). Define \( \ell^s(0) = (\ell^s_1(0), \ldots, \ell^s_d(0)) \) by:

\[
\ell^s_i(0) = \left[ \frac{\bar{s}}{s_i(2s + 1) \log n} \right], \quad i = 1 \ldots, d
\]
where $\lceil x \rceil$ denotes the least integer greater than or equal to $x$. Note that $h_i(\ell^*)$ is such that
\[ \frac{h_i^*(s)}{e} \leq h_i(\ell^*) \leq h_i^*(s) \]  
(16)
where
\[ h_i^*(s) = n^{-\frac{s}{r + s + 1}}. \]
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
\[ \| f - \widehat{f}_{\text{iso}} \|_p \leq \| f - \widehat{f}_{\ell} \|_p + 4 \left( \widehat{B}_p(\ell) + (1 + \tau)\widehat{M}_p(\ell) \right), \]
and then
\[ R_n^{(p,q)}(\widehat{f}_{\text{iso}}, f) \leq R_n^{(p,q)}(\widehat{f}_{\ell}, f) + 4 \left( \mathbf{E}^n \widehat{B}_p^q(\ell) \right)^{1/q} + 4(1 + \tau)\left( \mathbf{E}^n |\widehat{M}_p(\ell) - M_p(\ell)|^q \right)^{1/q} + 4(1 + \tau)M_p(\ell). \]

Following the same arguments of the proof of Theorem 1 (see the second paragraph), we have
\[ \left( \mathbf{E} |\widehat{M}_p(\ell) - M_p(\ell)|^q \right)^{1/q} \leq C \frac{\| W(\ell) \|_2}{(nV_h(\ell))^{1/2}}. \]
(17)
Applying (14), (15) and (17), we deduce the oracle inequality of Theorem 3.

### 4.7 Proof of Theorem 4

Set $s > 0$. Define:
\[ \ell^*(s) = \left\lceil \frac{1}{2s + d} \log n \right\rceil \quad \text{and} \quad h^*(s) = n^{-\frac{1}{s + 1}}. \]

Remark that
\[ \frac{h^*(s)}{e} \leq h(\ell^*) \leq h^*(s) \]
(18)
and
\[ s \leq m(\ell^*(s)) \leq s + 1. \]
(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):
\[ \max(\| W(\ell^*) \|_{2\Lambda p}, \| W(\ell^*) \|_{\infty}) \leq (s + 2)^{3d/2}. \]
(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:

\[
\|w_m\|_2^2 = (a^{(m)})^\top H_m a^{(m)} = (c_0^{(m)})^\top H_m^{-1} c_0^{(m)} = (m + 1)^2.
\]

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:

\[
\|w_m\|_p^p = \int_0^1 |w_m(u)|^p \, du \\
\leq \left( \int_0^1 |w_m(u)|^{2q} \, du \right)^{\frac{p}{2q}}.
\]

Now, it can be seen by induction that:

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

Indeed, let us denote

\[
u_q = \int_0^1 |w_m(u)|^{2q} \, du \]

\[
= \sum_{i_1, \ldots, i_{2q}} \frac{a_{i_1} \cdots a_{i_{2q}}}{1 + \sum_{k=1}^{2q} i_k} \leq 2^{q-1}(m + 1) \left( \sum_{i_1, \ldots, i_{2q-1}} \frac{a_{i_1} \cdots a_{i_{2q-1}}}{1 + \sum_{k=1}^{2q-1} i_k} \right) \left( \sum_{j_1, \ldots, j_{2q-1}} \frac{a_{j_1} \cdots a_{j_{2q-1}}}{1 + \sum_{k=1}^{2q-1} j_k} \right)
\]

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

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

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 \( \Delta_{d,\epsilon} \) equals to \( 2^{-d} \), we have

\[
E_f^n \|\xi_{W,h}\|_{r,\epsilon} \leq 2^{-\frac{d(2-r)}{2r}} E_f^n \|\xi_{W,h}\|_2 \leq 2^{-\frac{d(2-r)}{2r}} \|W\|_2.
\]
Let us now assume that \( r > 2 \). Using the Rosenthal inequality we have
\[
\mathbb{E}^n f |\xi_{W,h}(t)|^r \leq (C_r^*)^r (V_h)^{r/2} \left\{ (\mathbb{E}^n f K_{W,h}(t, X_1))^r/2 + 2^{r+1} n^{1-r/2} \mathbb{E}^n f |K_{W,h}(t, X_1)|^r \right\}.
\]

Using Jensen and Young inequalities we obtain:
\[
\mathbb{E}^n f \|\xi_{W,h}\|_{r,\varepsilon} \leq \left( \int_{\Delta_{d,\varepsilon}} \mathbb{E}^n f |\xi_{W,h}(t)|^r \, dt \right)^{1/r} \leq C_r^* \left\{ \Lambda_{\varepsilon}(W, h, r) + 2 \|W\|_r (n V_h)^{1 - 2/r} \right\} \leq C_r^* \left\{ \Lambda_{\varepsilon}(W, h, r) + 2 \|W\|_r \right\}.
\]

We have
\[
\Lambda_{\varepsilon}(W, h, r) \leq F_{\infty}^{1/2} \left( \int_{\Delta_{d,\varepsilon}} (V_h \int_{\Delta_d} K_{W,h}^2(t, x) \, dx)^{r/2} \, dt \right)^{1/r} \leq F_{\infty}^{1/2} \|W\|_2.
\]
As a consequence, for all \( r \geq 1 \), we have
\[
\Gamma_{\varepsilon}(W, h, r) \leq C \|W\|_{r,\sqrt{2}}
\]
where \( C \) depends on \( F_{\infty} \) and \( r \).

**4.10 Proof of Lemma 3**

In the following, \( \mathcal{L} \) is either \( \mathcal{L}_{ani} \) or \( \mathcal{L}_{iso} \). Let \( \ell \in \mathcal{L} \). Let \( W \in \{W(\ell), W^*(\ell)\} \). We denote by \( \mathbb{B}_{r'} \) the unit ball of \( L_{r'}(\Delta_{d,\varepsilon}) \) 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:
\[
\bar{g}_\lambda(x) = g_\lambda(x) - \mathbb{E}^n f g_\lambda(X_1) \quad \text{with} \quad g_\lambda(x) = V_h^{1/2} \int_{\Delta_{d,\varepsilon}} \lambda(t) K_{W,h}(t, x) \, dt.
\]

The variable \( Y = \|\xi_{W,h}\|_{r,\varepsilon} \) satisfies
\[
Y = \sup_{\|\lambda\|_{r',\varepsilon} \leq 1} \int_{\Delta_{d,\varepsilon}} \lambda(t) \xi_{W,h}(t) \, dt = \sup_{\|\lambda\|_{r',\varepsilon} \leq 1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{g}_\lambda(X_i)
\]

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:
\[
\|\bar{g}_{\lambda_k}\|_\infty \leq b(W, h, r) \quad (21)
\]
Using the Bousquet inequality, and denoting $\tau$ and $\mathbf{b}(W, h, r) = 2\|W\|_r V_h^{1/r-1/2}$. Now if $r_2$, which depends only on $r$, we have

$$\sup_{k \in \mathbb{N}} \mathbf{E}_k^n g_{\Delta_k}^2(X_1) \leq \sigma^2(W, h, r),$$

where

$$\sigma^2(W, h, r) = \sigma^2 = \begin{cases} \|W\|_r^2 V_h^{\frac{2}{r}-1} & \text{if } 1 \leq r < 2 \\ F_\infty\|W\|_r^2 V_h^{2} & \text{if } r \geq 2. \end{cases}$$

Using (23) we obtain:

$$\mathbf{P}\left( Y - \mathbf{E}_j^n Y \geq \frac{\tau}{2} x + x \right) \leq \exp \left( - \frac{x^2}{2\sigma^2 + \frac{b}{\sqrt{n}} \left( \Gamma_\mathcal{G} \left( \frac{12+\tau}{3} \right) + \frac{2x}{3} \right) } \right) \times \exp \left( - \frac{\tau \Gamma_\mathcal{G} x + \Gamma_\mathcal{G}^2 \tau^2 / 4 }{2\sigma^2 + \frac{b}{\sqrt{n}} \left( \Gamma_\mathcal{G} \left( \frac{12+\tau}{3} \right) + \frac{2x}{3} \right) } \right)$$

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

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

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 $c_1$ and $c_2$ that depend only on $d$, $\tau$, $r$ and $F_\infty$ such that

$$\frac{4 \left( 2\sigma^2 + \frac{b}{\sqrt{n}} \left( \Gamma_\mathcal{G} \left( \frac{12+\tau}{3} \right) + \frac{2x}{3} \right) \right) }{\Gamma_\mathcal{G}^2 \tau^2} \leq c_1(h^*)^d \tau^r + c_2(h^*)^d / r \tag{23}$$

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

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

Using (23) we obtain:

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

where $C_1$ is an absolute positive constant that depends only on $r$, $\tau$ and $F_\infty$.

Using Lemma 2, (21) and (22), we readily obtain that there exists an absolute constant $c_4$ which depends only on $F_\infty$, $\tau$ and $r$ such that:

$$2\sigma^2 + \frac{b}{\sqrt{n}} \left( \Gamma_\mathcal{G} \left( \frac{12+\tau}{3} \right) + \frac{2x}{3} \right) \leq c_4(1 + x)\|W\|_{\tau r^2}^2 \begin{cases} (h^*)^{d(\frac{2}{r}-1)} & \text{if } r < 2 \\ (h^*)^{\frac{2}{d}} & \text{if } r \geq 2 \end{cases}$$

Now if $W \in \{W(\ell), W^*(\ell)\}$, with $\ell \in \mathcal{L}_{\text{ani}}$, there exists a constant $C_2$ that depends on $r$, $\tau$, $F_\infty$ and $W^\circ$ such that

$$\mathbf{P}\left( Y - \mathbf{E}_j^n Y \geq \frac{\Gamma_\mathcal{G} x}{2} + x \right) \leq \exp \left( - \frac{x^2}{C_2(1 + x)} \right) \exp(-C_1 \alpha_n(r)). \tag{24}$$
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 $r$, $\tau$ and $F_\infty$ such that (24) holds. That concludes the lemma.

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

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

$$\|g_{\lambda_k}\|_\infty \leq 2\|g_{\lambda_k}\|_\infty \leq 2 \sup_{x \in \Delta_d} \frac{1}{2} \lambda_k(\cdot, x)_{r,\epsilon} \|K_{W,\ell}(\cdot, x)\|_{r,\epsilon} \leq 2\|W\|_r V_h^{1/r-1/2}.$$  

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

$$\sup_{k \in \mathbb{N}} \mathbb{E}_t^\gamma g_{\lambda_k}^2(X_1) = V_h \sup_{k \in \mathbb{N}} \int_{\Delta_d} \left( \int_{\Delta_d} \lambda_k(t)K_{W,\ell}(t, x) \, dt \right)^2 f(x) \, dx$$

$$\leq V_h \sup_{k \in \mathbb{N}} \int_{\Delta_d} \|K_{W,\lambda}(\cdot, x)\|_{r,\epsilon}^2 \lambda_k(\cdot, x)_{r,\epsilon} f(x) \, dx = V_h^{2/r-1}\|W\|^2.$$  

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

**4.11 Proof of Lemma 4**

We have

$$\mathbb{P} \left( \sum_{x \in \{0,1\}^d} \|\xi_{W^*(\ell),\lambda}\|_{r,\epsilon}^{1/2} \geq 2^d(nV_h)^{1/4} \right) \leq \sum_{x \in \{0,1\}^d} \mathbb{P}_x^\gamma \left( \|\xi_{W^*(\ell),\lambda}\|_{r,\epsilon} \geq 2^d(nV_h)^{1/2} \right)$$

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 $n \geq N_0$:

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

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 $n \geq N_0$:

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

Applying Lemma 3, we obtain the result of the lemma.
4.12 Proof of Lemma 5.

Let \((e_1, \ldots, e_d)\) be the canonical basis of \(\mathbb{R}^d\) and define
\[
v_i(u) = (t_1 + \eta_1 u_1, \ldots, t_i - 1 + \eta_{i-1} u_{i-1}, \quad t_i, t_{i+1} + h_{i+1} u_{i+1}, \ldots, t_d + h_d u_d).
\]
We can write:
\[
\begin{align*}
f(t + h \cdot u) - f(t + \eta \cdot u) &= \sum_{i=1}^{d} f(v_i(u) + h_i u_i e_i) - f(v_i(u) + \eta u_i e_i) \\
&= \sum_{i \in I} f(v_i(u) + h_i u_i e_i) - f(v_i(u)),
\end{align*}
\]
where \(I = \{i = 1, \ldots, d : \eta_i = 0\}\). Using a Taylor expansion of the function \(x \in \mathbb{R} \mapsto f(v_i(u) + x e_i)\) around 0, we obtain:
\[
f(t + h \cdot u) - f(t + \eta \cdot u) = \sum_{i \in I} \sum_{k=0}^{\lfloor s_i \rfloor} D_i^k f(v_i(u)) \frac{(h_i u_i)^k}{k!} + \sum_{i \in I} \frac{(h_i u_i)^{s_i}}{s_i!} \int_0^1 \left[ D_i^{[s_i]} f(v_i(u) + \tau h_i u_i) - D_i^{[s_i]} f(v_i(u)) \right] d\tau.
\]
Using the facts that \(v_i(u)\) does not depend on \(u_i\) and that \(\int_{\Delta_i} W_i(y) y^k dy = 0\) for any \(1 \leq k \leq [s_i]\), Fubini’s theorem implies that:
\[
S^*_W h, \eta (f) = \left( \int_{\Delta,0} \left| \int_{\Delta_d} \left( \prod_{j=1}^{d} W_j(u_j) \right) \sum_{i \in I} I_i(u, h) \, du \right|^p \right)^{1/p}
\]
where
\[
I_i(t, u, h) = \frac{(h_i u_i)^{[s_i]}}{[s_i]!} \int_0^1 \left[ D_i^{[s_i]} f(v_i(u) + \tau h_i u_i) - D_i^{[s_i]} f(v_i(u)) \right] d\tau.
\]
Using Jensen’s inequality and Fubini’s theorem we obtain that:
\[
S^*_W h, \eta (f) = (d \|W\|_1)^{1-1/p} \left( \int_{\Delta_d} \left| J(u, h) \prod_{j=1}^{d} W_j(u_j) \right| \, du \right)^{1/p},
\]
where \(J(u, h) = \sum_{i \in I} \int_{\Delta_{d,0}} |I_i(t, u, h)|^p \, dt\). Now, we study this last term:
\[
J(u, h) \leq \sum_{i \in I} \int_{\Delta_{d,0}} \frac{(h_i u_i)^{1+p s_i}}{[s_i]!} \int_0^1 \left| \frac{D_i^{[s_i]} f(v_i(u) + \tau h_i u_i) - D_i^{[s_i]} f(v_i(u))}{\tau h_i u_i} \right|^p \, d\tau \, dt.
\]
Using a simple change of variables, we obtain:
\[
J(u, h) \leq \sum_{i \in I} \frac{(h_i u_i)^{ps_i}}{[s_i]!} \int_{\Delta_d} \int_0^1 \frac{|D_i^{[s_i]} f(x_1, \ldots, x_{i-1}, \xi, x_{i+1}, \ldots, x_d) - D_i^{[s_i]} f(x)|^p}{|\xi - x_i|^{1+p(s_i - [s_i])}} \, d\xi \, dx.
\]
Since \(u_i \leq 1\) and \(f \in S_{k,p}(L)\) we have:
\[
S^*_W h, \eta (f) \leq d \|W\|_1 \kappa(s) L \left( \sum_{i \in I} h_i^{ps_i} \right)^{1/p}.
\]
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


